

VO service in Japan : Registry service based on Apache Solr and SIA v2 service for Japanese Facilities





Currently more than 20 thousands of VO services are registered in the VO registry database. Keyword Search is the most popular way to find a resource. There can be a lot of ways to implement this capability, and the performance depends on how to index the document describing the resource metadata.

Apache Solr is an open source search platform and uses the Lucene Java search library for full-text indexing. The document is indexed after the process that removes stop words (not adequate for a keyword e.g. a, the, and, is, are ...) and stems a word to a root word (e.g. clustering to cluster). Thus for the query of a keyword cluster, the Solr search system will return documents that contain a word cluster, clustering, or clustered. Solr also can handle single- or multi-token synonyms and abbreviation. For an example, SN, supernovae, and SNe are translated to supernova in the process of indexing. So it increases the probability for a given keyword to hit a desired resource metadata. In order to incorporate this feature, we upgraded the registry service behind the JVO portal to the one based on Solr.

The data from Nobeyama Legacy project were released from the JVO portal on 1st June in 2018. The data are now distributed also through the most recent VO standard interface called SIA-v2. The data of ALMA and Subaru telescope are also accessible through the SIA-v2. We are now working on distributing the data of Hyper Suprime-Cam SSP/DR1 and also the data of Hitomi satellite under the collaboration with C-SODA/JAXA.

http://jvo.nao.ac.jp/portal/ We present those our recent development on Japanese Virtual Observatory system.





1. Classification of VO resources

~ 20,000 resource metadata (RM) how to classify (grouping)? content metadata? some of them are poorly fulfilled.

auto-classification based on the full text in the RM.

Spectrum Super Nova Image AGN ALMA

2. How do we classify ?



define categories \rightarrow key words / phrases



- Solr-based RM registry is located behind the JVO portal.
- JVO portal provides a GUI for searching a resource.
- RM registry harvests RM from the publishing registries.

JVOIndex : Resource Metadata Indexing Service

236ABCDEFGHIKLMNOPQR<mark>S</mark>TUVWXYabcdefghijklmnopqrstuvwxyz

SAGE	Subaru Search									
	100	v records/page (tota	I = 211) Next	Skip to No	. 0 Skip					
SAINI galaxy survey	No.	Action	Bookmark	ID	Title	Туре	Activity	Reference URL	Access URL	Country
	0	Search Schema	-	More Info	Subaru Suprime-Cam image data Cutout service	Image	active	URL	URL	
SAO	1	Search Schema	-	More Info	Subaru HDS Spectrum data service	General	active	URL	URL	
	2	Search Schema	-	Me Info	Subaru Suprime-Cam data service	General	active	URL	URL	





extract key words and phrases from the RMs We used a Python implementation of the Rapid Automatic Keyword Extraction (RAKE) algorithm [1].

after long-long manual inspection ...



776 primary categories and 1,013 secondary categories are defined (as of 2018 Nov.).

The list of the categories needs to be updated regularly, but the manual intervention is a real bottleneck ... We need AI !

