

Building a smart portal for astronomy

Sébastien Derriere



Official definition of Portal

- 1 : door, entrance; especially : a grand or imposing one
- 2 : the whole architectural composition surrounding and including the doorways and porches of a church
- 3 : the approach or entrance to a bridge or tunnel
- 4 : a communicating part or area of an organism; specifically : the point at which something (as a pathogen) enters the body
- 5 : a site serving as a guide or point of entry to the World Wide Web and usually including a search engine or a collection of links to other sites arranged especially by topic

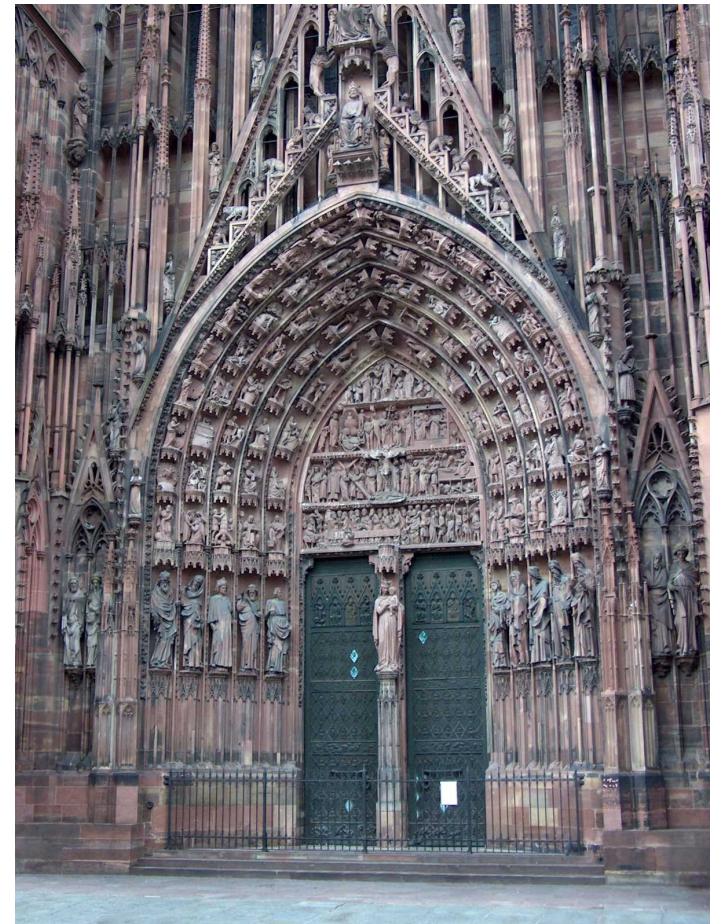
My definition of astronomy portal

- Web interface / application
- Data & Service discovery, query, analysis, workflows
- Customizable, modular (widgets), communicating
- Possible technologies : JS, Ajax, JSON, HTML5, VO protocols

My definition of astronomy portal

- Web interface / application
- Data & Service discovery, query, analysis, workflows
- Customizable, modular (widgets), communicating
- Possible technologies : JS, Ajax, JSON, HTML5, VO protocols

Simplicity vs complexity



VAO portal



Search All Virtual Observatory Collections:
[User Guide](#) | [Discovery Tool v1.3 \(4816\)...](#)

M101 Radius: 1 Arcmin Search
 Examples: [M101](#), [14 03 12.6 +54 20 56.7](#), [more...](#)

[Start Page](#)
[M101 r=1m](#)

296 Total Rows
[Export Table As...](#)
MESSIER 101 (RA: 14:03:12.545, Dec: +54:20:56.22), radius: 0.01667°

Filters
[Clear Filters](#)
[Edit Facets...](#)
[Help...](#)

Filter All Record Fields

Categories

- ☒ Catalog (255 of 255)
- ☒ Images (33 of 33)
- ☒ Spectra (3 of 3)

Waveband

- ☒ EUV (6 of 6)
- ☒ Gamma-ray (11 of 11)
- ☒ Infrared (71 of 71)
- ☒ Millimeter (6 of 6)
- ☒ Optical (161 of 161)
- ☒ Radio (29 of 29)
- ☒ UV (38 of 38)
- ☒ X-ray (52 of 52)

