

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 (4)

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

Synopsis

At Disaster Prevention Research Institute (DPRI), a large amount of disaster prevention research materials has stored since 1951. The aim of assignment research subject of Division of Technical Affairs is that these base materials are widely exhibited to society, and can be used. In this paper, we show digitization of the materials that is made by the DPRI, and construction of the search engine which can search the disaster prevention research materials on the Internet website at high speed.

In last year, we constructed the search engines to search data of the materials published by the DPRI for users efficiently and conveniently on the Internet website. Additionally, the system can search various (not limited to single) information media integrated in media of the format which conformed to an international standard.

Keywords: database; digitization; GoogleMaps; search engine; GPS

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 and multimedia technologies 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.

To date, DPRI has accumulated a large amount of disaster prevention materials. To further promote disaster prevention research, dissemination of Printed Matters in these materials is necessary. If these printed matters were available on the DPRI Internet website, that would be informative to the public. To disclose the disaster prevention research material that the disaster prevention laboratory possesses on Web, the material should be made an electronic file.

Therefore, we wrote this paper to the methods to exhibit the disaster prevention materials on Internet website, and to enable the users to acquire the materials at any time.

Specifically, user used various disaster prevention information so that we constructed of the data base where that accumulated the meta-data of the fruit of research concerning disaster prevention was

accumulated. Note particular bit of software (Acrobat Reader) is required to open the PDF file. Because PDF format is a *de facto* standard for electronic documents in Japan, we chose the PDF format. Acrobat Reader can be downloaded for free on the website of Adobe Systems Incorporated Company.

2. Construction of the search engines for browsing the disaster prevention information

2.1 The metadata database about the disaster prevention materials

In the DPRI, there are many disaster prevention materials, including publication (annual, newsletter, bulletin etc), seismic map, ground-motion waveform, disaster area map, photo in the disaster area. 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 the disaster prevention, in the present condition that disaster prevention study is not matured as an interdisciplinary field. Additionally, these materials of the disaster prevention study are preserved by two or more different format types.

Then, we draft the construction of the integrated database system as the place where various researchers exchanged information for disaster prevention by this research (Fig. 1).

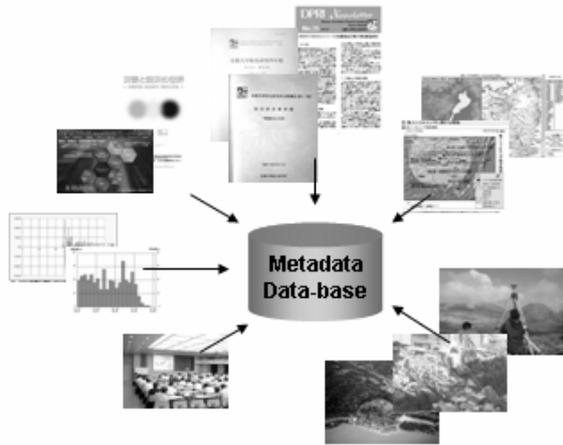


Fig.1 The metadata database about the disaster prevention materials

Fig.2 shows the key map of the metadata database about the disaster prevention materials.

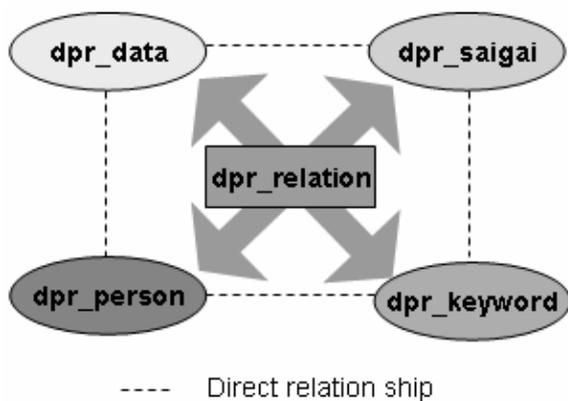


Fig. 2 the key map of the disaster prevention information database

The disaster prevention information database builds up with several database tables, involving “dpr_data” table, “dpr_saigai” table, “dpr_person” table, “dpr_keyword” table, etc. And “dpr_relation” table can correlate one table with the other one table. Here is the particularity of table that the disaster prevention database involved.

“dpr_data” table is kept the metadata of the disaster Prevention research materials, which includes “Title”, “Author”, “Synopsis”, “Data type”, “Media type”, “Location data”, accompanying information, etc.

“dpr_saigai” table is kept the pertinent data for major disaster.

