Numerical Representation of Web Sites of Remote Sensing Satellite Data Providers and Its Application to Knowledge Based Information Retrievals with Natural Language

Kohei Arai [1]
Graduate School of Science and Engineering
Saga University
Saga City, Japan

***Abstract*—A method for numerical expression of web site which is relating to satellite remote sensing and its application to knowledge based information retrieval system which allows retrievals with natural language is proposed and implemented. Through experiments with remote sensing related information, it is found that the proposed information retrieval system does work in particular for remote sensing satellite data retrievals with natural language.**

***Keywords*—Information retrieval; Knowledge based system; Natural language; Feature mapping**

I. Introduction

When a word or words are typed in search engines, a list of web sites that contain those words is displayed. The words you enter are known as a query [1]. Baeza-Yates and Ribeiro-Neto linked Information Retrieval to the user information needs which can be expressed as a query submitted to a search engine [2]. Search engines were also known as some of the brightest stars in the Internet investing frenzy that occurred in the late 1990s [3]. Although search engines are programmed to rank websites based on their popularity and relevancy, empirical studies indicate various political, economic, and social biases in the information they provide [4],[5].

There are great amount of information for remote sensing satellite data retrievals. Directory, inventory, catalog, and guide information are available in a worldwide basis. Smart search engine, therefore, is needed for remote sensing satellite data retrievals. In order to realize a smart search engine, knowledge base system has to be involved. Knowledge base system consists of knowledge base which includes object and attribute, and inference engine. Also, users would like to search remote sensing satellite data with natural language.

The next section describes the proposed knowledge based search engine which allows search appropriate URLs of remote sensing satellite data providers with natural language followed by some experimental results. Then conclusion is described together with some discussions.

II. Proposed Information Retrieval System

A. Knowledge Based System and Conventional Information Retrieval Systems

Figure 1 shows configuration of knowledge based system which consists of Inference Engine: IE, Knowledge Base database: KB, Knowledge Base Management System: KBMS and Knowledge Acquisition Module: KAM. When user submits query to the knowledge based system, previously acquired knowledge about remote sensing satellite data providers is used to output search results. Also Figure 2 shows the query system which allows distribution of multiple queries to the multiple databases which include different types of remote sensing satellite data through expanded query generator from a single query. Therefore, appropriate queries are submitted from expanded query generator by database system by database system.

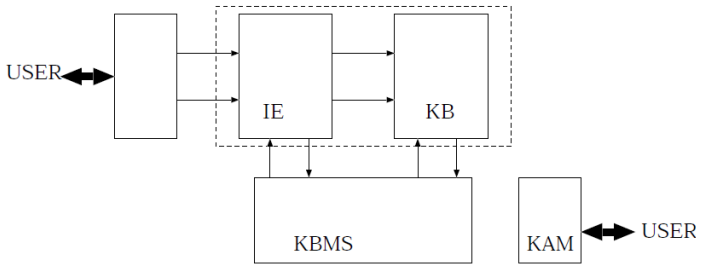

Fig. 1. Fundamental configuration of knowledge based system

There are distributed remote sensing satellite database systems created and managed by the data providers. Figure 3 shows assisted search module which allows distributed database servers search through internet. Only thing users have to do is to access the assisted search module. Then the module makes a search for appropriate database server from the distributed servers.

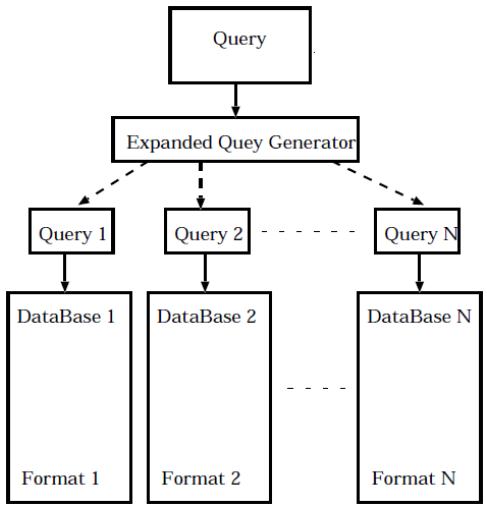

Fig. 2. Expanded query generator

There is NOAA: National Oceanic and Atmospheric Administration, EOSDIS: Earth Observation Satellite System of Data Information System, USGS: United States Geological Survey, DOE: Department of Energy, etc. as database servers of data providers. This assisted search module is the fundamental function of GCDIS-ASK: Global Change Data and Information System of Assisted Search for Knowledge.

