

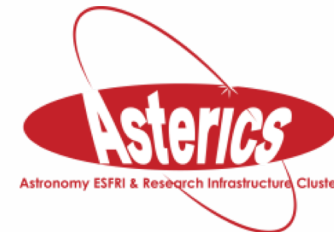
# Querying in NL on the semantics side.

---

André Schaaff, Thomas Boch, Sébastien Derriere  
Centre de Données astronomiques de Strasbourg

Alexis Guyot  
IUT Dijon

**IVOA Victoria, 28/05-01/06/2018**  
**Semantics Session**



# □ Purpose

- First presentation in Santiago Semantics WG to relate a R&D work done to evaluate the **Natural Language Processing** usage to the **query astronomical data services** (at CDS in a first step)
- Comments mainly **enthusiastic**
- It can be used for outreach but is it possible to reach query **results satisfying professional astronomers** ?
- How to reach this “level” ?

# □ We need the VO

- Link with the VO and with Semantics ?
  - We use VO standards like TAP, [UCDs](#), ...
- The VO enables the interoperability which is a mandatory backbone, helping us to query our services in NL and which will be useful in a step++ to query the whole VO through this way.

# □ What we are doing

- A **pragmatic** approach (more R&D than R), we have
  - authors in Simbad, VizieR
  - missions and **wavelengths** in VizieR
  - DJIN to recognize **identifiers** in a text
  - UCDs
  - a name resolver
  - ADQL / TAP
  - ...

What is the **effective temperature** of **Sirius**?

UCD :  
phys.temperature.effective

Identifier :  
\*alf CMa

# □ A set of queries in NL

1. **What is the redshift of 3C273? What is the redshift of the Virgo Cluster?**
2. **What is the parallax of Barnard's star? What is the distance of Barnard's star? What is the proper motion of Barnard's star?**
3. **What is the effective temperature of Sirius?**
4. **What are the galactic coordinates of Geminga?**
5. **Which galaxy interacts with NGC 4038?**
6. **Show me an image of the Pleiades in the K band**
7. **How many QSOs are there at redshift larger than 6? How many QSOs are there at  $z > 6$ ?**
8. **What is the redshift of galaxies members of the Virgo cluster?**
9. **Find globular clusters within  $3^\circ$  of M31. Find globular clusters in M31.**
10. **Query the latest Veron catalogue**
11. **What is the period of Algol? List of periods of Algol-type stars.**

# □ From Natural Language to ADQL

- In the literature, many examples of translation from NL to SQL.
- With ADQL :

List the **QSOs** at **Z** **> 6**.

```
SELECT main_id, oid, rvz_redshift
FROM basic
WHERE otype = -14680064 AND rvz_redshift > 6;
```

Simbad, TAP query

What is the **effective temperature** of **Sirius**?

```
SELECT "VI/137/gum_mw".Teff
FROM "VI/137/gum_mw"
WHERE 1 = CONTAINS(POINT('ICRS', "VI/137/gum_mw"."RAJ2000",
"VI/137gum_mw"."DEJ2000"), CIRCLE('ICRS', 101.287155333,
-16.716115861, 20/3600.));
```

VizieR, TAP Query

# □ New prototype engine: Intents

The screenshot displays the Dialogflow console interface. The main content area is titled "Intents" and features a "CREATE INTENT" button. Below the title is a search bar labeled "Search intents". A list of intents is shown, including "Default Fallback Intent", "Default Welcome Intent", and several action intents such as "get\_catalogue", "get\_children", "get\_count", "get\_measure", "get\_parents", "get\_siblings", "image.context", "measure.context", "show\_image", and "try\_again". The left sidebar contains navigation options: "Intents" (highlighted), "Entities", "Fulfillment", "Integrations", "Training", "History", "Analytics", "Prebuilt Agents", and "Small Talk". The right sidebar has a "Try it now" section with a microphone icon and a message: "Please use test console above to try a sentence." Below this is a link to "See how it works in Google Assistant." The browser address bar shows the URL: <https://console.dialogflow.com/api-client/#/agent/7e9a042d-3fa2-4d36-a752-19ff88b56e0b/intents>. The Windows taskbar at the bottom shows the date 25/05/2018 and time 12:55.

# □ New prototype engine: Training

The screenshot shows the Dialogflow Training console interface. The main content area displays a table of training data with columns for Conversation, Requests, No match, and Date. The table includes entries for 'Children of Electra', 'Children of Sirius', and 'Show me m51'. A sidebar on the left contains navigation options like Intents, Entities, Fulfillment, Integrations, Training (selected), History, Analytics, Prebuilt Agents, and Small Talk. The top right of the console has an 'UPLOAD' button and a 'Try it now' section with a microphone icon and a message: 'Please use test console above to try a sentence.' Below this is a link to 'See how it works in Google Assistant.' A 'NEXT' button is located at the bottom right of the table area.

Conversation	Requests	No match	Date	
Children of Electra	2	0	Today	>
Children of Sirius	14	0	Today	>
Children of Sirius	7	0	Today	>
Children of Sirius	6	0	Today	>
Show me m51	9	2	May 24	✓
m 51	12	2	May 24	✓



# New prototype engine: User query and intent

The screenshot shows the Dialogflow console interface. A modal window displays training data for an agent. The modal title is "Most popular catalogue measuring redshift and talking about QSOs" and includes an "APPROVE" button. Below the title, it shows "May 24", "12 REQUESTS", and "0 NO MATCH".

The first training example is:

USER SAYS: Most popular catalogue measuring redshift and talking about QSOs

PARAMETER NAME	ENTITY	RESOLVED VALUE	
catalog_desc	@catalog_desc	Most popular	x
meas	@meas	redshift	x
otype	@otype	QSOs	x

INTENT: `get_catalogue`

The second training example is:

USER SAYS: Query the latest Veron catalogue

INTENT: `get_catalogue`

# New prototype engine: Examples of sentences linked to get\_measure

The screenshot displays the Dialogflow console interface. On the left, a sidebar menu includes options like 'Intents', 'Entities', 'Fulfillment', 'Integrations', 'Training', 'History', 'Analytics', 'Prebuilt Agents', and 'Small Talk'. The main area is titled 'get\_measure' and contains a list of 13 training sentences. The first sentence is: "What is the parallax, the spectral type, the position, the proper motion, the edshift, the magnitudes and the distance of Sirius". Other sentences include: "Galactic coordinates and spectral type of Arcturus", "Tell me everything about T Tau", "Could you search the proper motion of pulsars", "Tell me everything about Andromeda", "Parallax of Sirius", "Parallax of Electra", "Parallax and redshift of Sirius", "Everything about Electra", and "What is the parallax of M31". A 'SAVE' button is visible at the top right of the main area. The bottom of the image shows a Windows taskbar with the search bar, system tray, and a clock showing 13:02 on 25/05/2018.

# □ Comments

- Vocabular “sensitive”
- A use case for Semantics ? Vocabularies ?
- Ideas ?
- Similar works ?
- Demo on Friday during the Apps session