“dpr_person” table is kept the data of author related to the disaster prevention research.

“dpr_keyword table” is kept the keyword which is relevant to the disaster prevention information.

“dpr_relation” is kept the relationship which there s interdependence between one table to the other one table.

2.2 The data type of the disaster prevention materials

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 10 categories. It is shown in Table 1. (document, study, model, voice,image,video, geospatial, internet, etc.)

Table 1 10 kinds of data resource types (XMDB conforming)

Data type	Example
Document	annual, newsletter
Image	photo, disaster prevention map
Audio	conference minutes, inter view
Video	video of natural disaster, three-dimensional numeric simulation
Internet	website for research organization
Model	mathematical simulation model, climate change simulation model
Study	academic conference, research project
Data	basic data for natural disaster, formula for the disaster prevention study
Event	lecture meeting, meeting for reading research papers, catastrophic disaster
Geospatial	Geographic information, land area

2.3 The search engines for browsing the disaster prevention information

Many of users want to retrieve and to acquire information and data concerning disaster prevention that is necessary with short time and easily. Besides, it seems that the users want to search various (not limited to single) information media integrated in media of the format which conformed to an international standard.

Then, to manage disaster prevention information integrated in this research, we constructed the system of the metadata database that was able to retrieve the link to the title, the writing origin, and the disaster prevention information origin etc. quickly (Fig.3 and Fig.4). Note that the data in the database need to be launched about copyright permission from the publishers, or something else entirely.

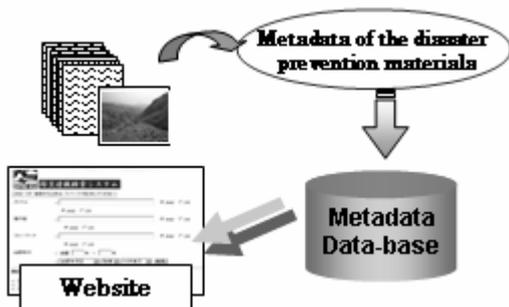


Fig.3 The key map that user retrieve and acquire information and data of the disaster prevention

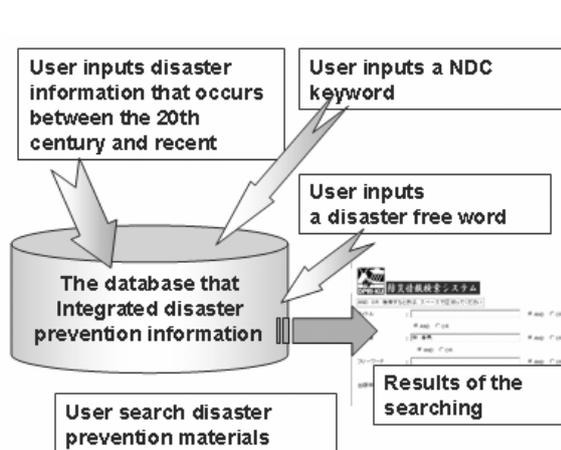


Fig. 4 Flow searching of the disaster prevention materials

The search engine constructed in the previous year, which can search disaster prevention, can be retrieved based on three elements information (Title, Author, and Freeword). After it retrieves, only satisfied disaster

prevention information is extracted by user, and those title and author are displayed on the screen as a retrieval result as shown in Figure 5.



Fig.5 The display in the website of the search engines for browsing the disaster prevention information

If the user wants to see original, detailed metadata of disaster prevention information about one of disaster prevention information in that, the metadata of the disaster prevention information can be displayed on the screen by clicking the display button (Fig.6).



Fig.6 display the results in the website, which of the search engines for browsing the disaster prevention information

In addition, the user can display the data of the source by clicking the link on the screen if the disaster prevention source exists on the website, and it exists in the state that can be displayed besides (Fig. 7).

