

A Prototype of PubMed Central Japan

Hideaki Sugawara¹

hsugawar@genes.nig.ac.jp

Takashi Gojobori¹

tgojobor@genes.nig.ac.jp

Takeshi Konno²

konno.takeshi@jp.fujitsu.com

Yasumasa Shigemoto²

shigemoto.yasum@jp.fujitsu.com

Masahito Yamaguchi²

yamaguchi.ma-07@jp.fujitsu.com

¹ Center for Information Biology and DDBJ, National Institute of Genetics (NIG), 1111 Yata, Mishima, Shizuoka 411-8540, Japan

² Fujitsu Limited, 1-17-25 Shin-kamata, Ohta-ku, Tokyo 144-8588, Japan

Keywords: database, XML, XSL, DTD, PDF, bibliography, electronic publishing

1 Introduction

“PubMed Central (PMC) is the U.S. National Library of Medicine’s digital archive of life sciences journal literature. Access to PMC is free and unrestricted.” [3]. It covers 78 journals and also 57 journals of BioMed Central [1]. It provides not only bibliographic information and abstracts but also full texts including figures and tables. Thus PMC is a rich and valuable information resource for research and development in biology and medicine.

2 Method and Results

The National Center for Biotechnology Information [2] provided us with files of 190 sample articles to set up PubMed Central Japan. They consist of XML files, Document Type Definition files, entity files, image files and PDF files. The total file size is 200Mbytes.

We developed a search engine by JAVA and implemented into a conventional Pentium machine. A list of articles hit by keyword(s) is displayed as shown in Fig. 1. From the list, the user can choose the abstract, the full text or the PDF to display. The full text is reconstructed from the XML document by use of XSL. The full text representation of the article 1 of Fig. 1 is partly introduced in Fig 2.1–2.3.

3 Discussion

The full set of PubMed Central will be open to the public from the DDBJ Web site after in-house evaluation. The Web services and user interface in Japanese will be prepared too. The total volume will be 1.5Tbytes to 2.0 Tbytes depending on the volume of image files, if all the articles are retrospectively digitized. The volume certainly expands yearly. The expansion rate will be even higher, if more journals join PMC.



Figure 1: A sample list of articles hit by keyword(s).

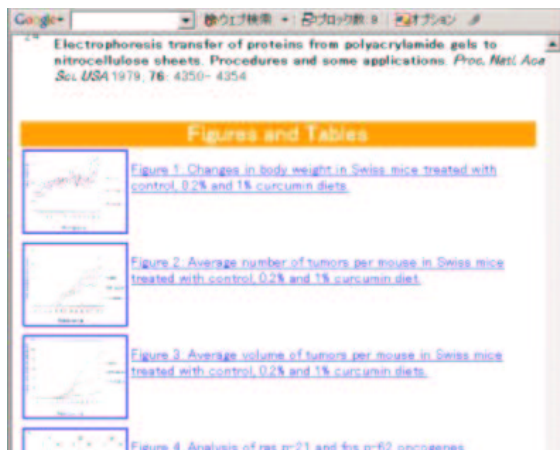


Figure 2-1 in the upper left:
The title page of the full text representation in PubMed Central Japan

Figure 2-2 in the upper right:
The last section of the full text representation in PubMed Central Japan

Figure 2-3 in the left:
In PubMed Central Japan, figures follow the last section in Fig. 2-2

Acknowledgments

The authors would like to express their sincere thanks to Drs. David Lipman and Ed Sequeira of the National Center for Biotechnology Information, US for their suggestions and supports.

This study is partly supported by the program "Research and Development of Biological Portal Site of the New Generation" through the Special Coordination Funds for Promoting Science and Technology from the Ministry of Education, Culture, Sports, Science and Technology, the Japanese Government in 2003.

References

- [1] <http://www.biomedcentral.com/>
- [2] <http://www.ncbi.nlm.nih.gov/>
- [3] <http://www.pubmedcentral.nih.gov/>