CiNii: Bringing Linked Data to Japan's Largest Scholarly Search Engine

Ikki Ohmukai
National Institute of Informatics,
Japan
i2k@nii.ac.jp

Hideaki Takeda
National Institute of Informatics,
Japan
takeda@nii.ac.jp

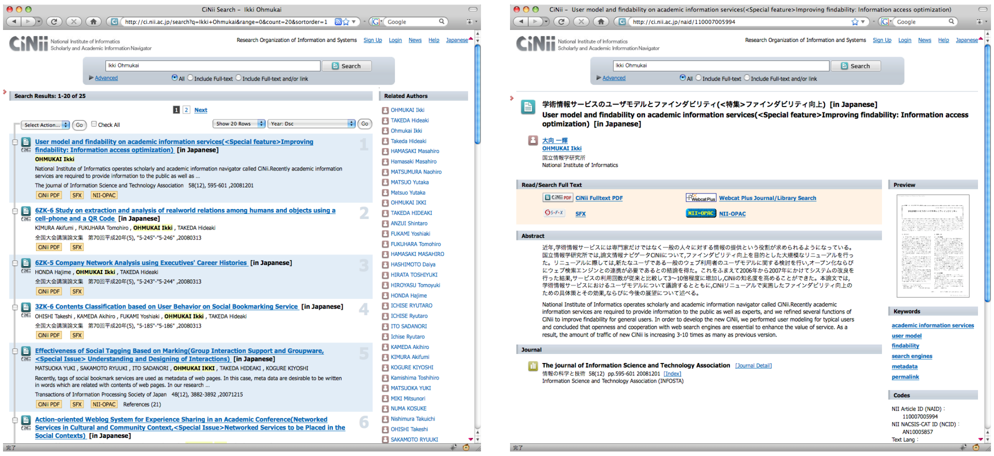
Keywords: metadata; linked data; academic information services; OpenSearch; RDF.

Abstract

National Institute of Informatics operates "CiNii" (http://ci.nii.ac.jp/), the largest scholarly search engine in Japan. CiNii is a database of journals and proceedings. It stores full text and bibliography of over 3 million articles. In cooperation with National Diet Library, electronic journal publishers and institutional repositories, CiNii also collects bibliographic metadata of 20 million articles. The number of page views was over 10 million in December 2008 and still increasing. Figure1 shows screenshots of CiNii.

CiNii provides search functions for scholarly articles. Search result shows a list of articles corresponding to the query. CiNii also publishes "Bibliography Permalink", which is a web page describing the bibliography of every article stored in the database. It shows not only title and authors, but also the abstract and a list of references and citations.

To become widely accepted by advanced users and developers, we are continuing to enhance CiNii, e.g., introducing permalinks and being the target of major search engines. In April 2009, CiNii redesigned and relaunched, and Linked Data (Berners-Lee, 2006) is offered for public use of scholarly information. In this paper we describe details of our Linked Data.


CiNii Search – Ikki Ohmukai
http://ci.nii.ac.jp/search?q=Ikki+Ohmukai&range=0&count=20&sortorder=1
Google
CiNii National Institute of Informatics Scholarly and Academic Information Navigator
Research Organization of Information and Systems
Sign Up
Login
News
Help
Japanese
Ikki Ohmukai
Search
Advanced
All
Include Full-text
Include Full-text and/or link
1
2
Next
Select Action...
Go
Check All
Show 20 Rows
Year: Dsc
Go
CiNii PDF
SFX
NII-OPAC
完了

Search Results: 1-20 of 25

1 User model and findability on academic information services(<Special feature>Improving findability: Information access optimization) [in Japanese]
OHMUKAI Ikki
National Institute of Informatics operates scholarly and academic information navigator called CiNii.Recently academic information services are required to provide information to the public as well as ...
The journal of Information Science and Technology Association 58(12), 595-601 ,20081201

2 6ZK-6 Study on extraction and analysis of realworld relations among humans and objects using a cell-phone and a QR Code [in Japanese]
KIMURA Akifumi , FUKUHARA Tomohiro , OHMUKAI Ikki , TAKEDA Hideaki
全国大会講演論文集 第70回平成20年(5), "5-245"-"5-246" ,20080313

3 6ZK-5 Company Network Analysis using Executives' Career Histories [in Japanese]
HONDA Hajime , OHMUKAI Ikki , TAKEDA Hideaki
全国大会講演論文集 第70回平成20年(5), "5-243"-"5-244" ,20080313