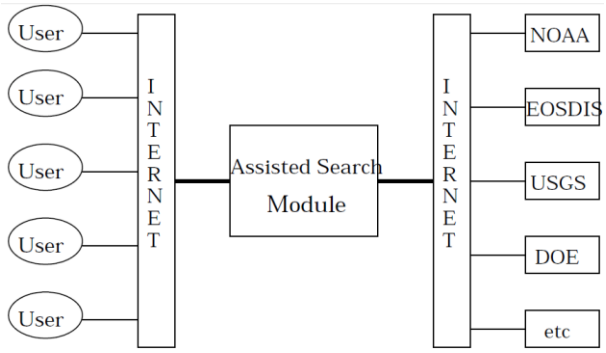

Fig. 3. Assisted search module

There are three basic components for GCDIS-ASK, client module, assisted search module and data collection module. Figure 4 shows system architecture of client module.

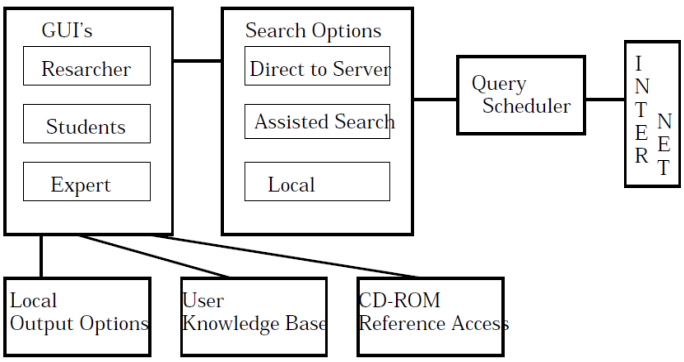

Fig. 4. Client module of GCDIS-ASK

When query is submitted from users, there are three options, direct search for the database, assisted search, and local search for the database under the GUI: Graphical User Interface.

One of the key features of assisted search module is Natural Language: NL search engine. Search can be done with a combination of statistical search and concept base search. The former is based on statistical variables, frequency of the query words, distance between query words, etc. On the other hands, the later uses concepts derived from expertise persons. Thus users can create concepts by using previously acquired knowledge and expertise in the knowledge base in order to improve search performance. Extendable knowledge base system makes such data and information search available. Under the extendable knowledge base system, there is NL search engine which consists of dictionary.

In order for that, smart query server and smart query scheduler are prepared as shown in Figure 5. There is specific database server under each smart query server. Search query scheduler monitors each smart query server. When user submits a query, client handler makes query generation and send query to appropriate smart query servers as shown in Figure 6.

Figure 7 shows more detailed architecture of assisted search module. Key issues here are NL search engine and user profiling. Users may use natural language in their queries. Users' profiles are archived and used for choosing information access options. Therefore, every time user access to the database, user profile is updated and thus information and data search can be done in much efficient way.

