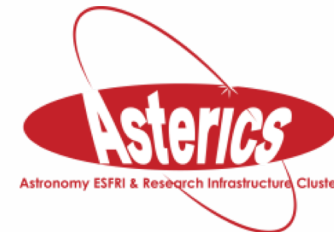# Querying in NL on the semantics side.

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° 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

Comments…

New prototype engine: Intents

New prototype engine: Training

New prototype engine: User query and intent

IVOA Victoria

New prototype engine: Examples of sentences linked to get_measure

Comments

- Vocabular "sensitive"

- A use case for Semantics ? Vocabularies ?

- Ideas ?

- Similar works ?

- Demo on Friday during the Apps session