

# Chatting with the services. VO Standards and NLP.

---

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

Alexis Guyot  
IUT Dijon

**IVOA Victoria, 28/05-01/06/2018**  
**Apps Session 2**



# □ Purpose

- First presentation in Santiago Semantics WG: the aim was to relate a budding R&D work, **Natural Language Processing** applied to the **querying of astronomical data services** and to collect comments (maybe **enthusiastic**), ideas, recommendations, or to initiate collaborations
- We are not geeks and we would like to propose this **way of interaction** in ... the future.
- Voice recognition ?
- ... is it possible to reach query **results satisfying professional astronomers** ?

# □ Comments

- Link with the VO ?
  - We use VO standards like TAP, UCDS, ...
  - We use tools based on ML, a link with KDD.
- 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.

# □ NL interaction is “common”

- Because it is natural (Apple Siri, Ok Google, Microsoft, etc.), assistants at your home to book a restaurant, find a medical appointment, ... in the every day life.
- We use forms (parameter fields, checkboxes, etc.) to query astronomical data services.

The Point Source catalogue of 470,992,970 sources. Please [acknowledge the usage of the 2MASS All-Sky Survey](#); see also the [introductory page](#).

Simple Constraint List Of Constraints

Query by [Constraints](#) applied on Columns (Output Order:  +  -)

Show	Sort	Column	Clear	Constraint	Explain (UCD)
<input checked="" type="checkbox"/>	<input type="radio"/>	RAJ2000		<a href="#">deg</a>	(ra) Right ascension (J2000) ( <a href="#">pos.eq.ra:meta.main</a> )
<input checked="" type="checkbox"/>	<input type="radio"/>	DEJ2000		<a href="#">deg</a>	(dec) Declination (J2000) (dec) ( <a href="#">pos.eq.dec:meta.main</a> )
<input type="checkbox"/>	<input type="radio"/>	errMaj		<a href="#">arcsec</a>	(err_maj) Semi-major axis of position error ellipse ( <a href="#">stat.error</a> )
<input type="checkbox"/>	<input type="radio"/>	errMin		<a href="#">arcsec</a>	(err_min) Semi-minor axis of position error ellipse ( <a href="#">stat.error</a> )
<input type="checkbox"/>	<input type="radio"/>	errPA		<a href="#">deg</a>	[0,180] (err_ang) Position angle of error ellipse major axis (E of N) ( <a href="#">stat.error</a> )
<input checked="" type="checkbox"/>	<input type="radio"/>	2MASS		(char)	(designation) Source designation ( <a href="#">Note 1</a> ) ( <a href="#">meta.id:meta.main</a> )
<input checked="" type="checkbox"/>	<input type="radio"/>	Jmag		<a href="#">mag</a>	<sup>(n)</sup> (j_m) J selected default magnitude ( <a href="#">Note 2</a> ) ( <a href="#">phot.mag:em_IR.J</a> )
<input type="checkbox"/>	<input type="radio"/>	Jmsig		<a href="#">mag</a>	<sup>(n)</sup> (j_cmsig) J default magnitude uncertainty ( <a href="#">Note 3</a> ) ( <a href="#">stat.error:phot.mag</a> )
<input checked="" type="checkbox"/>	<input type="radio"/>	e_Jmag		<a href="#">mag</a>	<sup>(n)</sup> (j_msigcom) J total magnitude uncertainty ( <a href="#">Note 4</a> ) ( <a href="#">stat.error:phot.mag:em_IR.J</a> )
<input type="checkbox"/>	<input type="radio"/>	Jsnr			<sup>(n)</sup> (i_snr) J Signal-to-noise ratio ( <a href="#">stat.snr</a> )

# □ What we are doing

- Learning about NLP (basis, tools, examples, ...)
  - Scope of the study: a first **set of queries**
  - Too large -> too much time and resources.
- 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 first 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

# □ Previous prototype

- Was based on Stanford NLP (POSTagger), DJIN, IVOA UCD and ADQL/TAP, ...
- Far from a chatbot

What is the effective temperature of Sirius ?

How many planets orbit Kepler 20 ?

What is the redshift of galaxies members of the Virgo cluster ?

Natural Language Processing for Astronomy

Which database will you Query ?

Simbad  
 VizieR  
 Simbad+VizieR

Query:

[Sirius] is a : [Identifier] Found as : [none] with Tag :Identified via Service

[effective] is a : [Unrecognised] Found as : [none] with Tag :J

[temperature] is a : [Unrecognised] Found as : [none] with Tag :NN

VizieR Results :

From Query : SELECT TOP 100 "III/193/catalog".theta FROM "III/193/catalog" WHERE 1 = CONTAINS(POINT('ICRS',"III/193/catalog"." RA","III/193/catalog"." DE"), CIRCLE('ICRS',101.287155333,-16.716115861, 20/3600.));

theta 0.51

From Query : SELECT TOP 100 "III/200B/fistars".Teff FROM "III/200B/fistars" WHERE 1 = CONTAINS(POINT('ICRS',"III/200B/fistars"." RA","III/200B/fistars"." DE"), CIRCLE('ICRS',101.287155333,-16.716115861, 20/3600.));

Teff 9333



# □ New prototype

- Chatbot approach to reduce the gap between “good” queries and unprecise / ambiguous queries.
- Based on Google Dialogflow
  - Many features
  - Machine learning
  - Engine is on the Google side, it is free, multiple users at the same time through a “Session” ID.