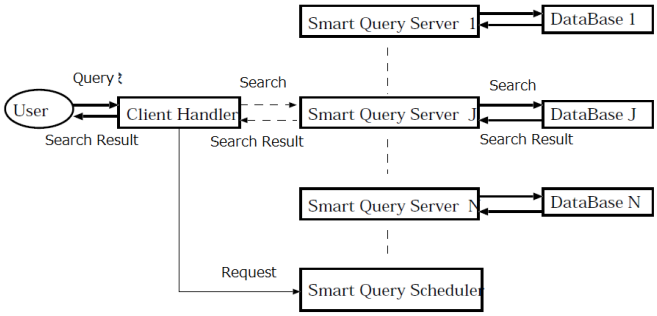

Fig. 6. Smart query scheduler and smart query servers

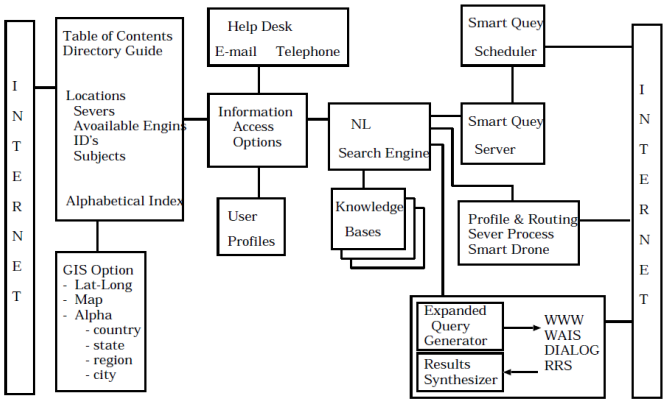

Fig. 7. Detailed architecture of assisted search module

Architecture of Data Collection Module: DCM is shown in Figure 8. Database search methods are different by data providers. User, however, can retrieve data and information from the different database servers in a unified way through this DCM.

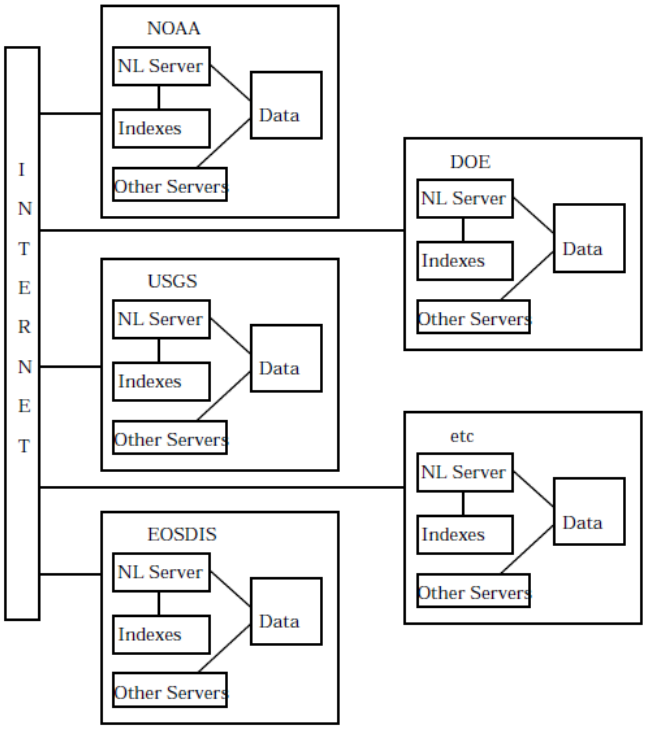

Fig. 8. Data Collection Module

B. *Proposed Information Retrieval Systems*

There are attribute information about data provider, keywords about URL, URL itself, data provider name, available data period (from when to when), information about data provider. There are attribute information about data, atmosphere, hydrosphere, cryosphere, geosphere, biosphere. Under the attribute information, there are many sensor names as shown in Table 1.

In the attribute information about data, there are observation target names, satellite names, sensor names, etc. as shown in Table 2. This is the example of NSIDC: National Snow and Ice Data Center.

Then it becomes possible to plot all URL in the five dimensional (attributes) vector space as shown in Figure 9. In the figure, x, y, and z axis are hydrosphere, cryosphere, and geosphere, respectively. Then the distance between URLs can be defined as shown in Figure 10. Angle between URLs can be calculated easily. Thus the smallest angle between input search query and the existing URL can be found followed by sending the closest URL to users as search result.

These attribute information can be classified as shown in Table 3. In the Table 3, number denotes the number of attribute and can be normalized as shown in the bottom row of the Table 3.

TABLE II. A VARIETY OF ATTRIBUTE INFORMATION ABOUT DATA WHICH ARE PROVIDED BY NSIDC

Attribution
snow
vegetation
river
sea ice
ice sheet
elevation
water vapor
avalanche
glacier
permafost
RADARSAT
NOAA
AVHRR
SSM/I
LAI
snowstorm
SST

TABLE III. CLASSIFIED ATTRIBUTE INFORMATION AND THE NUMBER OF ATTRIBUTES

Feature	Atmosphere	Hydosphere	Cryosphere	Geosphere	Biosphere
Attribute	water vapor NOAA	river sea ice NOAA AVHRR SSM/I SST	snow sea ice ice sheet avalanche glacier permafrost RADARSAT snowstorm	elevation NOAA RADARSAT	vegetation LAI
Number	2	6	8	3	2
			↓ Normalization ⇓		
	(0.184900,	0.554700,	0.739600,	0.277350,	0.184900)

