

# デジタルドキュメント(4)

高久雅生

2013年5月9日(木)3・4時限

# 前回の出席カードから（質疑）

- 第2回レポート課題の対象は、機関リポジトリの文献でもよいですか？
  - → はい、問題ありません。
- 授業開始
  - 遅れないよう注意します。
- 配布資料（レジュメ）
  - 両面印刷は行いません。

# 本日のお品書き

- 前回の復習
- 学術分野のデジタルドキュメント(続き)
  - 学術論文の事例にみるドキュメントのサービス機能とプラットフォーム
- 各種ドキュメントフォーマット

# (前回の復習 = ふりかえり)

- 学術分野のデジタルドキュメントの実例
- オンラインジャーナルの様態
  - ランディングページ
  - 全文ファイル(PDF, HTML)
  - 各種機能
  - 論文のID:DOI
- 事例(各社プラットフォーム)
- 第2回レポート課題:来週5/16〆切

# 学術分野のデジタルドキュメント (3)

# 学術分野におけるオンラインジャーナルの事例

- 例1:『Science』 The American Association for the Advancement of Science (AAAS)
  - ScienceMag.org
- 例2:『Nature』 Nature Publishing Group
  - Nature.com
- 例3:『Journal of American Society for Information Science and Technology』 American Society for Information Science and Technology (ASIS&T)
  - Wiley.com
- 例4:『Information Processing and Management』 Elsevier
  - Sciencedirect.com
- 例5:『Information Retrieval』 Springer
  - Springer.com
- 例6:『Proceedings of the IIiX 2008』 ACM
  - ACM.org
- 例7:『情報管理』科学技術振興機構 (JST)
  - Jstage.jst.go.jp
- 例8:『社会教育』全日本社会教育連合会
  - つくばリポジトリ – Tulips-R
- 例9:『図書館情報メディア研究』筑波大学図書館情報メディア研究科
  - つくばリポジトリ – Tulips-R
- 例10:『情報の科学と技術』情報科学技術協会
  - CiNii Articles

# オンラインジャーナルプラットフォーム の差異

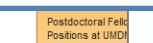
- ・ どのようなフォーマットで提供されているか？
- ・ プラットフォームは誰が運営しているか？
  - 論文の出版/刊行の主体と異なるか？
- ・ どのようなフローで作られているか？
  - 電子化は誰が担当しているか？
- ・ どのような機能があるか？

# 例 1

# Tomoki Nakamura, Takaaki Noguchi, Masahiko Tanaka, et al.: “Itokawa Dust Particles: A Direct Link Between S-Type Asteroids and Ordinary Chondrites”. *Science*, Vol.333, No.6046, 2011, p.1113-1116.

<http://dx.doi.org/10.1126/science.1207758>

Of the 40 particles removed by tapping (diameters ranging from 30 to 180  $\mu\text{m}$ ) that were analyzed by x-ray



**Fig. 1. (A to D)** Backscattered electron (BSE) images of RA-QD02-0030 (A), RA-QD02-0024 (B), RA-QD02-0013 (C), and RA-QD02-0027 (D).

www.sciencemag.org SCIENCE VOL 333 26 AUGUST 2011

卷之三

# 例2

## Reka Albert, Hawoong Jeong, Albert-Laszlo Barabasi: “Internet: Diameter of the World-Wide Web”. *Nature*, Vol.401, p.1113-1116.

We're throwing a PC party during Phusion Fest.

Full text access provided to University of Tsukuba Library by Library

My account | Submit manuscript | Log in | Subscribe

nature International weekly journal of science

Search Advanced search

Journal home > Archive > Brief Communications > Full Text

Brief Communications

Nature 401, 130-131 (9 September 1999) | doi:10.1038/43601

Internet: Diameter of the World-Wide Web

Reka Albert<sup>1</sup>, Hawoong Jeong<sup>1</sup> & Albert-László Barabási<sup>1</sup>

Despite its increasing role in communication, the World-Wide Web remains uncontrolled: any individual or institution can create a website with any number of documents and links. This unregulated growth leads to a huge and complex web, which becomes a large directed graph whose vertices are documents and whose edges are links (URLs) that point from one document to another. The topology of this graph determines the web's connectivity and consequently how effectively we can locate information on it. But its enormous size (estimated to be at least  $8 \times 10^8$  documents<sup>1</sup>) and the continual changing of documents and links make it impossible to catalogue all the vertices and edges.

The extent of the challenge in obtaining a complete topological map of the web is illustrated by the limitations of the commercial search engines: Northern Light, the search engine with the largest coverage, is estimated to index only 38% of the web<sup>2</sup>. Although much work has been done to map and characterize the Internet's infrastructure<sup>3</sup>, little is known about what really matters in the search for information — the topology of the web. Here we take a step towards filling this gap: we have used local connectivity measurements to construct a topological model of the World-Wide Web, which has enabled us to explore and characterize its large-scale properties.

To determine the local connectivity of the web, we constructed a robot that adds to its database all URLs found on a document and recursively follows these to retrieve the related documents and URLs. We used the data collected to determine the probabilities  $P_{\text{out}}(k)$  and  $P_{\text{in}}(k)$  that a document has  $k$  outgoing and incoming links, respectively. We find that both  $P_{\text{out}}(k)$  and  $P_{\text{in}}(k)$  follow a power law over several orders of magnitude, remarkably different not only from the Poisson distribution predicted by the classical theory of random graphs<sup>3,4</sup>, but also from the bounded distribution found in models of random networks<sup>5</sup>.

The power-law tail indicates that the probability of finding documents with a large number of links is significant, as the network connectivity is dominated by highly connected web pages. Similarly, for incoming links, the probability of finding very popular addresses, to which a large number of other documents point, is non-negligible, an indication of the flocking nature of the web. Furthermore, while the owner of each web page has complete freedom in choosing the number of links on a document and the addresses to which they point, the overall system obeys scaling laws characteristic only of highly interactive self-organized systems and critical phenomena<sup>6</sup>.

To investigate the connectivity and the large-scale topological properties of the

1. Chase, M. R., Möller, C., Keselci, R. & Bawa, K. S. *Nature* 383, 398–399 (1996).  
2. Naon, J. D., Allen-Herce, E. & Hamrick, J. L. *Nature* 391, 685–687 (1998).  
3. Crawford, T. J. *Hereditas* 52, 273–283 (1984).  
4. Chambers, J. Q., Higuchi, N. & Schimel, J. P. *Nature* 391, 135–136 (1998).  
5. Rebold, X. & Zeyl, C. *Hereditas* 72, 132–140 (1994).  
6. McCauley, D. E. *Trends Ecol. Evol.* 10, 198–202 (1995).  
7. Hamilton, M. B. *Mol. Ecol.* 8, 521–522 (1999).  
8. Taberlet, P. et al. *Plant Mol. Biol.* 17, 1105–1109 (1991).  
9. Weir, B. S. *Genetic Data Analysis II* (Sinauer, Sunderland, Massachusetts, 1996).  
10. Rand, D. M. *Conserv. Biol.* 10, 665–671 (1996).

## brief communications

1. Chase, M. R., Möller, C., Keselci, R. & Bawa, K. S. *Nature* 383, 398–399 (1996).  
2. Naon, J. D., Allen-Herce, E. & Hamrick, J. L. *Nature* 391, 685–687 (1998).  
3. Crawford, T. J. *Hereditas* 52, 273–283 (1984).  
4. Chambers, J. Q., Higuchi, N. & Schimel, J. P. *Nature* 391, 135–136 (1998).  
5. Rebold, X. & Zeyl, C. *Hereditas* 72, 132–140 (1994).  
6. McCauley, D. E. *Trends Ecol. Evol.* 10, 198–202 (1995).  
7. Hamilton, M. B. *Mol. Ecol.* 8, 521–522 (1999).  
8. Taberlet, P. et al. *Plant Mol. Biol.* 17, 1105–1109 (1991).  
9. Weir, B. S. *Genetic Data Analysis II* (Sinauer, Sunderland, Massachusetts, 1996).  
10. Rand, D. M. *Conserv. Biol.* 10, 665–671 (1996).

### Internet

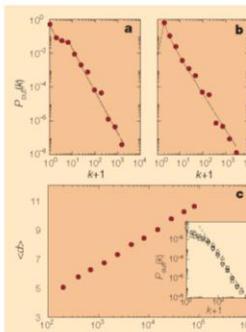
## Diameter of the World-Wide Web

Despite its increasing role in communication, the World-Wide Web remains uncontrolled: any individual or institution can create a website with any number of documents and links. This unregulated growth leads to a huge and complex web, which becomes a large directed graph whose vertices are documents and whose edges are links (URLs) that point from one document to another. The topology of this graph determines the web's connectivity and consequently how effectively we can locate information on it. But its enormous size (estimated to be at least  $8 \times 10^8$  documents<sup>1</sup>) and the continual changing of documents and links make it impossible to catalogue all the vertices and edges.

The extent of the challenge in obtaining a complete topological map of the web is illustrated by the limitations of the commercial search engines: Northern Light, the search engine with the largest coverage, is estimated to index only 38% of the web<sup>2</sup>. Although much work has been done to map and characterize the Internet's infrastructure<sup>3</sup>, little is known about what really matters in the search for information — the topology of the web. Here we take a step towards filling this gap: we have used local connectivity measurements to construct a topological model of the World-Wide Web, which has enabled us to explore and characterize its large-scale properties.