Publisher

- ☒ Canadian Astronomy Data Centre (3 of 3)
- ☒ CDS (156 of 156)
- ☒ Chandra X-ray Observatory (1 of 1)
- ☒ ESO (1 of 1)
- ☒ European Space Agency (1 of 1)

Short Name	Title
1 CADC	CADC Image Search
2 CADC/HST	CADC/HST Image Search
3 SwiftXRT	Swift XRT Instrument Log
4 XMMOMOBJ	XMM-Newton OM Objects
5 CADC/CFHT	CADC/CFHT Image Search
6 ADS	Astrophysics Data System
7 HLA [1]	Hubble Legacy Archive
8 Chan/XAssist	Chandra XAssist Source
9 Simbad	The SIMBAD astronomical
10 CIO	Catalog of Infrared Obser
11 SwiftUVOT	Swift UVOT Instrument Lo
12 SwiftBAT	Swift BAT Instrument Log
13 SuperCOSMOS [1]	SuperCOSMOS Science
14 SuperCOSMOS [2]	SuperCOSMOS Science
15 Chandra	Chandra X-Ray Observat
16 NED(sources)	The NASA/IPAC Extragala
17 B/cfht	Log of CFHT Exposures
18 MAST-Scrapbook	The MAST Image Scrapbo
19 ST-ECF/HLA/SIA	ST-ECF Hubble Legacy A
20 XMMOMSUSS	XMM-Newton Optical Mor
21 SkyView	SkyView Virtual Observat

AstroView

[RA] 14:04:17.023 [DEC] +54:12:25.717

[Bloquer...](#)

Simple input
Facets

ADS labs

ads labs NASA

ADS Labs Streamlined Search

Home Labs Home ADS Classic Help Sign on

intitle:"cloud" object:"Taurus" Astronomy SEARCH

Author First author Title Object Examples

Sort by

- ☒ Most recent
- ☐ Most relevant
- ☐ Most cited
- ☐ Most popular

Explore the field

- ☐ What people are reading
- ☐ What experts are citing
- ☐ Reviews and introductory papers

Return top 200 results.

myADS articles **Recently viewed articles**

Please login to enable personal content.

Categorization
Sorting
Login
Faceted search

ASCOT

[Viewport Sky](#) | [toggle](#)


[Name Resolve](#) | [toggle](#)

e.g. "M67" or "190 00 00 -11 37 21.2"
☒ RA: 132.8250°
DEC: 11.8000°

[Data Query](#) | [toggle](#)

Dataset Name:

☒ Returned 113 records.

Catalog:

Query String:

```
SELECT TOP 500 g.objid, g.ra, g.dec, s.z, g.modelMag_u, g.modelMag_g, g.modelMag_r,
g.modelMag_i, g.modelMag_z, (g.modelMag_u - g.modelMag_g) as u_g, (g.modelMag_g -
g.modelMag_r) as g_r, (g.modelMag_r - g.modelMag_i) as r_i, (g.modelMag_i -
g.modelMag_z) as i_z from Galaxy as g left outer join SpecObjAll as s on g.specobjID =
s.specobjID where g.ra between 132.8 and 132.9 and g.dec between 11.7 and 11.8 and
```

[Dataset Selector](#) | [toggle](#)

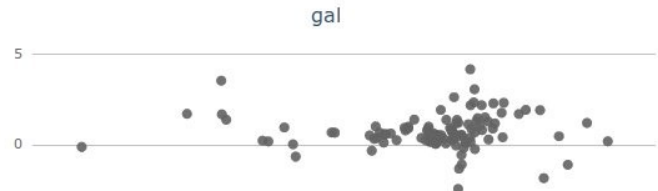
☒ gal (113)

☒ default (113)

[Plot](#) | [toggle](#)

Dataset: X Axis: Y Axis:

Selection Color:
☒ Select
☐ Zoom
Info: ☐



[Data Table](#) | [toggle](#)

VizieR

VizieR Service

Search Criteria

Preferences

max: 50

HTML Table

☐ All columns

[Compute](#)

Mirrors

CDS, France

Find catalogs among 10487 available

Clear Find...