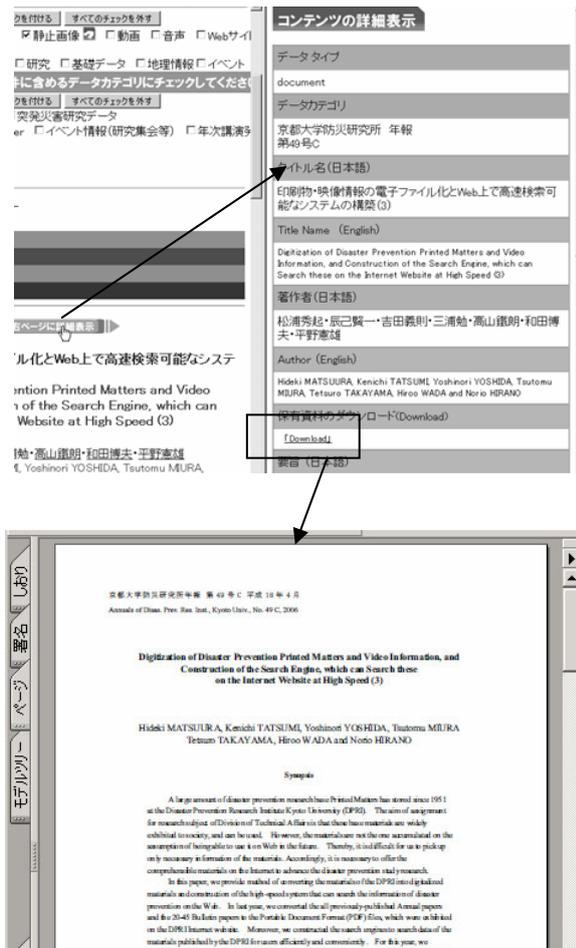


Fig.7 Metadata is shown on the right window, and the disaster prevention source of metadata is shown on the other window

2.4 Linkage with other database system

The disaster prevention information database retrieval system is enhanced linkages and cooperation among other database system, involving bibliography of staff in the DPRI, OPAC(Online Public Access Catalogue), “Next-generation scientific information resources infrastructure”.

The bibliography of staff in the DPRI is the website where it introduces a staff who is on the register in the disaster prevention laboratory, professors emeritus, and a successive head in the DPRI. If the author is an incumbent staff in the DPRI, the author details can be displayed from the link of the author of the retrieval result made by the search engines for browsing the disaster prevention information (Fig. 8).



Fig.8 Linkage with bibliography of staff in the DPRI

OPAC is a bibliography that can be retrieved by the computer-aided because of the abbreviation of Online Public Access Catalogue. The book on OPAC is retrieved automatically by the click of the OPAC retrieval button of the meta data when thinking that the user wants to look for a useful book for the understanding of the meta data of this retrieval result, and the corresponding book list can be extracted (Fig. 9).

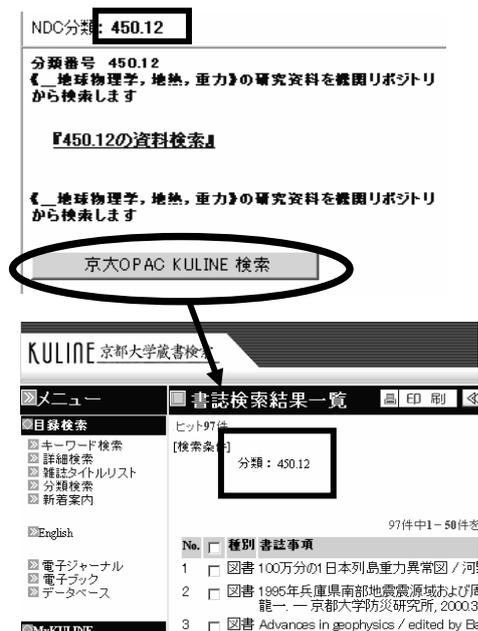


Fig.9 Linkage with OPAC, which is commonly known as KULINE, in the Kyoto university library

“Next-generation scientific information resources infrastructure” is the institutional repository-related projects by National Institute of Informatics (NII).

NII launches a collaborative program for institutional repositories with universities in Japan. NII launched in 2004 as a part of the Cyber Science Infrastructure (CSI) for institutional repository-related projects amework. In 2004, NII collaborated with 6 universities to introduce institutional repositories to Japanese universities. Those universities worked towards implementation and localization of institutional repository (IR) software, contents acquisition, and build experiences. In 2005, NII started a collaborative experiment with 19 university libraries. The project's purpose was deploying and coordinating institutional repositories in Japan. Now, halfway through 2006, 17 institutional repositories are running; these repositories hold a total of 62,423 items (as of June 28, 2006).

The metadata of the annual report that is the material of the disaster prevention laboratory is experimentally published in the Kyoto University Research Information Repository now (Fig. 10).