To determine the local connectivity of the web, we constructed a robot that adds to its database all URLs found on a document and recursively follows these to retrieve the related documents and URLs. We used the data collected to determine the probabilities  $P_{\text{out}}(k)$  and  $P_{\text{in}}(k)$  that a document has  $k$  outgoing and incoming links, respectively. We find that both  $P_{\text{out}}(k)$  and  $P_{\text{in}}(k)$  follow a power law over several orders of magnitude, remarkably different not only from the Poisson distribution predicted by the classical theory of random graphs<sup>3,4</sup>, but also from the bounded distribution found in models of random networks<sup>5</sup>.

To investigate the connectivity and the large-scale topological properties of the



**Figure 1** Distribution of links on the World-Wide Web. **a**, Outgoing links (URLs found on an HTML document); **b**, incoming links (URLs pointing to a certain HTML document). Data were obtained from the complete map of the nd.edu domain, which contains 325,729 documents and 1,469,680 links. Dotted lines represent analytical fits used as input distributions in constructing the topological model of the web; the tail of the distributions follows  $P(k) \propto k^{-\gamma}$ , with  $\gamma_{\text{out}} = 2.45$  and  $\gamma_{\text{in}} = 2.1$ . **c**, Average of the shortest path between two documents as a function of system size, as predicted by the model. To check the validity of our predictions, we determined  $\langle d_{\text{out,in}}$  for  $d$  documents in the domain nd.edu. The measured  $\langle d_{\text{out,in}} \rangle = 11.2$  agrees well with the prediction  $\langle d_{\text{out,in}} \rangle = 11.6$  obtained from our model. To show that the power-law tail of  $P(k)$  is a universal feature of the web, the inset shows  $P_{\text{out}}(k)$  obtained by staring from whitehouse.gov (squares), yahoo.com (triangles) and smu.ac.kr (inverted triangles). The slope of the dashed line is  $\gamma_{\text{out}} = 2.45$ , as obtained from nd.edu in **a**.

incoming links, the probability of finding very popular addresses, to which a large number of other documents point, is non-negligible, an indication of the flocking nature of the web. Furthermore, while the owner of each web page has complete freedom in choosing the number of links on a document and the addresses to which they point, the overall system obeys scaling laws characteristic only of highly interactive self-organized systems and critical phenomena<sup>6</sup>.

To investigate the connectivity and the large-scale topological properties of the web, we constructed a directed random

$k_i + 1$  outgoing (or incoming) links is less than  $N P_{\text{out}}(k_i + 1)$  (or  $N P_{\text{in}}(k_i + 1)$ ).

A particularly important quantity in a search process is the shortest path between two documents,  $d$ , defined as the smallest number of URL links that must be followed to navigate from one document to the other. We find that the average of  $d$  over all pairs of vertices is  $\langle d \rangle = 0.35 + 2.06 \log(N)$  (Fig. 1c), indicating that the web forms a small-world network<sup>5,7</sup>, which characterizes social or biological systems. For  $N = 8 \times 10^8$ ,  $\langle d_{\text{web}} \rangle = 18.59$ ; that is, two randomly chosen documents on the web are on average 19 clicks away from each other.

For a given  $N$ ,  $d$  follows a gaussian distribution so  $\langle d \rangle$  can be interpreted as the diameter of the web, a measure of the shortest distance between any two points in the system. Despite its huge size, our results indicate that the web is a highly connected graph with an average diameter of only 19 links. The logarithmic dependence of  $\langle d \rangle$  on  $N$  is important to the future potential of the web: we find that the expected 1,000% increase in the size of the web over the next few years will change  $\langle d \rangle$  very little, from 19 to only 21.

The relatively small value of  $\langle d \rangle$  indicates that an intelligent agent, who can interpret the links and follow only the relevant one, can find the desired information quickly by navigating the web. But this is not the case for a robot that locates the information based on matching strings. We find that such a robot, aiming to identify a document at distance  $\langle d \rangle$ , needs to search  $M(\langle d \rangle) \approx 0.53 \times 10^{0.92}$  documents, which, with  $N = 8 \times 10^8$ , leads to  $M = 8 \times 10^7$ , or 10% of the whole web. This indicates that robots cannot benefit from the highly connected nature of the web, their only successful strategy being to index as much of the web as possible.

The scale-free nature of the link distributions indicates that collective phenomena play a previously unsuspected role in the development of the web<sup>8</sup>, forcing us to look beyond the traditional random graph models<sup>5–7</sup>. A better understanding of the web's topology, aided by modelling efforts, is crucial in developing search algorithms or designing strategies for making information widely accessible on the World-Wide Web. Fortunately, the surprisingly small diameter of the web means that all that information is just a few clicks away.

Reka Albert, Hawoong Jeong,  
Albert-László Barabási  
Department of Physics, University of Notre Dame,  
Notre Dame, Indiana 46556, USA

<http://dx.doi.org/10.1038/43601>

**Figure 1: Distribution of links on the World-Wide Web.**

**a**, Outgoing links (URLs found on an HTML document); **b**, incoming links (URLs pointing to a certain HTML document). Data were obtained from the complete map of the nd.edu domain, which contains 325,729 documents and 1,469,680 links.

**NPG journals by subject area**

Chemistry  
Chemistry

# 例3 Natsuo Onodera, Mariko Iwasawa, Nobuyuki Midorikawa, et al.: “A method for eliminating articles by homonymous authors from the large number of articles retrieved by author search”. Journal of the American Society for Information Science and Technology, 2011, Vol.62, No.4, p.677-690

The screenshot shows the article "A method for eliminating articles by homonymous authors from the large number of articles retrieved by author search" by Natsuo Onodera, Mariko Iwasawa, Nobuyuki Midorikawa, and Fuyuki Yoshikane. The article was published in the Journal of the American Society for Information Science and Technology, Volume 62, Issue 4, pages 677-690, April 2011. The DOI is 10.1002/asi.21491. The page includes a sidebar with journal tools like Get New Content Alerts, Get RSS feed, Save to My Profile, Get Sample Copy, and Recommend to Your Librarian. It also has sections for Publications, Browse by Subject, Resources, and About Us.

## A Method for Eliminating Articles by Homonymous Authors From the Large Number of Articles Retrieved by Author Search

Natsuo Onodera, Mariko Iwasawa, Nobuyuki Midorikawa, and Fuyuki Yoshikane  
Graduate School of Library, Information and Media Studies, University of Tsukuba, 1-2, Kasuga, Tsukuba, Ibaraki 305-8550, Japan. E-mail: {onodera, miwasawa, midorika, fuyukij}@slis.tsukuba.ac.jp

Kou Amano  
Bioresource Information Division, RIKEN BioResource Center, 3-1-1, Koyadai, Tsukuba, Ibaraki 305-0074, Japan. E-mail: amano@brc.riken.jp

Yutaka Ootani and Tadashi Kodama  
Toho University Medical Media Center, 5-21-16, Omori-Nishi, Ota-ku, Tokyo 143-8540, Japan.  
E-mail: {y-ootani, kodamat}@mnc.toho-u.ac.jp

Yasuhiro Kiyama  
Juntendo University Library, 2-2-26, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.  
E-mail: kiyama@juntendo.ac.jp

Hiroyuki Tsunoda  
Department of Culture and Language, Shokei University, 6-5-1, Nirenoki, Kumamoto 861-8538, Japan.  
E-mail: tsunoda@shokei-gakuen.ac.jp

Shizuka Yamazaki  
International Medical Information Center, 35, Shinanomachi, Shinjuku-ku, Tokyo 160-0016, Japan.  
E-mail: yshizu@plum.ocn.ne.jp

This paper proposes a methodology which discriminates articles by the target authors ("true" articles) from those by other homonymous authors ("false" articles). Author name searches for 2,595 "source" authors in six subject fields retrieved about 629,000 articles. In order to extract true articles from the large amount of the retrieved articles, including many false ones, two filtering stages were applied. At the first stage any retrieved article was eliminated as false if either its affiliation addresses had little similarity to those of its source article or there was no citation relationship between the journal of the retrieved article and that of its source article. At the second stage, a sample of retrieved articles was subjected to manual judgment, and utilizing the judgment results, discrimination functions based on logistic regression were defined. These discrimination functions demonstrated both the recall ratio and the precision of about 95% and the accuracy (correct answer ratio) of 90–95%. Existence of common coauthor(s), address similarity, title words similarity, and interjournal citation relationships between the retrieved and source articles were found to be the effective discrimination predictors. Whether or not the source author was from a specific country was also one of the important predictors. Furthermore, it was shown that a retrieved article is almost certainly true if it was cited by, or cocited with, its source article. The method proposed in this study would be effective when dealing with a large number of articles whose subject fields and affiliation addresses vary widely.

<http://dx.doi.org/10.1002/asi.21491>

found from the articles published during the period between 1999 and 2002 in the Web of Science (WoS), and among these names 65% were Asian (54% Japanese). Obviously, many of the author names correspond to different authors. The problem with Western author names is comparatively less critical but nevertheless exists and cannot be ignored. Aksnes (2008) showed that if the 31,135 researchers registered in the Norwegian Research Personnel Register (Ver. 2005) were listed in the ISI style, 4,362 (14%) homonymous authors would be found.