3. Database



	_					Subaru Suprime-Cam data Service		Ocheral	active	OIL	UNL	•							
CAV	3	Search	Schema	-	(m)	Subaru MOIRCS data service		General	active	URL	URL								
	4	4 Search Schema - M fo Subaru/XMM-Newton Deep Survey v1.0						General	active	URL	URL								
	5	Search	Schema	-	More Info	Proper motions with Subaru. II. (Richmond+, 2010)		General	active	URL	URL								
SCUBA	6	Search	Schema	-	More Info	PNe in the Virgo elliptical galaxy NGC 4649 (Teodorescu+, 201	1)	General	active	URL	URL								
	7	Search	Schema	-	More Info	Subaru H images of UM673 (Koptelova+, 2014)		General	active	URL	URL								
2202	8	8 Search Schema - More Info Ly{alpha} emitter in the SUBARU Deep Field (Taniguchi+, 2005)						General	active	URL	URL								
3033	9	Search	Schema	-	More Info	Subaru/XMM-Newton reep survey IV. (SXDS) (Ouchi+, 2008)	General	active	URL	URL		and the second							
SDSS Quasar ca	SDSS Quasar ca ¹⁰ Search Schema -					Subaru/XMM-Newton deep survey (SXDS). II. (Furusawa+, 200	8)	General	active	URL	URL								
SDSS candidate (11	Search	Schema	-	More Info	Log of Subaru Prime Focus Camera Exposures (ADAC NAOJ, 2	2007)	General	active	URL	URL								
	12	Search	Schema	-	More Info	Narrow absorption lines of lensed QSO J1029+2623 (Misawa+,	2016)	General	inactive	URL	URL								
SED	13	Search	Schema	-	More Info	Subpreciog: Subaru Prime Focus Camera (Suprime-Cam) ES	(posures Log (LEDAS)) Coord Search	Inactive	URL	URL	215							
JED	14 Search Schema - More Info Subaru weak-lensing survey. I. (Miyazaki, 2007)				Subaru weak-lensing survey. I. (Miyazaki 2007)		General	active	URL	URL									
	15	Search	Schema		More Into	Subaru new Calinier transmission function (ELP+, 2013)	06)	General	nactive	URL	URL								
SEGUE	17	Search	Schema	-	More Info	Ey{aipha} eninters in the SOBARO Deep Field (Shi hasaku+, 20	00)	General	active	URL	URL								
	12	17 Search Schema - More Info Subaru-ODGS In the Coma cluster (Fagi+, 2016)						General	active	URL	URL								
SFR	10	Search	Schema	-	More Info	$1_{\text{(alpha)}}$ emitters at 2–2.25 (maisuda+, 2011)		General	active	URL	URL								
	20	Search Schema - More Info LIV selected sources in the COODS-S field (Bust- 2012)						General	active	URI	LIRI								
SGR	21	Search	Schema	-	More Info	Ly(alpha) emitters at z=6.5 in Subaru Deep Field (Kashikawa+	2006	General	active	URI	URI	11 - L							
	22	Search	Schema	-	More Info	UBVRIZ HK' photometry of 2Ms CDEN X-ray sources (Barger+	2003)	General	active	URI	URI								
SMC	23	Search	Schema		More Info	Proper motions with Subaru SDF (Richmond+, 2009)	2000)	General	active	URL	URL	11 - L							
	24	Search	Schema	-	More Info	Blue stragglers in NGC 2419 (Dalessandro+, 2008)		General	active	URL	URL	ii -							
SMC planatany po	25	Search	Schema	-	More Info	Gatties of J0454-0309 lensing fossil group (Schirmer+, 2010)		General	active	URL	URL	11							
Sivic planetary ne	26	Search	Schema	-	More Info	Imaging of high-redshift Lyman alpha emitters (Yamada+, 2005)	General	active	URL	URL	ii -							
	27	Search	Schema	-	More Info	Fai t companions around YSOs in TMC (Itoh+, 2008)													
SN	1				3 0	mprero galaxy	Resource N	Aetadata_											
SN la							Created	2008-04-10T0-00	:00+09	Updated	2018-08-08T15	:46:32+09	Status	activ					
CNI hash selever					St/	ernberg supernova catalogue	Identifier	ivo://jvo/subaru/sp	ram										
SIV DOSE DAIAXV																			
SN nost galaxy							Short Name	SUBARU_SUP					Subaru Suprime-Cam dat service						
SN nost galaxy SN survey							Short Name Title	SUBARU_SUP Subaru Suprime-C	am dat servi	се									
SN host galaxy SN survey					Sila	baru	Short Name Title ServiceType	SUBARU_SUP Subaru Suprime-C General	am de servi	Ce									
SN nost galaxy SN survey					<mark>∎ Տա</mark> "Ու	baru Subaru Deep Field	Short Name Title Service Type AccessURL Reference URL	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp/ http://subarutelesc	am dat servi /skynode/do/ta ope.org/Obse	ce ap/spcam rving/Instrument	s/SCam/index.htr	nl							
SN host galaxy SN survey					<mark>والہ</mark>	Subaru Deep Field	Short Name Title ServiceType AccessURL ReferenceURL	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp/ http://subarutelesc	am da servi /skynode/do/ta ope.org/Obse	ce ap/spcam rving/Instrument	s/SCam/index.htr	nl							
SN host galaxy SN survey SNR SNR					الم ک ر الم	Subaru Deep Field	Short Name Title Service Type AccessURL ReferenceURL Country	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp. http://subarutelesc	am da servi /skynode/do/ta ope.org/Obse	ce ap/spcam rving/Instrument	s/SCam/index.htr	nl							
SN host galaxy SN survey SNR SNR					∎ Sw ch ∎ Su	Subaru Deep Field	Short Name Title Service Type AccessURL ReferenceURL Country Harvested	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp. http://subarutelesc	am dat servi /skynode/do/ta ope.org/Obse	ce ap/spcam rving/Instrument	s/SCam/index.htr	nl							
SN host galaxy SN survey SNR SNe SNe la					■ Sw ch ■ Su	Subaru Deep Field	Short Name Title Service Type AccessURL ReferenceURL Country Harvested from	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing	am dat servi /skynode/do/ta ope.org/Obse	ce ap/spcam rving/Instrument	s/SCam/index.htr	nl							