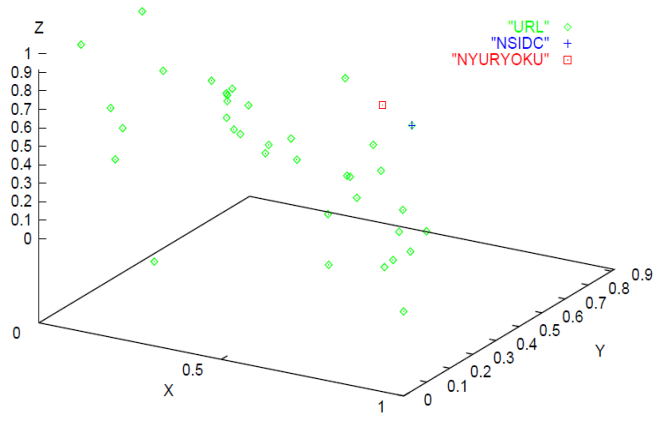

Fig. 9. URL distribution in the feature space of attributes (Hydrosphere, Cryosphere, and Geosphere)

TABLE I. SENSOR NAMES UNDER THE ATTRIBUTION INFORMATION

Atmosphere	Hydosphere	Cryosphere	Geosphere	Biosphere
volcanic smoke	sea wind	snow	wild fire	vegetation
volcanic ash	river	ice	forest fire	oilfire
cyclone	rain	snow covered area	LST	NPP
ozone	sea ice	glaial landforms	volcanic eruption	desertification
sea wind	sea	ice sheet	land	acid rain
cloud	ocean color	ENVISAT	earthquake	tree crown
hurricane	SST	X-SAR	elevatition	vegetation index
storm	El Nino	RADARSAT	liquefaction	tropical rain forest
water vapor	oil slick	sea ice	fire burn area	NDVI
air	chlorophyll	avalanche	carbonatite	vegetation cover rate
humidity	flood	glacier	crater temperature	vegetation community
aerosol	acid rain	permafrost	volcanic landforms	biomass
El Nino	ASTER	atomaspheric ice	igneous rock	LAI
ozone layer	AMI	MSR	drought	red edge
total ozone	ALMAZ	ADEOS-2	desertification	HIRS/2
global warming	EOS	ILAS-2	iron oxide	PRIRODA
fumaroic gas	AVHRR	MISR	SPOT	OKEAN
UV	PR	MOPITT	X-SAR	VCL
AMI	OCTS	RA	JERS-1	disaster
EOS	MOS-1	ATSR-M	sedimentary rock	environment
atomashere	GOES	ATSR-2	fault	resource of agiculture
albedo	SEASAT	MERIS	TM	
HCMM	SeaStar	AATSR	stratum	
GOME	DMSP	IRS-1C/1D	mud flow	
greenhouse effect	TOPEX/POSEIDON	WiFS	debris flow	
GOES	NIMBUS	PAN	lava	
GMS	CZCS	disaster	ophiolite	
DMSP	TRMM	snow cover	ASTER	
NINBUS	NOAA	NOAA	ETM+	

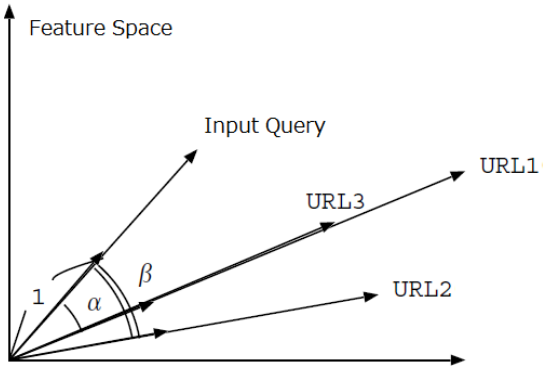

Fig. 10. Relation between input query and the existing URLs

Figure 11 shows architecture of the proposed remote sensing satellite data and information retrieval system.

Query from users is written in text format with natural language. Then angle between URLs can be calculated easily. Thus the smallest angle between input search query and the existing URL can be found followed by sending the closest URL to users as search result.