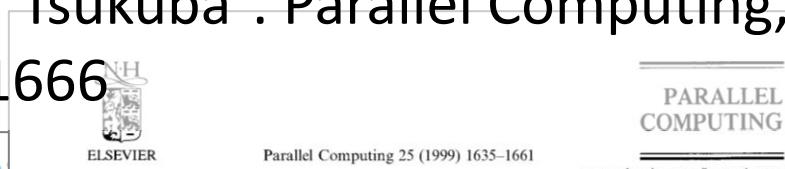
A number of methods exist to discriminate the wanted articles from the unwanted ones by homonymous authors among the articles retrieved through author search. The best method is to obtain a list of the papers published by the target research group and compare the

defined. These discrimination functions demonstrated both the recall ratio and the precision of about 95% and the accuracy (correct answer ratio) of 90–95%. Existence of common coauthor(s), address similarity, title words similarity, and interjournal citation relationships between the retrieved and source articles were found to be the effective discrimination predictors. Whether or not the source author was from a specific country was also one of the important predictors. Furthermore, it was shown that a retrieved article is almost certainly true if it was cited by, or cocited with, its source article. The method proposed in this study would be effective when dealing with a large number of articles whose subject fields and affiliation addresses vary widely.

例4

Kisaburo Nakazawa, Hiroshi Nakamura, Taisuke Boku, Ikuo Nakata, Yoshiyuki Yamashita: “CP-PACS: A massively parallel processor at the University of Tsukuba”. *Parallel Computing*, 1999, Vol.25, No.13, p.1635-1666

The screenshot shows the ScienceDirect website interface. At the top, there are links for Hub, ScienceDirect, Scopus, AI, Register, Login, and Go to. Below that, it says "Brought to you by: University of Tsukuba Li". The main navigation bar includes Home, Publications, Search, My settings, My alerts, Help, Export citation, PDF (1227 K), More options..., and a search bar for "Search ScienceDirect". The article title "Parallel Computing" is displayed, along with the volume information "Volume 25, Issues 13–14, December 1999, Pages 1635–1661". The authors' names are listed as Kisaburo Nakazawa<sup>a,\*</sup>, Hiroshi Nakamura<sup>b</sup>, Taisuke Boku<sup>c</sup>, Ikuo Nakata<sup>d</sup>, Yoshiyuki Yamashita<sup>c</sup>. The abstract, keywords, and references sections are also visible.



## CP-PACS: A massively parallel processor at the University of Tsukuba

Kisaburo Nakazawa<sup>a,\*</sup>, Hiroshi Nakamura<sup>b</sup>, Taisuke Boku<sup>c</sup>, Ikuo Nakata<sup>d</sup>, Yoshiyuki Yamashita<sup>c</sup>

<sup>a</sup> Department of Electronics and Computer Science, Meisei University, 2-590 Nagafuchi, Ohme, Tokyo 198-8655, Japan

<sup>b</sup> Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan

<sup>c</sup> Institute of Information Sciences and Electronics, University of Tsukuba, 1-1-1 Tennohdai, Tsukuba, Ibaraki 305-8577, Japan

<sup>d</sup> University of Library and Information Science, 1-2 Kasuga, Tsukuba, Ibaraki 305-8550, Japan

Received 2 October 1998; received in revised form 16 December 1998

### Abstract

Computational Physics by Parallel Array Computer System (CP-PACS) is a massively parallel processor developed and in full operation at the Center for Computational Physics at the University of Tsukuba. It is an MIMD machine with a distributed memory, equipped with 2048 processing units and 128 GB of main memory. The theoretical peak performance of CP-PACS is 614.4 Gflops. CP-PACS achieved 368.2 Gflops with the Linpack benchmark in 1996, which at that time was the fastest Gflops rating in the world.

CP-PACS has two remarkable features. Pseudo Vector Processing feature (PVP-SW) on each node processor, which can perform high speed vector processing on a single chip superscalar microprocessor; and a 3-dimensional Hyper-Crossbar (3-D HXB) Interconnection network, which provides high speed and flexible communication among node processors.

In this article, we present the overview of CP-PACS, the architectural topics, some details of hardware and support software, and several performance results. © 1999 Elsevier Science B.V. All rights reserved.

**Keywords:** Massively parallel processor; Distributed memory; Processor architecture; Interconnection network; Benchmark

[http://dx.doi.org/10.1016/S0167-8191\(99\)00078-2](http://dx.doi.org/10.1016/S0167-8191(99)00078-2)

\* Corresponding author.

# 例5 Omar Alonso: “Implementing crowdsourcing-based relevance experimentation: an industrial perspective”. Information Retrieval, 2013, Vol.16, No.2p.101-120

SpringerLink

Search

Home • Contact Us

» Download PDF (669 KB)   » View Article

Information Retrieval  
April 2013, Volume 16, Issue 2, pp 101–120

## Implementing crowdsourcing-based relevance experimentation: an industrial perspective

Omar Alonso

» Download PDF (669 KB)   » View Article

### Abstract

Crowdsourcing has emerged as a viable platform for conducting different types of relevance evaluation. The main reason behind this trend is that it makes possible to conduct experiments extremely fast, with good results at a low cost. However, like in any experiment, there are several implementation details that would make an experiment work or fail. To gather useful results, clear instructions, user interface guidelines, content quality, inter-rater agreement metrics, work quality, and worker feedback are important characteristics of a successful crowdsourcing experiment. Furthermore, designing and implementing experiments that require thousands or millions of labels is different than conducting small scale research investigations. In this paper we outline a framework for conducting continuous crowdsourcing experiments, emphasizing aspects that should be of importance for all sorts of tasks. We illustrate the value of characteristics that can impact the overall outcome using examples based on TREC, INEX, and Wikipedia data sets.

Share

Facebook Twitter LinkedIn

Within this Article

- » Introduction
- » Related work
- » Development framework
- » Operational considerations for experiments in production
- » Experiment design
- » Quality control
- » Content aspects
- » Concluding remarks and outlook
- » References
- » References

Other actions

- » Export citations
- » Register for Journal Updates
- » About This Journal
- » Reprints and Permissions

Sign in English Academic

Inf Retrieval (2013) 16:101–120  
DOI 10.1007/s10791-012-9204-1

CROWD SOURCING

## Implementing crowdsourcing-based relevance experimentation: an industrial perspective

Omar Alonso

Received: 16 May 2011 / Accepted: 21 June 2012 / Published online: 20 July 2012  
© Springer Science+Business Media, LLC 2012

### Abstract

Crowdsourcing has emerged as a viable platform for conducting different types of relevance evaluation. The main reason behind this trend is that it makes possible to conduct experiments extremely fast, with good results at a low cost. However, like in any experiment, there are several implementation details that would make an experiment work or fail. To gather useful results, clear instructions, user interface guidelines, content quality, inter-rater agreement metrics, work quality, and worker feedback are important characteristics of a successful crowdsourcing experiment. Furthermore, designing and implementing experiments that require thousands or millions of labels is different than conducting small scale research investigations. In this paper we outline a framework for conducting continuous crowdsourcing experiments, emphasizing aspects that should be of importance for all sorts of tasks. We illustrate the value of characteristics that can impact the overall outcome using examples based on TREC, INEX, and Wikipedia data sets.

**Keywords** Relevance assessment & evaluation · Crowdsourcing · Experiment design · Methodology

### 1 Introduction

Crowdsourcing has been used for a wide range of applications, from relevance evaluation (Alonso and Mizzaro 2012), machine learning (Alonso et al. 2009) and natural language processing (Snow et al. 2008), just to name a few. The cost of running experiments in conjunction with the flexibility of the editorial approach at a larger scale, makes this approach very attractive for quickly testing new ideas. It is also possible to introduce experimentation early in the system development cycle.

Now that crowdsourcing is being adopted by industry and academia, people are noticing that its deployment in practice is not that simple. Tasks have to be designed carefully with

<http://dx.doi.org/10.1007/s10791-012-9204-1>

#### About this Article

**Title**  
Implementing crowdsourcing-based relevance experimentation: an industrial perspective

**Topics**  
» Information Storage and Retrieval  
» Document Preparation and Text Processing  
» Data Mining  
» UK e-Learning

**Authors**  
Omar Alonso  (1)  
**Author Affiliations**

# 例6 Hitoshi Terai, Hitomi Saito, Yuka Egusa, Masao Takaku, Makiko Miwa, Noriko Kando: “Differences between informational and transactional tasks in information seeking on the web”.