Expand search ☐

Wavelength

Radio	AKARI	AGN
IR	ANS	Abundances
optical	ASCA	Ages
UV	BeppoSAX	Associations
EUV	CGRO	Atomic_Data
X-ray	COBE	BL_Lac_objects
Gamma-ray	Chandra	Binaries:cataclysmic

Search by Position across 10939 tables

Target Name (resolved by [Sesame](#)) or Position:

Clear J2000

Target dimension: 2 arcmin

☒ Radius ☐ Box size

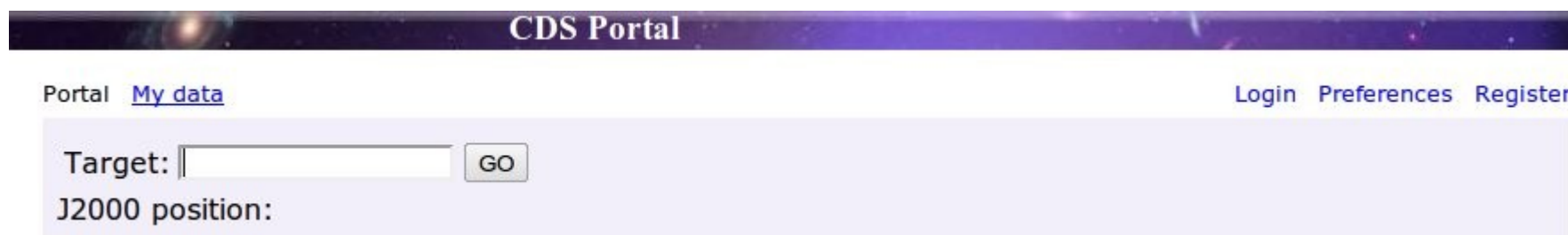
[More about VizieR](#)

[Find Catalogs](#)

CDS portal

<http://cdsportal.u-strasbg.fr/>

- Simple interface
- Aggregate results from SIMBAD, Aladin, VizieR...
- Input currently limited to simple input
 - Object name
 - Position



The screenshot shows the CDS Portal interface. At the top is a dark purple header with the text "CDS Portal" in white. Below the header, on the left, is the text "Portal" followed by a link "My data". On the right, there are links "Login", "Preferences", and "Register". Below these links is a search area with a label "Target:" followed by a text input field and a "GO" button. Below the search area is a label "J2000 position:" followed by a large empty text area.

CDS portal

- Planned improvements
 - Data & service discovery using additional inputs
 - Smart interpretation of input
 - Keep it simple

CDS Portal

Portal [My data](#) seb Preferences Admin Logout

Target:

J2000 position for VV 245b (NGC 4038): 12 01 52.48 -18 52 02.9

Object identifiers, measurements and bibliography

- Object type: Galaxy in Pair of Galaxies
- Morphological type: Sc
- [More SIMBAD data for NGC 4038](#)
- [1026 bibliographic references](#)
- [338 objects within 2'](#)
- [Display map around NGC 4038](#)
- [Display SimPlay interactive map around NGC 4038](#)
- [Related objects in bibliography](#)

Images for NGC 4038

- [Display region in Aladin \(Web Start\)](#)

Aladin images

Survey	Band	λ (nm)	Size	Epoch	Resolu
DEHS	I	0.79	12.5' x 12.4'	1996-03-30	0.9" / p
DEHS	J	1.23	12.6' x 12.7'	1996-03-30	0.9" / p
DEHS	K	2.16	12.6' x 12.7'	1996-03-30	0.9" / p
2MASS	K	2.16	8.5' x 17.0'	1999-03-20	0.9" / p
2MASS	H	1.65	8.5' x 17.0'	1999-03-20	0.9" / p
2MASS	J	1.24	8.5' x 17.0'	1999-03-20	0.9" / p
AAO	R	0.63	12.9' x 12.9'	1996-02-26	1.0" / p
SERC	I	0.80	12.9' x 12.9'	1996-05-25	1.0" / pixel

[JPG](#) [FITS](#)

Catalogues for NGC 4038

- [6 catalogues with 'NGC 4038' keyword](#)
- [147 catalogues around NGC 4038:](#)