SN host galaxy SN survey SNR SNe SNe la type la SNe					= Sw ch = Su = Su	ibaru Subaru Deep Field nyaev nyaev-Zeldovich	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function large scale	am dat servi /skynode/do/ta ope.org/Obse registry S,galaxies: evo	ce ap/spcam rving/Instrument	s/SCam/index.htr formation,galaxie	nl s: luminosit	r function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe					= Sw Ch = Su = Su	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo	ce ap/spcam rving/Instrument blution,galaxies: universe	s/SCam/index.htr formation,galaxie	nl s: luminosit;	∕ function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO					= Sw Ch = Su = Su	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope,	am da servi /skynode/do/ta ope.org/Obse registry 8,galaxies: evo e structure of Suprime-Cam	ce ap/spcam rving/Instrument blution,galaxies: universe	s/SCam/index.htr formation,galaxie	nl s: luminosit	/ function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO					= Sw Ch = Su = Su	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument Subject	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope, GALAXIES,STARS	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo e structure of Suprime-Cam 6,galaxies: evo	ce ap/spcam rving/Instrument blution,galaxies: universe blution,galaxies:	s/SCam/index.htr formation,galaxie formation,galaxie	nl s: luminosit s: luminosit	/ function,mass / function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO					= Su Ch = Su = Su = Su	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster perCOSMOS	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument Subject Contributor	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope, GALAXIES,STARS function,large-scale	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo e structure of Suprime-Cam 6,galaxies: evo e structure of	ce ap/spcam rving/instrument blution,galaxies: universe blution,galaxies: universe	s/SCam/index.htr formation,galaxie formation,galaxie	nl s: luminosit; s: luminosit;	√ function,mass √ function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO SPIRE					■ Sw Ch ■ Su ■ Su ■ Su	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster perCOSMOS	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument Subject Contributor Creator	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope, GALAXIES,STARS function,large-scale	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo e structure of Suprime-Cam 6,galaxies: evo e structure of	ce ap/spcam rving/Instrument plution,galaxies: universe plution,galaxies: universe	s/SCam/index.htr formation,galaxie formation,galaxie	nl s: luminosit; s: luminosit;	γ function,mass γ function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO SPIRE					= Sw 2 5 5 5 5 5 5 5 5 5 5 5 5 5	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster perCOSMOS	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument Subject Contributor Creator Curation Date	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope, GALAXIES,STARS function,large-scale	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo e structure of Suprime-Cam 6,galaxies: evo e structure of	ce ap/spcam rving/Instrument plution,galaxies: universe plution,galaxies: universe	s/SCam/index.htr formation,galaxie formation,galaxie Curatio	nl s: luminosit; s: luminosit; n Version	γ function,mass γ function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO SPIRE SPT					= Sw = Su = Su = Su = Sw	baru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster perCOSMOS rift	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument Subject Contributor Creator Curation Date Publisher	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope, GALAXIES,STARS function,large-scale	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo e structure of Suprime-Cam 6,galaxies: evo e structure of	ce ap/spcam rving/Instrument plution,galaxies: universe plution,galaxies: universe	s/SCam/index.htr formation,galaxie formation,galaxie Curatio	nl s: luminosit; s: luminosit; n Version	γ function,mass γ function,mass						
SN host galaxy SN survey SNR SNe SNe la type la SNe SOHO SPIRE SPT					= Sw = Su = Su = Sw = Sw	subaru Subaru Deep Field nyaev nyaev-Zeldovich Sunyaev-Zeldovich Cluster perCOSMOS rift	Short Name Title ServiceType AccessURL ReferenceURL Country Harvested from Content Type Facility and Instrument Subject Contributor Creator Curation Date Publisher Right	SUBARU_SUP Subaru Suprime-C General http://jvo.nao.ac.jp http://subarutelesc ivo://jvo/publishing GALAXIES,STARS function,large-scale Subaru Telescope, GALAXIES,STARS function,large-scale	am da servi /skynode/do/ta ope.org/Obse registry 6,galaxies: evo e structure of Suprime-Cam 6,galaxies: evo e structure of	ce ap/spcam rving/Instrument plution,galaxies: universe	s/SCam/index.htr formation,galaxie formation,galaxie Curatio	nl s: luminosit s: luminosit n Version	y function,mass y function,mass						