ACM DIGITAL LIBRARY University of Tsukuba

Differences between informational and transactional tasks in information seeking on the Web

Full Text: [PDF](#)

Authors: Hitoshi Terai Tokyo Denki University, Chiba, Japan  
Hitomi Saito Aichi University, Nagoya, Japan  
Yuka Egusa National Institute for Educational Policy Research, Tokyo, Japan  
Masao Takaku Research Organization of Information and Systems, Tokyo, Japan  
Makiko Miwa National Institute of Multimedia Education, Chiba, Japan  
Noriko Kando National Institute of Informatics, Tokyo, Japan

Published in:  
• Proceeding  
[IIIX '08 Proceedings of the second international symposium on Information interaction in context](#)  
Pages 152-159  
ACM New York, NY, USA ©2008  
[table of contents](#) ISBN: 978-1-60558-310-5 doi:>[10.1145/1414694.1414728](https://doi.org/10.1145/1414694.1414728)

Bibliometrics  
- Downloads (6 Weeks): 5  
- Downloads (12 Months): 68  
- Downloads (cumulative): 442  
- Citation Count: 6

Tags:  
analysis  
retrieval  
behavior  
web search

SIGN IN SIGN UP

Proceedings of the second international symposium on Information interaction in context, 2008, p.152-159

## Differences between Informational and Transactional Tasks in Information Seeking on the Web

Hitoshi Terai  
Tokyo Denki University  
2-1200 Muzai Gakuen-dai  
Inzai-shi, Chiba 207-1382,  
Japan  
terai@sie.dendai.ac.jp

Hitomi Saito  
Aichi University of Education  
1 Hirosawa, Iiga-cho  
Kariya-shi, Aichi 448-8542,  
Japan  
hsaito@aeucc.aichi-edu.ac.jp

Yuka Egusa  
National Institute for  
Educational Policy Research  
3-2-2 Kasumigaseki  
Chiyoda-ku, Tokyo 100-8951,  
Japan  
yuka@nier.go.jp

Masao Takaku  
Research Organization of  
Information and Systems  
2-1-2 Hitotsubashi  
Chiyoda-ku, Tokyo 101-8430,  
Japan  
masao@nii.ac.jp

Makiko Miwa  
National Institute of  
Multimedia Education  
2-12 Wakaba  
Mihamachi, Chiba 261-0014,  
Japan  
miwamaki@nime.ac.jp

Noriko Kando  
National Institute of  
Informatics  
2-1-2 Hitotsubashi  
Chiyoda-ku, Tokyo 101-8430,  
Japan  
kando@nii.ac.jp

 Feedback | Switch to [single page view](#) (no tabs)

[Abstract](#) [Authors](#) [References](#) [Cited By](#) [Index Terms](#) [Publication](#) [Reviews](#) [Comments](#) [Table of Contents](#)

We examine the influence of task types on information-seeking behaviors on the Web by using screen capture logs and data. Eleven participants performed two different types of web search, an informational task and a transactional task, aloud protocols and behaviors were recorded. Analyses of the screen capture logs showed that the task type affected information-seeking behaviors. In the transactional task, participants visited more web pages than for the informational task, reading time for each page was shorter than in the informational task. A preliminary analysis of eye-movement data for revealed characteristics of the scanpaths followed in search result pages as well as the distribution of lookzones for each

## ABSTRACT

We examine the influence of task types on information-seeking behaviors on the Web by using screen capture logs and eye-movement data. Eleven participants performed two different types of web search, an informational task and a transactional task, and their aloud protocols and behaviors were recorded. Analyses of the screen capture logs showed that the task type affected the participants' information-seeking behaviors. In the transactional task, participants visited more web pages than for the informational task, but their reading time for each page was shorter than in the informational task. A preliminary analysis of eye-movement data for nine participants revealed characteristics of the scanpaths followed in search result pages as well as the distribution of lookzones for each task.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

## Keywords

information-seeking behavior, task type, eye-movements analysis, user studies, Web search

so on. Understanding the information-seeking behavior of users is critical for service providers and search engine developers. How users search for information on the Web depends on their particular purposes, so it is necessary to consider task characteristics when studying information-seeking behavior on the Web. We chose to capture exploratory search processes which reflect naturally occurring information needs [12].

Prior studies that investigated the relations between characteristics of tasks and information-seeking behavior have used various tasks, such as known-item, subject, fact-finding, and information gathering. These tasks considered to be classified into "navigational" or "informational" categories in Broder's taxonomy [2]. However, Broder [2] and Keller, Watters, and Shepard [7] suggest that transactional information must occupy a stable fraction of the information needs in Web searches. Therefore, it could be useful to investigate users' Web searches in tasks to obtain transactional information. In this study, we compare the information-seeking behaviors between an informational task and a transactional task.

Moreover, we analyze information-seeking behaviors by using eye-tracking data as well as the protocol and log data. Eye-tracking data has been used in only a few prior studies. We think it is necessary to refine the methodology by further accumulating research findings. The prior studies using eye-tracking data analyzed the interaction of user and search engine. In contrast, we investigate not only the interaction of user and search engine, but also the interaction of user and the entire Web space encompassing the search for individual Web pages presented in the search results. To achieve such a goal, however, it is necessary to clarify the interaction of user and search engine and the interaction of user and individual Web pages. Therefore, in this study, we first analyze the influence of task characteristics on information-seeking behavior on the Web based on log data. Moreover, as a preliminary analysis of eye-tracking data, we analyze eye-tracking during the browsing of search results pages.

In the rest of the paper, Chapter 2 describes related works

<http://dx.doi.org/10.1145/1414694.1414728>

# 例7 高久雅生, 谷藤幹子: “材料系研究所における機関リポジトリ NIMS eSciDoc の開発から応用まで: 研究者総覧 SAMURAI と研究ライブラリコレクション”. 情報管理, 2012, Vol.55, No.1, p.29-41

J-STAGE  
情報管理 Journal of Information Science and Management  
科学技術振興機構 Japan Science and Technology Agency

ONLINE ISSN: 1347-1597 PRINT ISSN: 0022-2988 2012年5月02日現在 収録数: 5,621記事

閲覧する 論文投稿する 発行機関について 最新巻号 J-STAGE トップ > 資料トップ > 書誌事項

情報管理 Vol. 55 (2012) No. 1 P 29-41  
DOI: http://dx.doi.org/10.1241/johokanri.55.29 JST J-STAGE/johokanri/55.29

材料系研究所における機関リポジトリ NIMS eSciDoc の開発から応用まで 研究者総覧 SAMURAI と研究ライブラリコレクション  
高久 雅生<sup>1)</sup>, 谷藤 幹子<sup>1)</sup>  
1) 独立行政法人物質・材料研究機構 企画部門 科学情報室  
公開日 2012/04/01  
キーワード: 機関リポジトリ, 研究者総覧, eSciDoc, SAMURAI, 研究ワークフロー

全文HTML 本文PDF [3368K]

抄録 引用文献(35) 被引用文献(2)

物質・材料研究機構 (NIMS) では、2008年よりデジタルライブラリー構想に基づく機関リポジトリ NIMS eSciDoc の開発と運用を始めた。eSciDocは柔軟な拡張可能性と豊富な Web API を併せ持つドイツ製のオープンソースのリポジトリソフトウェアであり、単に文献リポジトリにとどまらず、eサイエンスのための汎用ツールとしての機能を持ち合わせている。このような利点を活かして開発、運用してきた機関リポジトリ NIMS eSciDoc の現状と課題を報告する。あわせて、機関リポジトリと対をなして取り組んでいる研究者総覧 SAMURAI についても報告する。SAMURAI は、NIMS 研究者約 500 人を対象に、その連絡先や業績文献、研究内容などをわかりやすく伝えるサービスとして、機関リポジトリや外部データベースと密に連携しながら、2010 年より運用を開始した。本報告では、これらのサービス内容と利用動向とともに、今後の展開について述べる。

記事言語: Japanese 前の記事 | 次の記事

記事ツール  
お気に入り登録  
被引用アラート  
認証解除アラート  
追加情報アラート  
URLコピー  
著者にメール  
書誌事項をダウンロード  
RIS  
BibTeX  
問い合わせ  
この記事を共有  
ツイッター フェイスブック メールで知らせる

材料系研究所における機関リポジトリ NIMS eSciDoc の開発から応用まで 研究者総覧 SAMURAI と研究ライブラリコレクション  
Developing institutional repository at National Institute for Materials Science  
Researchers directory service "SAMURAI" and Research Collection Library  
高久 雅生<sup>1)</sup> | 谷藤 幹子<sup>1)</sup>  
TAKAKU Masao<sup>1)</sup>; TANIFUJI Mikiko<sup>1)</sup>  
1) 独立行政法人物質・材料研究機構 企画部門 科学情報室 (〒305-0047 茨城県つくば市千現1-2-1) Tel:029-859-2813  
E-mail: TAKAKU.Masao@nims.go.jp  
1 Scientific Information Office, Planning Division, National Institute for Materials Science  
(1-2-1 Sengen Tsukuba-shi, Ibaraki 305-0047)  
原稿受理 (2012-02-13)  
情報管理 55(1), 029-041, doi:10.1241/johokanri.55.29 (http://dx.doi.org/10.1241/johokanri.55.29)  
著者抄録  
物質・材料研究機構 (NIMS) では、2008 年よりデジタルライブラリー構想に基づく機関リポジトリ NIMS eSciDoc の開発と運用を始めた。eSciDoc は柔軟な拡張可能性と豊富な Web API を併せ持つドイツ製のオープンソースのリポジトリソフトウェアであり、単に文献リポジトリにとどまらず、e サイエンスのための汎用ツールとしての機能を持ち合わせている。このような利点を活かして開発、運用してきた機関リポジトリ NIMS eSciDoc の現状と課題を報告する。あわせて、機関リポジトリと対をなして取り組んでいる研究者総覧 SAMURAI についても報告する。SAMURAI は、NIMS 研究者約 500 人を対象に、その連絡先や業績文献、研究内容などをわかりやすく伝えるサービスとして、機関リポジトリや外部データベースと密に連携しながら、2010 年より運用を開始した。本報告では、これらのサービス内容と利用動向とともに、今後の展開について述べる。  
キーワード  
機関リポジトリ, 研究者総覧, eSciDoc, SAMURAI, 研究ワークフロー  
における機関リポジトリと情報ベース化と社会公開の歴史は古い。例を挙げると、機関リポジトリ NIMS eSciDoc<sup>1)</sup>、前身の NIMS 研究データベース<sup>2)</sup>、放射線医学総合研究所の発表論文等データベース<sup>3)</sup>、宇宙航空研究開発機構の JAXA リポジトリ  
<http://dx.doi.org/10.1241/johokanri.55.29>  
Copyright © Japan Science and Technology Agency 2012  
▲ ページトップへ  
研究者という学術機関における研究成果のデータベース化と社会公開の歴史は古い。例を挙げると、機関リポジトリ NIMS eSciDoc<sup>1)</sup>、前身の NIMS 研究データベース<sup>2)</sup>、放射線医学総合研究所の発表論文等データベース<sup>3)</sup>、宇宙航空研究開発機構の JAXA リポジトリ  
情報管理 vol.55 no.1 2012 29