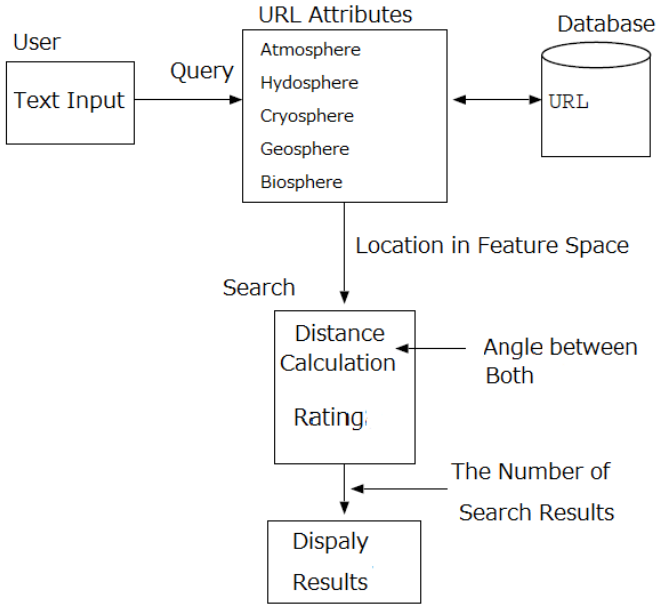

Fig. 11. Architecture of the proposed remote sensing satellite data and information retrieval system

III. Implementation and Experiments

A. Implementation

Using netscape environment, web design is performed with PHP. Top page of the proposed search system is shown in Figure 12.

B. Search Example

In the example of Figure 12, search request is done with the following natural language,

"I would like to get images of areas suffered from heavy snow. I would like to know situation of iceberg in the Antarctic Ocean using data from Polar 1km AVHRR dataset. I would like to know about icy content mapped from space with RADARSAT."

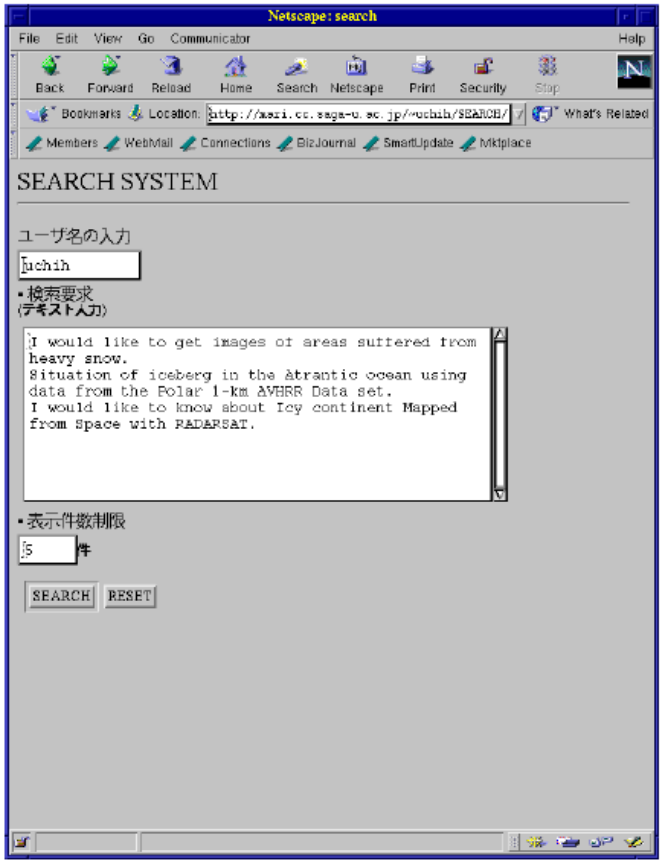

Fig. 12. Query input web page

When users submit the query together with users ID and the maximum number of search results, then the search result is returned as shown in Figure 13. For the example, the top five closest data providers to the query are output as search result with URL and the detailed information. These are aligned in accordance with the distance (angle) between query and the attribute information about data provider of URLs.

Users can refine the search results by reselecting much appropriate wording for query as shown in Figure 14. Then users can get much suitable URLs. Users' satisfaction is evaluated through questionnaire with the ten students and compares the evaluation result to the conventional keyword search. As the result, all students prefer the proposed natural language search rather than the conventional keyword search. Hit ratio is also evaluated with ten students and compare to the keyword search. It is found that approximately 10 points improvement is confirmed for the proposed search system in comparison to the conventional keyword search.