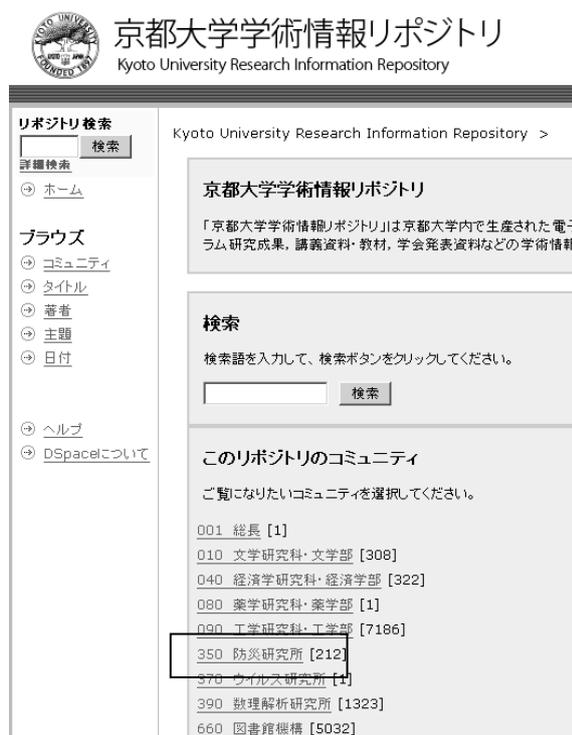


Fig.10 Linkage with Kyoto University Research Information Repository

2.5 Utilization of GoogleMaps

Many of information (a cause and measures) of the disaster that the disaster prevention research institute offers are related to the place where a disaster happened. In a word, the disaster investigation and the research relate to the place where the disaster happens are done. In the research process, information on the acquired real world is analyzed by the system that can integrate and analyze it on the digital map like Geographical Information System (GIS:Geographic Information System) etc. The information that the disaster prevention research institute got by the investigation and a study is analyzed by a geographical information system, which can display it on Web.

However, many of disaster investigation reports, disaster local photographs, and the fruit of research offered the society by the disaster prevention laboratory through the publication and the homepage are often offered by a text document and a digital image. Most of information (disaster investigation report, local photograph and fruit of research) offered, through a publication and a homepage, to society from the disaster prevention research institute are a text document and digital images conventionally.

Then, we constructed the system that was able to offer disaster prevention information by making the position comprehensible in the sight by using GoogleMaps that was one of the systems that used the Ajax(Asynchronous JavaScript+XML) technology that had come to be used actively in recent years.

The system can acquires the location information of the preserved metadata from the metadata database, and can mark the position of the metadata searched by user on GoogleMaps. Concretely, when user wants to know the location information of the meta data in the retrieval result user clicks "「GoogleMaps」で地点表示", so that the position of the disaster prevention information is marked on GoogleMaps (Fig.11)

In this research, we did system construction that was able to do a dynamic mapping of point information on GoogleMaps by making GoogleMaps, PHP, and the Web data base cooperate. And we think about the construction of the system that improves mail with a digital image to the data base automatically in the futures (Fig. 12).



Fig.11 Utilization of GoogleMaps with search engine

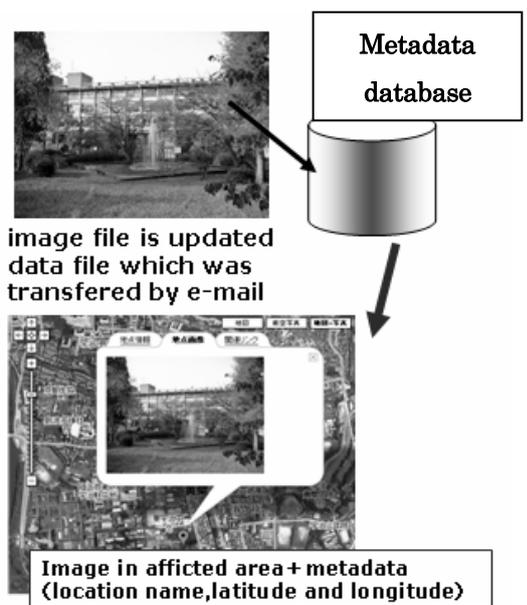


Fig.12 The key map that image file is updated data file which was transferred by e-mail