## 例8

薬袋秀樹：“図書館協議会の可能性－草の根からの図書館振興”. 社会教育. 2012, No.792, p.20-25

**Tulips<sup>R</sup>**  
University of  
Tsukuba Library

つくばリポジトリ

検索  検索

詳細検索

④ ホーム

ブラウズ

④ コミュニティ & コレクション

④ 発行日

④ 著者

④ タイトル

④ 主題

つくばリポジトリ (Tulips-R) >  
0 コンテンツタイプ別 (Content type) >  
01 雑誌発表論文等 (Journal article, etc.) >  
社会教育 (Social education) >

このアイテムの引用には次の識別子を使用してください: <http://hdl.handle.net/2207/1000>

タイトル: 図書館協議会の可能性: 草の根の視点

著者: 菊袋 秀樹  
[Minai, Hideki](#)  
ミナイ, ヒデキ

発行日: 6月 - 2012

出版者: 全日本社会教育連合会

誌名: 社会教育

号: 792

開始ページ: 20

終了ページ: 25

URI: <http://hdl.handle.net/2207/1000>

テキストバージョン: publisher

出現コレクション: [菊袋 秀樹 \(Minai Hideki\)](#)  
[社会教育 \(Social education\)](#)

このアイテムのファイル:

ファイル	記述	サイズ	フォーマット
<a href="#">社会教育_792.pdf</a>		672.23 kB	Adobe PDF

<http://hdl.handle.net/2241/117267>

はじめに

近年、公立図書館における図書館協議会の活動に関するニュースを聞くことが多い。以前と比べて、図書館協議会の活動は活発化していると言われるが、課題も多い。

図書館協議会の設置の現状については、三年ごとに、文部科学省の「社会教育調査」で県別、地方公共団体の種類別に設置図書館数の調査が行われている(注1)。

全国の図書館協議会の詳しい実態については、一九八五年に日本図書館協会(注2)、二〇一二年に平山陽菜・池内淳(筑波大学図書館情報メディア系)が調査を行っている(注3)。公立図書館の特定事項に関する調査を取り上げられる場合もある。

図書館協議会の現状、可能性、課題について、関係文献と筆者の協議会委員の経験をもとに論じてみたい。

(一) 現行図書館法(二〇一一年改正)  
図書館協議会設置の趣旨  
第一四一一六条で定められている。その任務は「図書館の運営に関し館長の諮詢問に応ずるとともに、図書館の行うべき事務につき、館長に対して意見を述べる」ことである(一四条)。  
図書館法制定時の解説書では、「住民の具体的な図書館に対する要望なり意見なりを、図書館委員会を実施する責任者とも言うべき館長に対して反映せしめるために置かれる」と書かれている(註)。  
図書館協議会の設置は任意で(一四条)、協議会の委員は、当該図書館を設置する地方公共団体の教育委員会が任命する(一五条)。協議会の設置、委員の定数、任期等必要な事項は地方公共団体

（二）――年に、地域の自主性及び自立性を高めるための改革の推進を図るための関係法律の整備に関する法律（平成二年法律第105号）によつて、図書館法が改正され、「委員の任命の基準については、文部科学省令で定める基準を參照するものとする」が付け加えられた（六条）。

併せて、図書館法施行規則（文部科学省令）が改正され、參照すべき基準として、「学校教育及び社会教育の関係者、家庭教育の向上に資する活動を行う者並びに学識経験のある者の中から任命することとする」と定められた（二二条）。

（二）図書館法の改正（二〇〇八年）

二〇〇八年の図書館法が改章するにあたっては、第三

# ARTICLE

## 図書館協議会の可能性

筑波大学図書館情報学系教授  
著者秀樹

# 例9 柳玄姫, 葉袋秀樹: “韓国の図書館法と社会的背景”. 図書館情報メディア研究, 2013, Vol.10, No.2, p.1-17.

**Tulips®**  
University of Tsukuba Library  
つくばリポジトリ

検索

詳細検索

ホーム

ブラウズ

- ④ コミュニティ & コレクション
- ④ 発行日
- ④ 著者
- ④ タイトル
- ④ 主題

つばりポジトリ (Tulips-R) >  
○コンテンツタイプ別 (Content type) >  
05 紀要論文 (Departmental bulletin paper) >  
図書館情報メディア研究 >

このアイテムの引用には次の識別子を使用してください: <http://hdl.handle.net/2241/118746>

タイトル: 〈原著論文〉韓国の図書館法と社会的背景  
その他のタイトル: 〈Papers〉The Korean Library Laws and their social background  
著者: 柳, 玄姫  
葉袋, 秀樹  
[Ryu, Hyeonsook](#)  
[Minai, Hideki](#)  
リュウ, ヒョンスク  
ミナイ, ヒデキ  
[筑波大学図書館情報メディア研究科](#)  
[筑波大学図書館情報メディア系](#)  
[Graduate School of Library, Information and Media Studies, University of Tsukuba](#)  
[Faculty of Library, Information and Media Science, University of Tsukuba](#)

発行日: 29-3月-2013  
出版者: 「図書館情報メディア研究」編集委員会  
誌名: 図書館情報メディア研究  
巻: 10  
号: 2  
開始ページ: 1  
終了ページ: 17  
URI: <http://hdl.handle.net/2241/118746>  
フルテキスト: <http://www.tulips.tsukuba.ac.jp/mylimedio/dl/page.do?issueid=1135979&tocid=100102294&>

<http://hdl.handle.net/2241/118746>

このアイテムのファイル:

このアイテムに関連するファイルはありません。

図書館情報メディア研究第10卷2号1~17ページ

2012年

## 韓国の図書館法と社会的背景

柳 玄姫\*, 葉袋 秀樹\*\*

The Korean Library Laws and their social background

Hyeonsook RYU, Hideki MINAI

### 抄録

韓国では、1963年に図書館法が制定され、1987年に改正されたが、1991年に廃止されて、代わりに図書館振興法が、1994年に図書館・読書振興法が制定され、2006年には、再び図書館法が制定された。1987年の改正図書館法を含めて、5つの法律が存在してきたが、これらの基本的内容が共通するものであることをから、これらの法律を5つの図書館法ととらえることができる。本研究の目的は、それらがどのような社会的背景のもとで、制定・改正されてきたのかを考察することである。

本研究では、まず、韓国の図書館法の制定や改正の社会的背景について、政治・経済、教育・文化の4つの面から考察した。次に、法律の改正や制定の変遷の内容と特徴を考察し、法律の時代区分を試み、3期に分け、各期の法律の特徴を分析した。そして、図書館法の変遷と社会背景の関係について考察した。

その結果、韓国の図書館法は、政治の民主化、経済の発達、公教育の普及、文化政策の確立などを背景として制定・改正されてきたことが明らかになった。

### Abstract

In Korea, the Library Law was adopted in 1963. It has been amended in 1987, but was abolished in 1991. Instead, Library Promotion Law has been enacted as Library and Reading Promotion Law in 1994. In 2006, Library Law has again been adopted. Including the Amended Library Law of 1987, a total of 5 laws have existed. Owing to the mutual contents of these, it is possible to treat them as 5 different Library Laws. The object of this paper is to examine the social background, bringing about their enactments and amendments.

Firstly, this paper examines the social background of Korean Library Laws adoption and amendments from four perspectives: political, economic, educational and cultural. Secondly, it considers the contents and characteristics of amendments and enactment changes. Furthermore, this paper attempts a periodization of the 5 laws, dividing them into 3 terms and analysing their characteristics in every period. Then, the paper examines the relationship between the social background and the changes of the Library Laws.

The main conclusions drawn from this paper argue that democratisation of politics, economic development, prevalence of public education and the policy on culture constitute the background for the adoption and amendments of Korean Library Laws.