VizieR catalogues

Name	Description	gth	Popularity	Coverage map
JApJS /154/519	Properties of ultraluminous X-ray candidates (Swartz+, 2004) [ReadMe]		33	
Bxmm	XMM-Newton Observation Log (XMM-Newton Science Operation Center, ...)		55	
VIRICA	SPECIFIC V2.0 Catalog of radio continuum spectra (Vollmer+ 2009) [ReadMe]		70	
W277	UBVRI photometry of faint field stars (Shiff, 2007) [ReadMe]		55	
Chandra	The Chandra Archive Log (CXC, 1999-) [ReadMe]		56	
IX25	Einstein Observatory IPC Parameters (HEAD-2, 1978-1981) [ReadMe]	4	X-ray	53
Jother /ApJ/30.1	Southern catalogue of isolated pairs extension (Reduzzi+ 1995) [ReadMe]	3	optical	42
JAZH78/675	Quasar-galaxy associations (Bukharina, 2001) [ReadMe]	3	optical	44
JMRAS /28/1181	Extragalactic supernovae (Iwaszardian+, 2001) [ReadMe]	3	optical	37
JAS /127/117	Kinematics of galaxies (Frugniel+ 1998) [ReadMe]	3	optical	45

Displaying 31 - 40 of 147

© CDS/CNRS Contact: [✉](#)

Use cases

- Redshift of 3C273
- Globular clusters in M31
- Proper motion of brown dwarves
- SED for Vega
- Veron catalogue
- QSOs with redshift > 5
- Cross match SDSS and 2MASS

Exemple 1 : Google



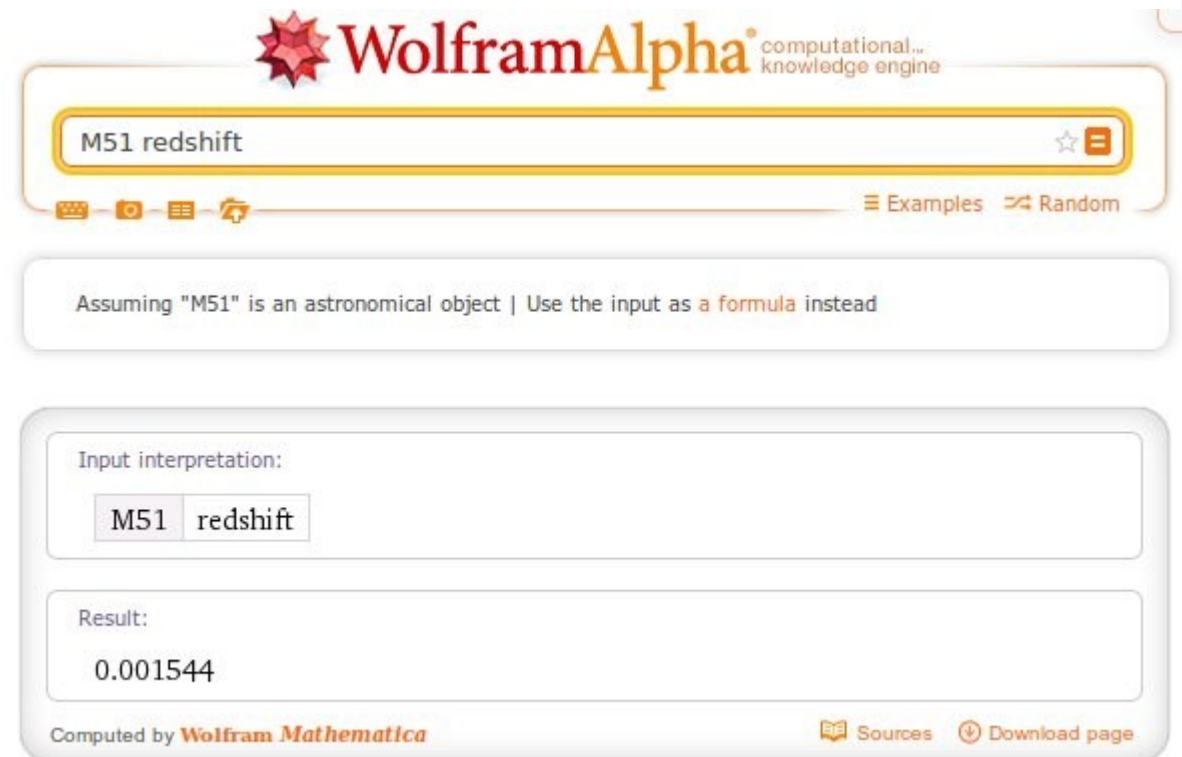
Google Search

I'm Feeling Lucky

- User-side
 - Simple
 - Suggestions
 - Sorted results
- Server-side
 - Indexing
 - Fast
 - Fuzzy search