4 3ZK-6 Contents Classification based on User Behavior on Social Bookmarking Service [in Japanese]
OHISHI Takeshi , KAMEDA Akihiro , FUKAMI Yoshiaki , OHMUKAI Ikki , TAKEDA Hideaki
全国大会講演論文集 第70回平成20年(5), "5-185"-"5-186" ,20080313

5 Effectiveness of Social Tagging Based on Marking(Group Interaction Support and Groupware, <Special Issue> Understanding and Designing of Interactions) [in Japanese]
MATSUOKA YUKI , SAKAMOTO RYUUKI , ITO SADANORI , OHMUKAI IKKI , TAKEDA HIDEAKI , KOGURE KIYOSHI
Recently, tags of social bookmark services are used as metadata of web pages. In this case, meta data are desirable to be written in words which are related with contents of web pages. In our research ...
Transactions of Information Processing Society of Japan 48(12), 3882-3892 ,20071215
References (21)

6 Action-oriented Weblog System for Experience Sharing in an Academic Conference(Networked Services in Cultural and Community Context,<Special Issue>Networked Services to be Placed in the Social Contexts) [in Japanese]

Related Authors
OHMUKAI Ikki
TAKEDA Hideaki
Ohmukai Ikki
Takeda Hideaki
HAMASAKI Masahiro
Hamasaki Masahiro
MATSUMURA Naohiro
MATSUO Yutaka
Matsuo Yutaka
OHMUKAI IKKI
TAKEDA HIDEAKI
ANZUI Shintaro
FUKAMI Yoshiaki
FUKUHARA Tomohiro
HAMASAKI MASAHIRO
HASHIMOTO Daiya
HIRATA TOSHIYUKI
HIROYASU Tomoyuki
HONDA Hajime
ICHISE RYUTARO
ICHISE Ryutaro
ITO SADANORI
Ichise Ryutaro
KAMEDA Akihiro
KIMURA Akifumi
KOGURE KIYOSHI
Kamishima Toshihiro
MATSUOKA YUKI
MIKI Mitsunori
NUMA KOSUKE
Nishimura Takuichi
OHISHI Takeshi
SAKAMOTO RYUUKI

CiNii – User model and findability on academic information services(<Special feature>Improving findability: Information access optimization)
http://ci.nii.ac.jp/naid/110007005994
Google
CiNii National Institute of Informatics Scholarly and Academic Information Navigator
Research Organization of Information and Systems
Sign Up
Login
News
Help
Japanese
Ikki Ohmukai
Search
Advanced
All
Include Full-text
Include Full-text and/or link
Read/Search Full Text
CiNii PDF
CiNii Fulltext PDF
Webcat Plus
Webcat Plus Journal/Library Search
S·F·X
SFX
NII-OPAC
NII-OPAC
Preview
完了

学術情報サービスのユーザモデルとファインダビリティ(<特集>ファインダビリティ向上) [in Japanese]
User model and findability on academic information services(<Special feature>Improving findability: Information access optimization) [in Japanese]

大向 一輝
OHMUKAI Ikki
国立情報学研究所
National Institute of Informatics

Abstract

近年,学術情報サービスには専門家だけではなく一般の人々に対する情報の提供という役割が求められるようになっている。国立情報学研究所では,論文情報ナビゲータCiNiiについて,ファインダビリティ向上を目的とした大規模なリニューアルを行った。リニューアルに際しては,新たなユーザである一般のウェブ利用者のユーザモデルに関する検討を行い,オープン化ならびにウェブ検索エンジンとの連携が必要であるとの結論を得た。これをふまえて2006年から2007年にかけてシステムの改良を行った結果,サービスの利用回数が従来と比較して3〜10倍程度に増加し,CiNiiの知名度を高めることができた。本論文では,学術情報サービスにおけるユーザモデルについて議論するとともに,CiNiiリニューアルで実施したファインダビリティ向上のための具体策とその効果,ならびに今後の展望について述べる。

National Institute of Informatics operates scholarly and academic information navigator called CiNii.Recently academic information services are required to provide information to the public as well as experts, and we refined several functions of CiNii to improve findability for general users. In order to develop the new CiNii, we performed user modeling for typical users and concluded that openness and cooperation with web search engines are essential to enhance the value of service. As a result, the amount of traffic of new CiNii is increasing 3-10 times as many as previous version.