\* 筑波大学大学院図書館情報メディア研究科研究生  
Graduate School of Library, Information and Media Studies, University of Tsukuba  
\*\* 筑波大学図書館情報メディア系  
Faculty of Library, Information and Media Science, University of Tsukuba

# 例10 大向一輝: “CiNii Articlesのシステムデザインとデータモデル”. 情報の科学と技術, 2012, Vol.62, No.11, p.473-477.

CiNii 日本の論文をさがす 大学図書館の本をさがす 新規登録 ログイン English

論文検索 著者検索 全文検索 (beta) 大学図書館の本をさがす >

▼詳細検索

論文名  
著者名  
ISSN  
出版者

著者所属  
卷  
参考文献

刊行物名  
号  
出版年 年から 年まで

ページ

○すべて ○CiNiiに本文あり ○CiNiiに本文あり、または連携サービスへのリンクあり

検索

**CiNii Articlesのシステムデザインとデータモデル(<特集>データベース構築の今)**  
**System design and data modeling of CiNii articles(<Special feature>Database design and construction)**

太向一輝  
OHMUKAI Ikkai

国立情報学研究所  
National Institute of Informatics

この論文を読む/探す

CiNii PDF CiNii 論文PDF - 定額アクセス可能  
CiNii Books CiNii Books - 大学図書館でさがす

抄録

学術情報サービスにおいて、コスト面での制約がある中で大量のアクセスを高速に処理するためには、サービスが備えるべき機能を精査し、その機能の実現に適したシステム設計を行う必要がある。CiNii Articlesでは月間3500万～5000万のアクセスに対応するため、機能要件を検索と書誌表示に限定し、高速な検索エンジンと単純な処理のみを行うRDBを組み合わせることで性能要件を達成した。また、書誌IDを維持・管理するシステムを構築することで信頼性の高い情報サービスの提供を行っている。

In order to process large amount of access at low cost, it is necessary to design the system considering important functions to be provided. In CiNii Articles, we have achieved the performance requirements by using search engine and simple DBMS. We also provide reliable information service by the bibliographic ID management system.

収録刊行物

<http://ci.nii.ac.jp/naid/110009544496>

NII-ELS

書き出し

RefWorks書き出し  
EndNote書き出し

Information Science and Technology Association

特集 : データベース構築の今

## CiNii Articles のシステムデザインとデータモデル

大向 一輝\*

学術情報サービスにおいて、コスト面での制約がある中で大量のアクセスを高速に処理するためには、サービスが備えるべき機能を精査し、その機能の実現に適したシステム設計を行う必要がある。CiNii Articles では月間3500万～5000万のアクセスに対応するため、機能要件を検索と書誌表示に限定し、高速な検索エンジンと単純な処理のみを行うRDBを組み合わせることで性能要件を達成した。また、書誌IDを維持・管理するシステムを構築することで信頼性の高い情報サービスの提供を行っている。

キーワード : システムデザイン, 検索エンジン, データベース管理システム, ユニーク ID, 名寄せ

### 1. はじめに

人々の情報収集の手段として、ウェブは極めて重要な存在となった。学術情報流通の分野においても、ウェブを通じた情報提供サービスの比重が高まっており、研究者や学生にとってなくてはならないインフラとして認識されつつある。実際、サービスへのアクセス数や論文ファイルのダウンロード数は毎年最高値を記録しており、その勢いが衰える気配はない。その意味で、サービス提供者の責任は日増しに大きくなっている。一方、国内では学術情報サービスは公的機関が多く役割を担っているが、昨今の財政状況の中では、需要が伸びていると言えども、それに応じた予算を投入することが難しくなっている。

このように、限られた資源の中で、増え続けるアクセスにどう対応していくかはサービス提供者が抱える共通の課題である。この課題に応えるためには、サービスの要件を明確にするだけでなく、各種のソフトウェアあるいはハードウェアの特徴や制約を知り、それらを生かしたシステムを設計しなければならない。とくに、学術情報サービスは大規模なデータを取り扱うことから、そのデータの構造や管理手法について熟知しておく必要がある。

本稿では、筆者らが開発・運用を行っている学術情報サービスである「CiNii（サイニイ）」を取り上げ、その設計方針と実装について述べる。2012年9月の時点では、CiNii のサービスには国内の論文情報を提供する「CiNii Articles」と、大学図書館の図書・雑誌の情報を提供するという2種類のシステムが併存している。CiNii Articlesについて述べる。なお、CiNii でも CiNii Articles と同様の設計方針を踏襲

### 2. データベースとしての CiNii Articles

#### 2.1 CiNii Articles の概要

CiNii Articles は国立情報学研究所（NII）が2005年から運営している国内最大規模の学術論文検索・提供サービスである。学術会や大学との連携によって NII が電子化を行った約400万件の論文情報を加え、国立国会図書館の雑誌記事索引、科学技術振興機構の J-Stage、各大学・研究機関が運営する機関リポジトリなどの論文情報を合わせて約1,500万件の論文・記事が検索対象となっている。NII が電子化した論文は CiNii Articles 上で本文ファイルを提供し、外部のサービスに存在している論文についてはリンクを表示する。また、NII が構築している引用文献索引データベースの情報を用いて、論文の引用・被引用関係を表示することができる。

#### 2.2 機能要件

ユーザー側から見た CiNii Articles の主な機能としては、入力されたキーワードに合致する論文の一覧を表示する検索機能と、個々の論文の書誌を表示する書誌表示機能、本文ファイルの提供機能、刊行物・巻・号ごとに論文の一覧を表示するディレクトリ機能などがある。いずれの機能も、ユーザーの操作によってシステム内のデータが書き換えられることはなく、ユーザーはまっさらな情報を受け取るのみである。一方、CiNii Articles で扱うデータは、後述のパックエンドシステムで作成され、毎週1回の頻度でパッチ処理によって更新が行われる。

このように、ユーザーに対しては表示のみを行い、パッチ処理によってデータを更新するという構成は情報システムとして特段珍しいものではない。また、論文の書誌データはあらかじめ構造化されているため、リレーショナルデータベース（RDB）になじみやすい。単純化すれば、論文情報を提供するサービスは、書誌データを RDB に一括登録し、検索や書誌表示といった要求に応じて問い合わせを行い、その結果を表示するようなシステムであればよい。その意味では CiNii Articles は典型的なデータベースシステム

	雑誌名	出版社	プラットフォーム	提供フォーマット	電子化の主体	機能
例1	Science	AAAS	AAAS	PDF, HTML, +α	出版社	◎
例2	Nature	Nature	Nature	PDF, HTML, +α	出版社	◎
例3	JASIS&T	ASIS&T	Wiley	PDF, HTML, +α	Wiley	◎
例4	IPM	Elsevier	Elsevier	PDF, HTML, +α	出版社	◎
例5	IR	Springer	Springer	PDF, HTML, +α	出版社	◎
例6	IIiX	ACM	ACM	PDF, HTML, +α	出版社	◎
例7	情報管理	JST	J-STAGE	PDF, HTML, +α	JST	◎
例8	社会教育	全日本社会教育連合会	つくばリポジトリ	PDF(スキャン)	つくばリポジトリ	△
例9	図書館情報 メディア研究科	筑波大学	つくばリポジトリ	PDF(スキャン)	つくばリポジトリ	△
例10	情報の科学と技術	情報科学技術協会	CiNii	PDF(スキャン)	CiNii	△

# 学術分野におけるデジタルドキュメント

## —雑誌論文以外に流通する文献—

- Eジャーナル
- Eブック
  - 専門書
  - 事典
  - レファレンスブック
  - データシート
- 技術報告(テクニカルレポート; Technical report)
- 会議論文集(プロシードィングス; Proceedings)
- 口頭発表資料
  - スライド
  - ポスター
- 教材

# Eブックの例

John Domingue, Dieter Fensel, James A. Hendler, Eds. "Handbook of Semantic Web Technologies". Springer, 2011, 1056p.  
(eISBN:978-3-540-92913-0)



Search



Home • Contact Us



Search within this reference work



2011

## Handbook of Semantic Web Technologies

Editors: John Domingue, Dieter Fensel, James A. Hendler

ISBN: 978-3-540-92912-3 (Print) 978-3-540-92913-0 (Online)

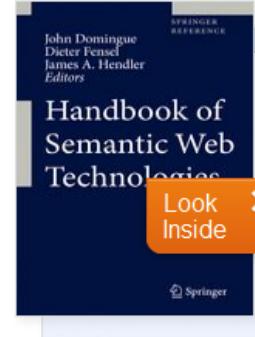
### Table of contents (22 entries)

◀ Page 1 of 2 ▶

#### Front Matter

» Download PDF (136KB)

Pages i-xx



#### Share



#### Other actions

» About this Reference Work

<http://dx.doi.org/10.1007/978-3-540-92913-0>

#### Foundations and Technologies

Reference Work Entry

##### Introduction to the Semantic Web Technologies

John Domingue, Dieter Fensel, James A. Hendler

» Download PDF (788KB)

» View Reference Work Entry

Pages 1-41

Reference Work Entry

##### Semantic Web Architecture

Andreas Harth, Maciej Janik, Steffen Staab

» Download PDF (696KB)

» View Reference Work Entry

Pages 43-75

# Eブックの例(事典)

“Encyclopedia of Library and Information Sciences”, Third Edition. Taylor & Francis, 2012.  
(eISBN:978-0-8493-9711-0)

Taylor & Francis Online Access provided by Tsukuba University

Register | Sign in | Mobile | Librarians | Authors & Editors | Societies

Home | Browse | Products | Reference | Journals | eBooks | Tandfonline.com

The online platform for Taylor & Francis Group content

Search Advanced Search

Within current reference work  Entire site

Home > Reference Works > Encyclopedia of Library and Information Sciences, Third Edition

Browse Reference Work

By Title

By Contributor

By Date

Description

Editor Biographies

Subscribe

Encyclopedia of Library and Information Sciences, Third Edition

DOI: 10.1081/E-ELIS3  
ISBN: 978-0-8493-9712-7; eISBN: 978-0-8493-9711-0  
Last Updated: 10 Aug 2012

Subject: Information Science, Librarianship,  
Publisher: Taylor & Francis

Related Articles

Most read  Most cited

There are no chapters to display.