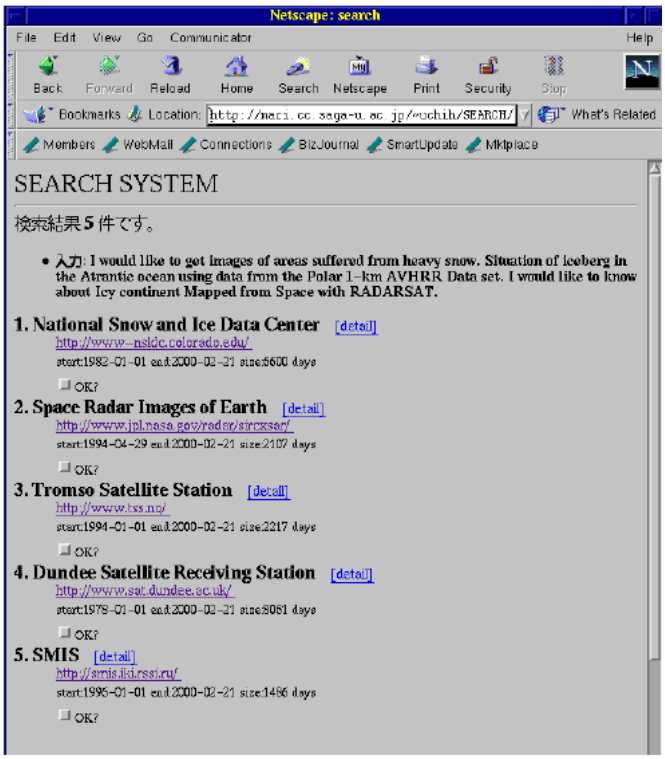

Fig. 13. Search result for the query of with the following natural language, "I would like to get images of areas suffered from heavy snow. I would like to know situation of iceberg in the Antarctic Ocean using data from Polar 1km AVHRR dataset. I would like to know about icy content mapped from space with RADARSAT."

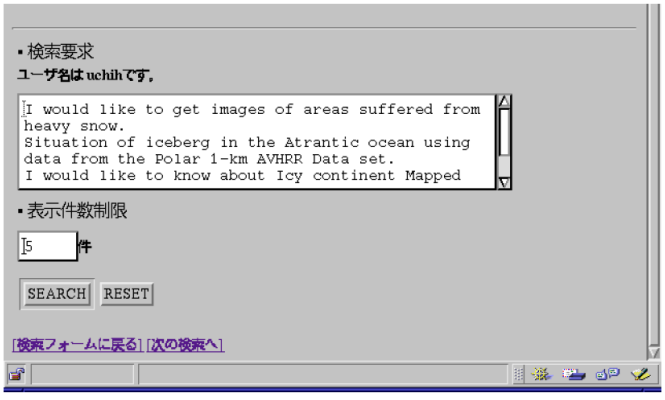

Fig. 14. Refinement of the search result by submit query with modified natural language again.

IV. Conclusion

A method for numerical expression of web site which is relating to satellite remote sensing and its application to knowledge based information retrieval system which allows retrievals with natural language is proposed and implemented.

Through experiments with remote sensing related information, it is found that the proposed information retrieval system does work in particular for remote sensing satellite data retrievals with natural language

Users' satisfaction is evaluated through questionnaire with the ten students and compares the evaluation result to the conventional keyword search. As the result, all students prefer the proposed natural language search rather than the conventional keyword search. Hit ratio is also evaluated with ten students and compare to the keyword search. It is found that approximately 10 points improvement is confirmed for the proposed search system in comparison to the conventional keyword search.

ACKNOWLEDGMENT

The author would like to thank Mr. Fumihiro Uchihashi for his effort to implement the proposed search engine and to conduct experiments with ten students.

REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. In WWW, 1998.

[2] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern Information Retrieval. Addison Wesley, May 1999.

[3] Gandal, Neil (2001). "The dynamics of competition in the internet search engine market". *International Journal of Industrial Organization* **19** (7): 1103–1117

[4] Segev, Elad (2010). Google and the Digital Divide: The Biases of Online Knowledge, Oxford: Chandos Publishing

[5] Vaughan, L. & Thelwall, M. (2004). Search engine coverage bias: evidence and possible causes, Information Processing & Management, 40(4), 693-707

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission "A" of ICSU/COSPAR since 2008. He wrote 30 books and published 322 journal papers