Journal

The journal of Information Science and Technology Association [Journal Detail]
情報の科学と技術 58(12) pp.595-601 20081201 [Index]
Information Science and Technology Association (INFOSTA)

Keywords
academic information services
user model
findability
search engines
metadata
permalink

Codes
NII Article ID (NAID) :
110007005994
NII NACSIS-CAT ID (NCID) :
AN10005857
Text Lang :

FIG. 1. Screenshots of CiNii.

Summary of Linked Data in CiNii is shown in Figure 2.

Search function in the new CiNii is compatible with the OpenSearch specification. Response formats of OpenSearch consist of XHTML, RSS 1.0 and Atom 1.0. To improve interoperability with other Web services, CiNii does not use original modules, and only uses standard vocabularies such as Dublin Core and PRISM.

In OpenSearch, RSS has a link to bibliography RDF by `rdfs:seeAlso` so that software can obtain machine-readable bibliography without parsing XHTML. Bibliographic metadata is

https://doi.org/10.23106/dcmi.952109685

described with RDF. Like OpenSearch responses, Bibliography RDF does not contain original vocabulary.

There are several problems with representing bibliography by RDF. CiNii cannot know exact URI of full text PDF, which corresponds to the subject of an RDF triple, because the download site might be a different place for different users. So we consider a virtual URI which is combination of bibliography permalink and `#article` fragment. There is also no reference to URI of RDF itself, but we resolve this problem by introducing `foaf:isPrimaryTopicOf` to indicate RDF URI.

Author information is expressed by not only `dc:creator` but also FOAF terms. In FOAF specification, it is impossible to describe that "Person A belongs to Organization B". We use PIM vocabulary (W3C, 2006) to illustrate that statement. There are still some problems. For example, information for more than one person with the same name would be mixed because authors in CiNii do not have unique URI. In the case of affiliation, we define organization name as a resource tentatively because it is not likely for multiple organizations with same name to exist.

We embedded several microformats in XHTML files for quick hacking. xFolk is used in search results, and hCard/hAtom are used in bibliography permalinks.

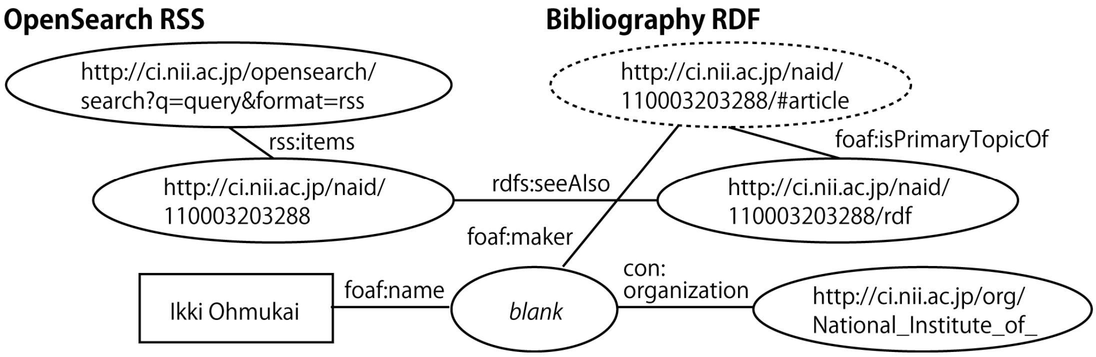

OpenSearch RSS
Bibliography RDF
http://ci.nii.ac.jp/opensearch/search?q=query&format=rss
http://ci.nii.ac.jp/naid/110003203288/#article
rss:items
foaf:isPrimaryTopicOf
http://ci.nii.ac.jp/naid/110003203288
rdfs:seeAlso
http://ci.nii.ac.jp/naid/110003203288/rdf
foaf:maker
con:organization
Ikki Ohmukai
foaf:name
blank
http://ci.nii.ac.jp/org/National_Institute_of_

FIG. 2. Model of Linked Data in CiNii.

In this poster we present our efforts to create a large repository of Linked Data from Japan's largest scholarly search engine. We are going to operate the new CiNii and obtain feedback from the users and third-party developers. Trust in our data is still low because of a lack of unique URIs about authors and organizations. We will collaborate with information sources to resolve this problem.

References

Berners-Lee, Tim. (2006). Linked Data - Design Issues. Retrieved May 7, 2009, from http://www.w3.org/DesignIssues/LinkedData.html.

W3C. (2006). SWAP Personal Information Markup. Retrieved May 7, 2009, from http://www.w3.org/2000/10/swap/pim/contact#.

https://doi.org/10.23106/dcmi.952109685