Add to shortlist | Recommend to: A friend | A librarian

By Title  By Contributor  By Date

Choose an action ▾

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Academic Librarianship  
Barbara B. Moran, Elisabeth Leonard  
Published: 9 Dec 2009 | Abstract | References | Open URL

Academic Libraries  
Susan Carol Curran, Jessie Quiñónez-Skrivner

View full text  
Download full text  
Full access

<http://www.tandfonline.com/doi/book/10.1081/E-ELIS3>

# 会議論文集の例

IEEE computer society

Japan Consortium - Tsukuba University

Home Digital Library Home Search Resources Institutional and Library Resources Newsletter About the Digital Library

> I > ICDE > 2012 > 2012 IEEE 28th International Conference on Data Engineering

> Abstract - Viewing the Web as a Distributed Knowledge Base

Quick Search This Publication Digital Library Advanced Search

2012 IEEE 28th International Conference on Data Engineering  
Arlington, Virginia USA  
April 01-April 05  
ISBN: 978-0-7695-4747-3

Table of Contents

Papers

- [Front cover] (PDF)  
pp. C1  
 PDF  

- [Title page i] (PDF)  
pp. i  
 PDF  

- [Title page iii] (PDF)  
pp. iii  
 PDF  

- Copyright Page (PDF)  
pp. iv  
 PDF  

- Table of contents (PDF)  
pp. v-xii  
 PDF  

- Message from the Program Chairs and the General Chair (PDF)  
pp. xvii-xviii  
 PDF  

- Conference Organization (PDF)  
pp. viii-xviii  
 PDF  
  
  


ONLINE M.S. IN DATA ANALYTICS  
LEARN MORE

IEEE computer society

Japan Consortium - Tsukuba University

Home Digital Library Home Search Resources Institutional and Library Resources Newsletter About the Digital Library

> I > ICDE > 2012 > 2012 IEEE 28th International Conference on Data Engineering

> Abstract - Viewing the Web as a Distributed Knowledge Base

Quick Search This Publication Digital Library Advanced Search

2012 IEEE 28th International Conference on Data Engineering  
Arlington, Virginia USA  
April 01-April 05  
ISBN: 978-0-7695-4747-3

Viewing the Web as a Distributed Knowledge Base (PDF)

Serge Abiteboul  
Émilien Antoine  
Julia Stoyanovich

DOI Bookmark: <http://doi.ieeecomputersociety.org/10.1109/ICDE.2012.150>

ABSTRACT

This paper addresses the challenges faced by everyday Web users, who interact with inherently heterogeneous and distributed information. Managing such data is currently beyond the skills of casual users. We describe ongoing work that has as its goal the development of foundations for declarative distributed data management. In this approach, we see the Web as a knowledge base consisting of distributed logical facts and rules. Our objective is to enable automated reasoning over this knowledge base, ultimately improving the quality of service and of data. For this, we use Webdamlog, a Datalog-style language with rule delegation. We outline ongoing efforts on the Web dam Exchange platform that combines Webdamlog evaluation with communication and security protocols.

ADDITIONAL INFO

Citation:  
Serge Abiteboul, Émilien Antoine  
International Conference on Data Engineering

Usage of this product signifies acceptance of the terms and conditions

<http://www.computer.org/cSDL/proceedings/icde/2012/4747/00/index.html>

“Proceedings of the 28<sup>th</sup> International Conference on Data Engineering”, 2012, Arlington, Virginia USA, IEEE. (ISBN: 978-0-7695-4747-3)

IEEE computer society

Japan Consortium - Tsukuba University

Home Digital Library Home Search Resources Institutional and Library Resources Newsletter About the Digital Library

> I > ICDE > 2012 > 2012 IEEE 28th International Conference on Data Engineering

> Abstract - Viewing the Web as a Distributed Knowledge Base

Quick Search This Publication Digital Library Advanced Search

2012 IEEE 28th International Conference on Data Engineering  
Arlington, Virginia USA  
April 01-April 05  
ISBN: 978-0-7695-4747-3

Viewing the Web as a Distributed Knowledge Base (PDF)

Serge Abiteboul  
Émilien Antoine  
Julia Stoyanovich

DOI Bookmark: <http://doi.ieeecomputersociety.org/10.1109/ICDE.2012.150>

ABSTRACT

This paper addresses the challenges faced by everyday Web users, who interact with inherently heterogeneous and distributed information. Managing such data is currently beyond the skills of casual users. We describe ongoing work that has as its goal the development of foundations for declarative distributed data management. In this approach, we see the Web as a knowledge base consisting of distributed logical facts and rules. Our objective is to enable automated reasoning over this knowledge base, ultimately improving the quality of service and of data. For this, we use Webdamlog, a Datalog-style language with rule delegation. We outline ongoing efforts on the Web dam Exchange platform that combines Webdamlog evaluation with communication and security protocols.

ADDITIONAL INFO

Citation:  
Serge Abiteboul, Émilien Antoine  
International Conference on Data Engineering

Usage of this product signifies acceptance of the terms and conditions

<http://www.computer.org/cSDL/proceedings/icde/2012/4747/00/4747a001-abs.html>

Cybersecurity.

# テクニカルレポートの例

## NII Technical Reports

National Institute of Informatics  
2-1-2 Hitotsubashi, Chiyoda, Tokyo, 101-8430, Japan  
Print ISSN: 1346-5597  
Online ISSN: 1884-0795

[Japanese Here](#)

[2001](#) [2002](#) [2003](#) [2004](#) [2005](#) [2006](#) [2007](#) [2008](#) [2009](#) [2010](#) [2011](#) [2012](#) [2013](#)

[About NII Technical Reports](#)  
[Authors' Page \(Staff Only\)](#)

2013

No.	Authors & Title	Published
<a href="#">NII-2013-001E</a>	Kugamoorthy GAJANANAN <i>A Short Tutorial of the Scenario Markup Language</i>	Feb 8, 2013

2012

No.	Authors & Title	Published
<a href="#">NII-2012-005E</a>	Keiichi Morikuni and Ken Hayami <i>Convergence of inner-iteration GMRES methods for least squares problems</i>	Aug 3, 2012
<a href="#">NII-2012-004E</a>	Michael E. Houle and Michael Nett <i>Rank-Based Similarity Search: Reducing the Dimensional Dependence</i>	Jun 18, 2012
<a href="#">NII-2012-003E</a>	Makoto Kanazawa <i>Almost Affine Lambda Terms</i>	May 25, 2012
<a href="#">NII-2012-002E</a>	Makoto Tatsuta <i>Formalization of Lemma for Adjacent Replacement Paths</i>	Feb 17, 2012
<a href="#">NII-2012-001E</a>	Keiichi Morikuni, Lothar Reichel, Ken Hayami <i>FGMRES for linear discrete ill-posed problems</i>	Jan 12, 2012

2011

No.	Authors & Title	Published
<a href="#">NII-2011-002E</a>	Yasunori Aoki, Ken Hayami, Hans De Sterck and Akihiko Konagaya <i>Cluster Newton Method for Sampling Multiple Solutions of an Underdetermined Inverse Problem: Parameter Identification for Pharmacokinetics</i>	Aug 22, 2011
<a href="#">NII-2011-001E</a>	Keiichi Morikuni and Ken Hayami <i>Inner-Iteration Krylov Subspace Methods for Least Squares Problems</i>	Apr 04, 2011

2010

No.	Authors & Title	Published
<a href="#">NII-2010-003E</a>	Makoto Tatsuta <i>Translation of Multi-Staged Language</i>	Jul 20, 2010
<a href="#">NII-2010-002E</a>	Liang Xu, Hideaki Takeda, Masahiro Hamasaki, and Huayu Wu <i>Typing Software Articles with Wikipedia Category Structure</i>	Jul 12, 2010
<a href="#">NII-2010-001E</a>	Shaoying Liu and Shin Nakajima <i>A Decompositional Approach to Automatic Test Case Generation Based on Formal Specifications</i>	Jan 12, 2010

<http://www.nii.ac.jp/TechReports/>

## Stanford InfoLab Publication Server



[Home](#) | [Browse by Year](#) | [Browse by Project](#) | [Browse by Author](#) | [Statistics](#) | [Advanced Search](#) | [Help](#)  
[Login](#) | [Create Account](#)

### The PageRank Citation Ranking: Bringing Order to the Web.

Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) *The PageRank Citation Ranking: Bringing Order to the Web*. Technical Report. Stanford InfoLab.

[BibTeX](#) [DublinCore](#) [EndNote](#) [HTML](#)



PDF  
299Kb

#### Abstract

The importance of a Web page is an inherently subjective matter, which depends on the readers interests, knowledge and attitudes. But there is still much that can be said objectively about the relative importance of Web pages. This paper describes PageRank, a method for rating Web pages objectively and mechanically, effectively measuring the human interest and attention devoted to them. We compare PageRank to an idealized random Web surfer. We show how to efficiently compute PageRank for large numbers of pages. And, we show how to apply PageRank to search and to user navigation.