The system structure is “Apache+PHP +MySQL + Fetchmail+ Procmail”, the soft are the open source software. MySQL is easy to use compared with expensive “business” databases like Oracle or SQLServer. The server script language “PHP” which can directly connected to the database operates as a module of Apache. The structure has high affinity with database, and has smaller burdens than the CGI, which runs with a different process. “Fetchmail+Procmail” can be software to be able to retrieve and to acquire email from various mailserver, and the email are distributed. We constructed the system that can preserve the GPS data and the image data in the data base, when E-mail that attached the GPS data and the digital image data was sent to a specific E-mail Address, The system structure of database. The system structure is ”Apache+PHP+MySQL+ Fetchmail+ Procmail” which is combined both of them, is supported throughout the world, because the two are open source software. GoogleMaps is "Digital map that can be scrolled with the mouse drug" that the Google Co.released free of charge in April, 2005. The improvement is done repeatedly when initial, and the mapping of the point can be done on the map by using GoogleMapsAPI if latitude and the longitude of the system of the survey a land in the world in a certain point are understood now. For instance, when the map where a certain point was done in the mapping can be displayed on Web, the name of the point, latitude, and point information in the longitude etc are necessary. Moreover, if point information used beforehand is stored in the Web data base, the location information mark system that cooperates with the data base search engine can be constructed.

3. Digitizing the disaster prevention research materials

It is necessary to have converted all the disaster prevention research base materials into the form of the electronic file to deliver the materials on the Web. Note that the sequence of digitizing materials of the paper made in the DPRI is outlined below:

- [1] Each page of the papers is converted into digital image (bit map image format) files on an image-scanner and is saved.
- [2] The blots of Digital image files are eliminated on the Picture edit software.
- [3] If the data type of materials are document, ”Digital

image files made in process [1] and [2] are converted to PDF files on WinReader8.0 (a type of Windows OCR application software). And the character strings of the materials are extracted from digital image files on WinReader8.0. The strings are saved in the text file. In addition, we input strings data into a database.

If the data type of materials are image, "Digital image files made in process [1] and [2] are converted to JPEG files. In addition, we stored the metadata that was able to be stored in the disaster prevention information database on the character data of the paper medium in the data base.

4. Conclusions

We constructed the search engines to search data of the materials published by the DPRI for users efficiently and conveniently on the Internet website by the end of 2006 fiscal year. Additionally, the system can search various (not limited to single) information media integrated in media of the format which conformed to an international standard. We make a database of the metadata of the paper medium disaster prevention basic material. And, we constructed the system that was able to offer disaster prevention information by making the position comprehensible in the sight by using GoogleMaps that was one of the systems that used the Ajax(Asynchronous JavaScript+XML) technology that had come to be used actively in recent years.

As a result, any disaster prevention information was efficiently preserved, and the data base system that the user was able to retrieve easily was completed. We have improved previous search engine so that would be able to cooperation with other data bases, involving XMDB, SAIGAI, OPAC, bibliography of staff in the DPRI, Kyoto University Research Information Repository, GoogleMaps, 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. Norio MAKI 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.
- Hideki MATSUURA, Kenichi TATSUMI, Hideo TAGAWA, Yoshinori YOSHIDA, Tsutomu MIURA, Tetsuro TAKAYAMA, Hiroo WADA and Norio, (2005): 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 (2), Annuals of DPRI, Kyoto Univ., No. 48 C, pp. 125-132.
- 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), Annuals of DPRI, Kyoto Univ., No. 49 C, pp.75-80.
- 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 .

Hironori KAWAKATA, Paul YOSHITOMI, Go URAKAWA, Kelly CHAN, Hideki MATSUURA, Kenichi TATSUMI, Takeshi HARA, Munenari AGUSA, Haruo HAYASHI and Yoshiaki KAWATA, (2006):

Building an Integrated Database System of Information on Disaster Hazard, Risk, and Recovery Process-Cross-

Media Database (2), Annuals of DPRI, Kyoto Univ., No. 48 C, pp. 53-60.

Hironori KAWAKATA, Paul YOSHITOMI, Go URAKAWA, Kelly CHAN, Hideki MATSUURA, Kenichi TATSUMI, Takeshi HARA, Munenari AGUSA, Haruo HAYASHI and Yoshiaki KAWATA, (2006):

Building an Integrated Database System of Information on Disaster Hazard, Risk, and Recovery Process-Cross-Media Database (3), Annuals of DPRI, Kyoto Univ., No. 49 C, pp. 69-73.

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 (4)

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18年度では、統合型防災情報検索システムの完成度を高め、職員総覧、OPAC、機関リポジトリといった、他のデータベースとも連携できるように構築を行った。また、データの位置情報表記として、GoogleMapsを採用し、データの地理的な表現方法の新しい可能性を示した。

キーワード:データベース; 電子ファイル化; GoogleMaps; 検索システム; GPS