Exemple 2 : Wolfram alpha

- User-side
 - Simple
 - « Smart »
- Server-side
 - Interpretation
 - Knowledge base

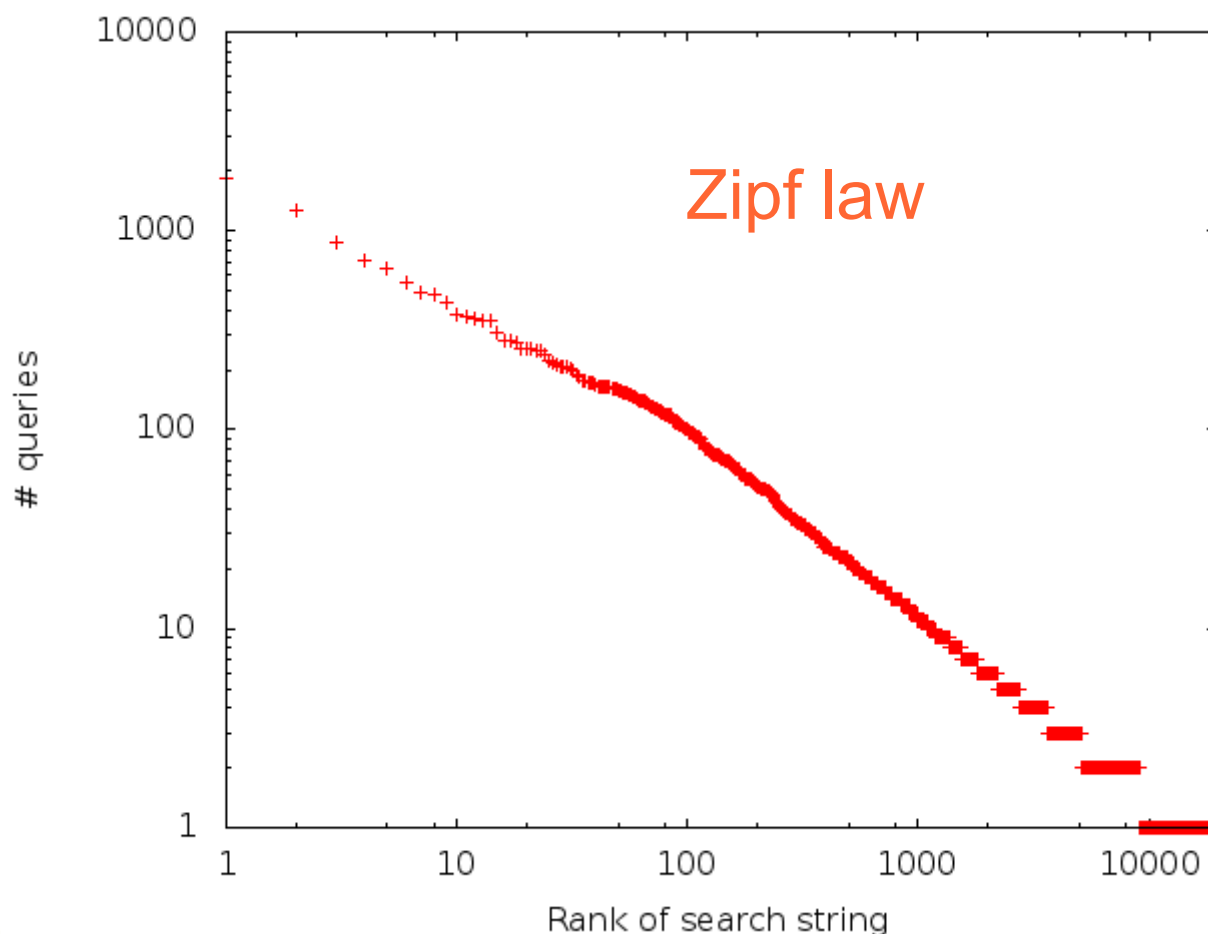


Basic steps for smart portal

- What do you mean ?
 - Stemming (star vs stars), lemmatization (quasar vs QSO)
 - Named-entity recognition (NER)
- What can I do with that ?
 - Match the tagged search string to search templates in various contexts
 - Translate the query in each context