**Item Type:** Techreport (Technical Report)

**Additional Information:** Previous number = SIDL-WP-1999-0120

**Subjects:** Computer Science > Digital Libraries

**Projects:** [Digital Libraries](#)

**Related URLs:** Project Homepage

<http://www-diglib.stanford.edu/diglib/pub/>

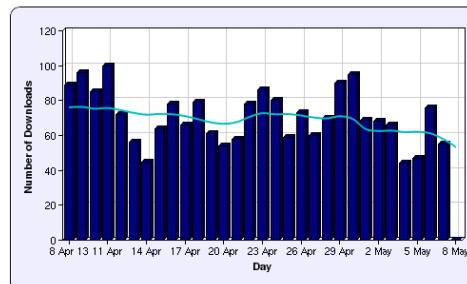
**ID Code:** 422

**Deposited By:** Import Account

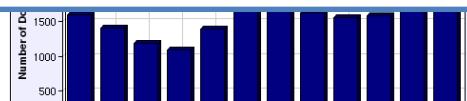
**Deposited On:** 30 Oct 2001 16:00

**Last Modified:** 28 Dec 2008 09:56

#### Download statistics



<http://ilpubs.stanford.edu:8090/422/>



# さて、ここで改めて…

学術分野のドキュメントの場合、デジタルドキュメントならでは特徴とは、一体なんだろうか？

# オンラインジャーナルの動向

- オンラインのみの雑誌(E-only)
  - 例1: Dlib Magazine( <http://www.dlib.org/> )
    - 1995年創刊
  - 例2: PLoS One( <http://www.plosone.org/> )
    - 2006年創刊
  - 例3: PeerJ ( <https://peerj.com/> )
    - 2013年創刊
- 紙メディアからの離脱
  - △ Issue
  - △ Volume
  - ? ページ
  - ○ Article number

# 論文の要素、構造

- (書誌情報)
    - タイトル
    - 著者
    - 抄録(要旨)
    - キーワード
  - 本文
    - 章
      - 節
      - 段落
        - » 文
  - 図
  - 表
  - 参考文献(リスト)
  - 付録資料  
(Supplementary material)
    - 図表
    - ビデオ
    - 音声
    - ...
- デジタルドキュメントならではの要素が多く出てくる
- Science誌での例から:  
<http://dx.doi.org/10.1126/science.1207758>

# オンラインジャーナルの動向 －オープンサイエンス(Open Science)－

- 背景
  - オープンアクセス
  - オープンデータ
- インフォーマルコミュニケーションの補完
- ウェブ上でのコメント、査読機能
- (査読とはどうあるべきか？)
- (科学的であるとはどういうものか？)



## Exposed proteins of the *Schistosoma japonicum* tegument

Jason Mulvenna<sup>1,2</sup>\*, Luke Moerdyk<sup>3</sup>, Malcolm K. Jones<sup>4,5</sup>, Supavee Nawaratna<sup>6</sup>, Erica M. Lovis<sup>1</sup>, Geoffrey N. Cobert<sup>6</sup>, Michelle Colgrave<sup>6</sup>, Alan Jones<sup>7</sup>, Alex Loukas<sup>8</sup>, Donald P. McManus<sup>9</sup>

\*Corresponding author. Tel.: +61 7 3845 2726; fax: +61 7 3845 2997.

<sup>1</sup>Molecular Biology Laboratory, Division of Infectious Diseases, Queensland Institute of Medical Research, Qld 4029, Australia

<sup>2</sup>Molecular Parasitology Laboratory, Division of Infectious Diseases, Queensland Institute of Medical Research, Qld 4029, Australia

<sup>3</sup>Parasite Cell Biology Laboratory, Division of Infectious Diseases, Queensland Institute of Medical Research, Qld 4029, Australia

<sup>4</sup>The University of Queensland, School of Veterinary Sciences, Qld 4072, Australia

<sup>5</sup>CSIRO Livestock Industries, Brisbanes, Qld 4037, Australia

<sup>6</sup>The University of Queensland, Institute for Molecular Biosciences, Qld 4072, Australia

<sup>7</sup>These authors contributed equally to the manuscript.

### Research highlights

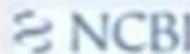
- Proteins exposed on the surface of parasitic worms are an important source of novel drug and vaccine targets.
- These proteins are the most accessible to the host and likely to possess functions important for the survival of the worm.
- Biotinylation (labelling of lysine side-chain residues with biotin) of whole worms is a useful technique for separating exposed proteins from other protein constituents of the tegument.
- Using this technique in combination with LC-MS/MS we identified 54 proteins as putatively host-exposed in *Schistosoma japonicum*.
- Using confocal and electron microscopy, the immunolocalisation of these proteins was observed.

### Subcellular content

... proteins

5 of 54 (100%)

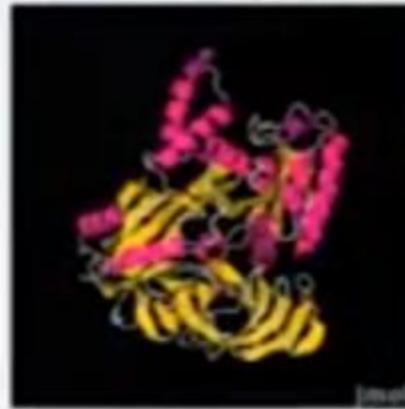
All 54 (100%)



elongation factor 1-Schistosoma japonicum]

348 aa protein

Connexin-26-interacting protein (NCBP1)



1 msd

Download movie (100%)

### Subcellular Location

Cytosol/Plasma

Accessories

AAK57918

Version:

AAK57918.1 (GJ14258828)

...

<http://www.articleofthefuture.com/>

# 一つの論文原稿の複数バージョン

## - 版と種類 -

- 作成過程、執筆過程による差異
  - 投稿原稿
  - 著者最終稿：改訂原稿
  - (プレプリント)
  - (早期公開版)
  - 出版社刊行版
  - (再利用版：機関リポジトリなど)
- 電子化による差異
  - 紙→スキャン
  - ボーンデジタル

# オンラインジャーナルのアクセス管理

- 購読者に閲覧を許す仕組み
  - IPアドレス単位による組織単位の認証
  - ユーザ・パスワード方式による個人(組織)認証
- アクセス提供、閲覧利用をライセンスする



As a service to the community, AAAS/Science has made this

User Login

User Name

Password

Remember me

[Forgot your user/password?](#)

**SUBMIT**

[Register for Free](#)   [Join/Subsribe](#)

I want to purchase this article

Price: ¥3300

In order to purchase this article you must be a registered user.

**Register now**

I want to buy this article via ReadCube

Rent: \$3.99\*  
Purchase: \$6.99\*  
\*Printing and sharing restrictions apply

**Purchase now**

I want to subscribe to *Nature*

Price: US\$199

This includes a free subscription to *Nature News* together with *Nature Journal*.

**Subscribe now**

Personal subscribers to *Nature* can view articles published from 1997 to the

access Online Article

**JP-PACS: A massively parallel processor at Tsukuba** Original Research Article

Parallel Computing, Volume 25, Issues 13–14, December 1999, Pages 1635–1656  
Isamu Nakazawa, Hiroshi Nakamura, Taisuke Boku, Ikuo Nakata, Yoshiyuki Y...

options for accessing this content:  
If you have access to this content through a society membership, please first log in via other institutional login options <http://onlinelibrary.wiley.com/login-options>. If you would like institutional access to this content, please recommend the title to your institution.  
If you already have a Wiley Online Library or Wiley InterScience user account, you can purchase online access to this Article for a 24-hour period (price varies by article).  
New Users: Please register, then proceed to purchase the article.

Login via your institution  
Athens / Institutional login  
Remote access  
Click here to activate  
registered Users please login:  
Access your saved publications, articles and searches  
Manage your email alerts, orders and subscriptions  
Change your contact information, including your password

e-mail:   
password:

# オンラインジャーナルのアクセス管理

## - 著作権管理 -

- オープンになるか？
- 著作権の移譲処理
  - 著作財産権が出版社に移管されることが通常（ライセンスアグリーメント）
  - 販売する権利
  - → 電子化する権利
  - → 配信する権利
- 著者の権利
  - 機関リポジトリ

# オンラインジャーナルの保存/保管

- 大半のオンラインジャーナルは学術機関の図書館購読の契約に依存している
- アーカイブ権(いったん購読すれば、その契約期間分のアクセス閲覧は保証される)
- ただし、出版社の手違い、プラットフォームの閉鎖や倒産等により、データが失われることを避けるための仕組み
  - CROCKSS
  - 分散ミラーリング(サイトまるごとをコピーしたものを作成しておく)
  - 平時にはコピーするだけ

# まとめ

- ・出版社サイト等での事例を元に、オンラインジャーナルプラットフォームがもつ機能、提供形態、フォーマット、その周辺について考えてみた
  - PDF, HTML, XML, etc.
  - ドキュメントフォーマット
  - 制作の過程
  - 著作権
  - 最近の動向
- ・次回は、さらに詳しくドキュメントフォーマットについてさらに考えてみたいと思います。

# 出席票提出

- 最後に、出席票に日付・氏名・所属、感想・質問等(あれば)を記入の上、忘れず提出してください。

提出位置：

3編生

2013xxxxx,

2012xxxxx

3年次

(知識)

2011xxxxx

3年次

(創成)

2011xxxxx

4年次

2010xxxxx,

etc.