Result

Using the key words & phrases extracted from the RMs, indexing service "JVO Index" was developed.

 JVO Index service provides an overview of the resources which is available in the VO

world.



RMs belonging to a given category are searched for with an backend database upon a request.

We selected Apache Solr for the database.

Apache Solr is an open source search platform.

Customizable text analyzer build from a sequence of tokenizers and filters.

An example of how the text "many Galaxies clustered around AGN" is indexed.

6. VO Services of Japanese Facilities

Major data resources published from the JVO and the adapted standards are summarized in the table below. JVO VO service toolkit [2] was used to build these services.

Title	SIA v1	SIA v2	SSA 1.1	CS	ΤΑΡ	Obs Core
ALMA VO Service		0			0	1.1

tokenizer	many	Galaxies	clustered	tered around AGN				Subaru Suprime-Cam data service	0 0		0	1.1
CORCINZET	↓ ↓	↓ ↓	V	↓ ↓	, tert		input text is split into tokens.	Subaru MOIRCS data service	0		0	
stop filter	X	Galaxies ↓	clustered	X	AGN	↓	stop words ^{*1} are removed.	Subaru HDS Spectrum data service		0	0	
synonym filter		Galaxies	clustered	active	galactic	nucleus	synonyms are replaced with a representative single- or multi- token.	Nobeyama Radio Telescope FITS archive	0		0	1.1
stem filter		galaxy	cluster	active	galactic	nucleus	tokens are replaced with a stem word of lowercase.	AKARI Far-infrared All-Sky Survey Maps	0		0	
		These wo	rds are used	as indexes		AKARI Point Source Catalog Public Release 1			0			
Query input is active galaction	s analy c nucle	yzed in tl ei" will m	he same v natch with	References [1] <u>https://github.com/anees</u> [2] Shirasaki, Y. et al. 2012, A	<mark>sha/RAKE</mark> SP Conf Vol 4	461, p4	51, ADASS XX	<				

^{*1} stop words are words which are the most common words and not useful for indexing. ^{*2} it is possible to use different analysis patterns for indexing of text and query.