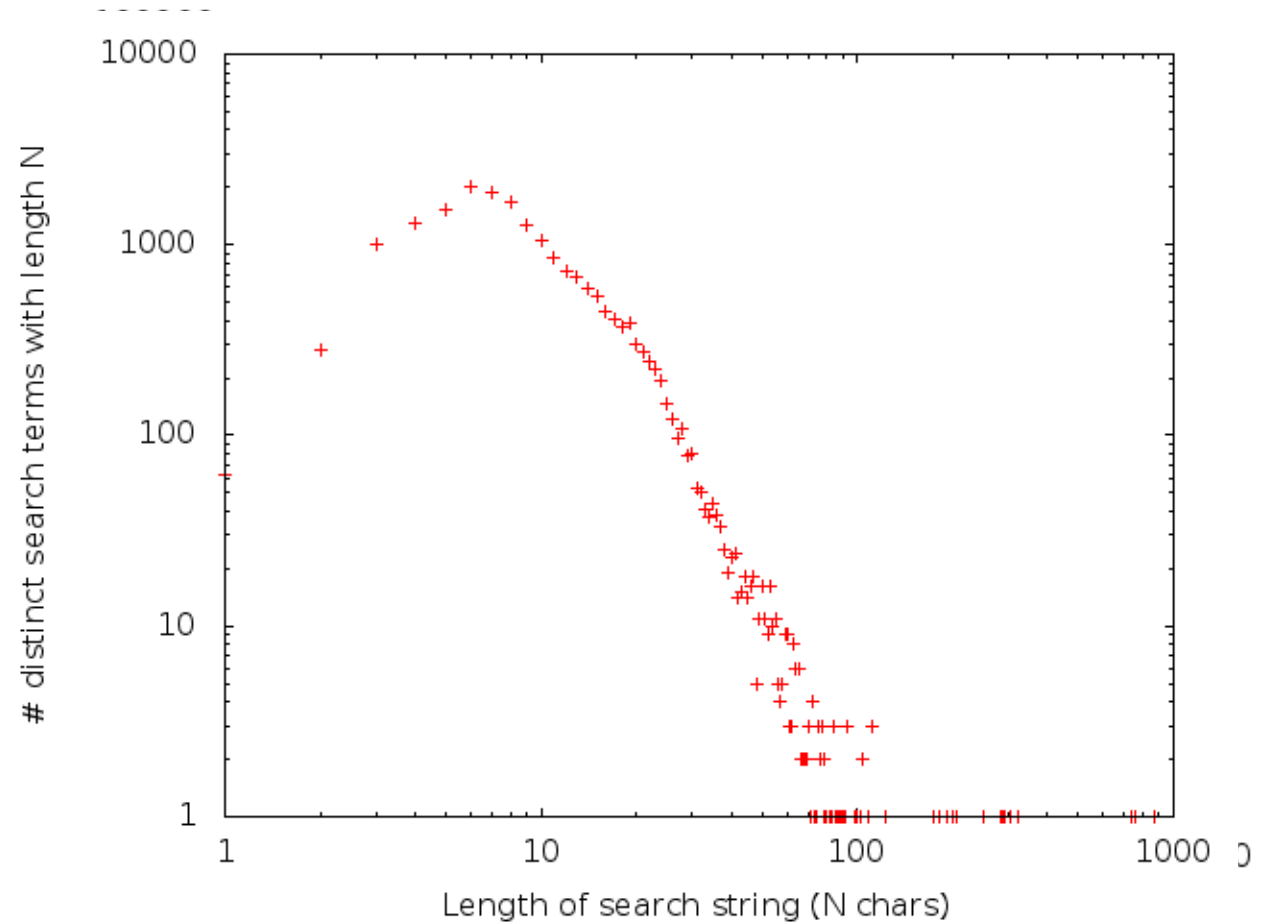
What do astronomers search ?

- Analysis of VizieR logs on « open » search
 - Expressions searched



Length of search strings

5 or 6
chars !!