# Training

Dialogflow console interface showing the Training page. The page displays a table of training data with columns for Conversation, Requests, No match, and Date. The table lists several training examples, including 'Children of Electra', 'Children of Sirius', and 'Show me m51'. A sidebar on the left contains navigation options like Intents, Entities, Fulfillment, Integrations, Training, 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. The bottom of the image shows a Windows taskbar with the date 1/06/18 and the system tray showing 12:55 on 25/05/2018.

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	✓

# Intents

The screenshot shows the Dialogflow console interface. The browser address bar displays the URL: <https://console.dialogflow.com/api-client/#/agent/7e9a042d-3fa2-4d36-a752-19ff88b56e0b/intents>. The main content area is titled "Intents" and features a "CREATE INTENT" button. A search bar labeled "Search intents" is positioned above a list of intent names. The list includes:

- Default Fallback Intent
- Default Welcome Intent
- get\_catalogue
- get\_children
- get\_count
- get\_measure
- get\_parents
- get\_siblings
- image.context
- measure.context
- show\_image
- try\_again

On the right side of the console, there is a "Try it now" section with a microphone icon and a message: "Please use test console above to try a sentence." Below this, there is a link: "See how it works in Google Assistant."

The left sidebar contains navigation options: Intents (selected), Entities, Fulfillment, Integrations, Training, History, Analytics, Prebuilt Agents, and Small Talk. The user profile "astroboy" and language "en" are also visible.

# □ User query and intent

Dialogflow console showing training data for an agent. The main window displays a user query: "Most popular catalogue measuring redshift and talking about QSOs". Below the query is a table of extracted entities and their resolved values. The intent is identified as "get\_catalogue". A second example query is also shown: "Query the latest Veron catalogue".

Most popular catalogue measuring redshift and talking about QSOs

May 24 12 REQUESTS 0 NO MATCH

APPROVE

USER SAYS

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

INTENT `get_catalogue`

USER SAYS

Query the latest Veron catalogue

INTENT `get_catalogue`

# Examples of sentences linked to get\_measure

The screenshot displays the Dialogflow console for the 'get\_measure' intent. The left sidebar shows the navigation menu with 'Intents' selected. The main content area is divided into sections: 'Contexts', 'Events', and 'Training phrases'. Under 'Contexts', there are two output contexts: '5 searched\_object' and '5 searched\_meas'. Under 'Training phrases', a sample sentence is shown with highlighted entities: 'parallax', 'spectral type', 'position', 'proper motion', 'redshift', 'magnitudes', and 'distance of Sirius'. The right sidebar contains 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'.

# Examples of sentences linked to get\_measure (2)

The screenshot displays the Dialogflow console interface. The main area shows a list of 13 example sentences for the 'get\_measure' intent, with the first one selected. The examples are:

- What is the parallax, the spectral type, the position, the proper motion, the redshift, the magnitudes and the distance of Sirius
- 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
- What is the parallax of M31

The interface includes a left sidebar with navigation options (Intents, Entities, Fulfillment, Integrations, Training, History, Analytics, Prebuilt Agents, Small Talk), a top navigation bar with 'Dialogflow' and 'get\_measure' tabs, and a right sidebar with a 'Try it now' section. The bottom of the image shows a Windows taskbar with the time 13:02 and date 25/05/2018.

# Parameters

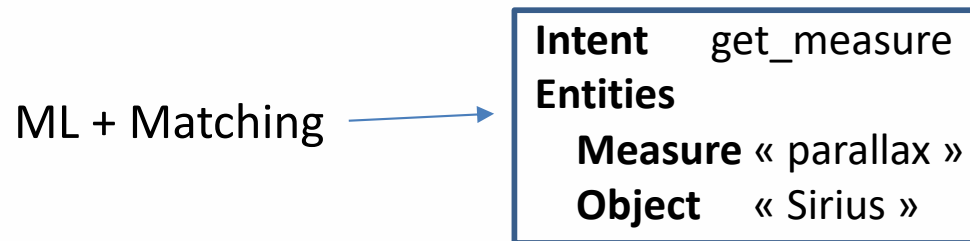
The screenshot shows the Dialogflow console interface. The left sidebar contains navigation options: astroboy, Intents (selected), Entities, Fulfillment, Integrations, Training, History, Analytics, Prebuilt Agents, and Small Talk. The main area displays the configuration for the intent 'image.context'. Under the 'Action and parameters' section, there is a table of parameters:

REQUIRED	PARAMETER NAME	ENTITY	VALUE	IS LIST
<input type="checkbox"/>	oid	Enter entity	#searched.o id	<input type="checkbox"/>
<input type="checkbox"/>	wavelength	@wavelength	\$wavelength	<input type="checkbox"/>
<input type="checkbox"/>	Enter name	Enter entity	Enter value	<input type="checkbox"/>

Below the table is a '+ New parameter' link. The right sidebar shows a 'Try it now' button and a message: 'Please use test console above to try a sentence.' with a link to 'See how it works in Google Assistant.' The Windows taskbar at the bottom shows the date 1/06/18, the time 13:03, and the date 25/05/2018.

# □ Before the demo

**What is the parallax of Sirius ?**



Object name or type ?

TAP query to Simbad to get back the parallax

If no result in Simbad, TAP query to VizieR

The result is « dressed » with Natural Language

**The parallax of Sirius is ...**





# □ Demo

- Only the user interface is on my laptop

# □ Please, give us a feedback

- Comments ?
  - Enthusiastic like in Santiago ?
- Ideas ?
- Recommendations ?
- Similar works ?

# □ Conclusion

- We are testing and improving step by step.
- We are (somewhere) on a way to **interact with data services** differently.
- We will enjoy to have a few volunteers to try it and to give us a feedback.