Categorization of search terms

- Catalogue
- Mission, instrument, observatory
- Author
- Measurement
- Object type
- Object name, position
- Data product

Use cases

- Redshift of 3C273
- Globular clusters in M31
- Proper motion of brown dwarves
- SED for Vega
- Veron catalogue
- QSOs with redshift > 5
- Cross match SDSS and 2MASS

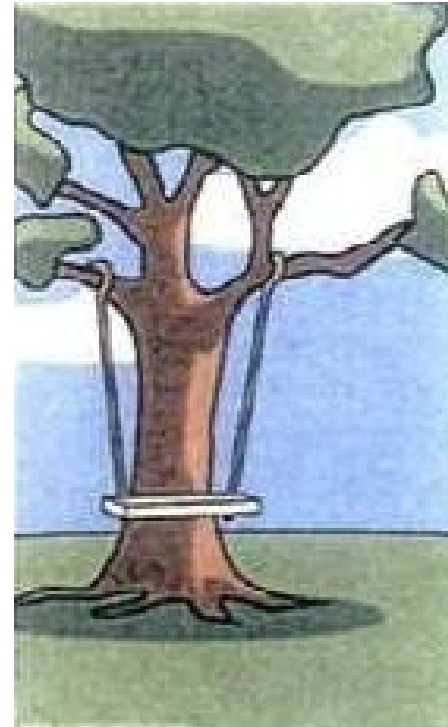
Smart portal



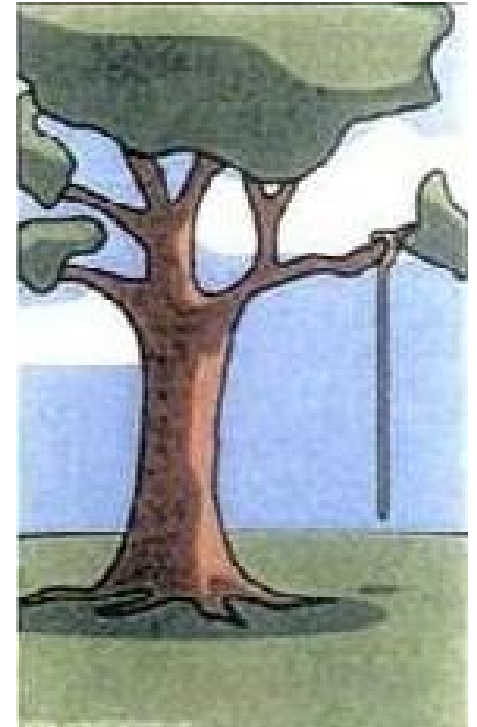
What the
user wants



How he
expresses it



How the
portal
understands
it



What the
portal
returns

Interpret & tag user input

- The simple way :
 - let the user do the tagging
 - Ex : Google `site:cds.unistra.fr` or
ADS `author:"Veron"`
- The hard way :
 - Go figure out
 - Use vocabularies (but not only)
- In both cases :
 - Do it on the fly & provide suggestions

Examples :

CDS Smart Portal Demo

Try searching for: cat, mission, author, quantity, otype, target, or product.

Q TARGET: 3c273 QUANTITY: redshift

CDS Smart Portal Demo

Try searching for: cat, mission, author, quantity, otype, target, or product.

Q TARGET: 3c273 QUANTITY: a

absolute magnitude
abundance
age

Suggesting
means sorting &
filtering

Tagging & suggesting

- Use reference vocabularies, thesauri
 - Object types, measurements
- But also :
 - Author names (papers, catalogues...)
 - Object names
 - > 12 million identifiers in SIMBAD
 - ~ 10 billion in VizieR
 - Dictionary of nomenclature
- Search it all at once !

Sorting things out

- Mandatory to face data avalanche
- Make use of usage statistics
 - Most queried catalogues
 - Papers with more citations
 - ... bias ? Winner take all ?
- Use trends
 - Short-term statistics
 - High citation rate vs total citations
- Computing relevance
 - e.g. Object name cited in title of paper

One query in various contexts

- Quasars with redshift > 5
- SIMBAD
 - otypes=QSO
 - redshift>5
- VizieR : relevant catalogues
 - kw.Astronomy=QSOs
 - UCD = src.redshift*
- Metadata mappings

Summary

- Smart portal for data, services & new features discovery
- Extreme text mining
- Learn from your logs
- There's no such thing as too many metadata
- Sorting, suggesting
- Learn from portal usage
 - Customized output for registered users
- Things are changing : voc, new data, obsolete
- Do not underestimate carbon-based intelligence