CENTRE DE DONNÉES
ASTRONOMIQUES DE STRASBOURG

# Investigating a user-registry search engine

G.Landais

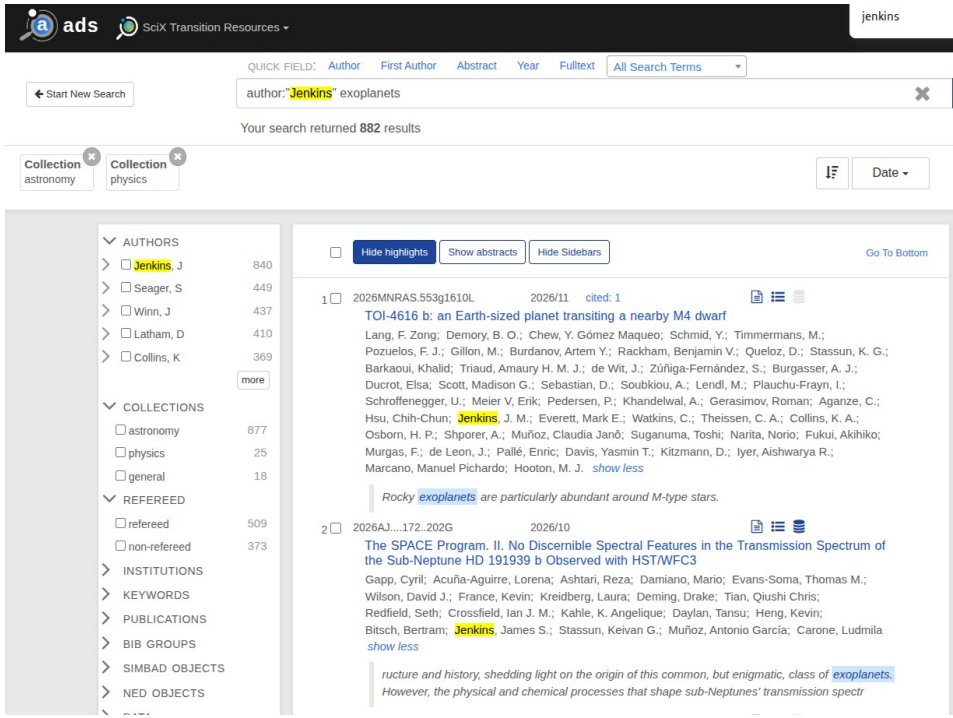


# Search-engine motivations

regTAP is the norm and is well desgigned for sofware
… but not really adapted for human.

**Give an Human-search alternative to discover resources available in the VO registry**

- A search engine like ADS bumblebee (or prototype developped by J. Lenorman , Paris Observatory)
- A query-language allowing powefull textual search:
  - Applied on specific or on whole fields
    eg: content.subject: “multiple-stars”
    eg: “proper motion”
  - Allowing wildcards, fuzzy or exact search
    eg: author: schmitt~
  - Grammar integrated (tokenization) :
    rotation, totate, rotations → rotat
  - Accept Numeric / Date constraint:
    eg: date:>2026-01-01
- User-friendly WEB Interface



# Search-engine Prototype

## Testing Opensearch

A prototype in CDS with content coming from CDS registry
(compatible with any registry: GAVO, etc)

- Transform VO Registry record: (OAI-PMH) XML → JSON
- Populate a JSON-OpenSearch Database
- Tuning the configuration based on weights applied to (JSON) fields
- Query the database with OpenSearch API
  example:
  OpenSearch API:
  http://localhost:9200/registry/search?q=proper+motion

OpenSearch
https://opensearch.org/

**OpenSearch**

- Apache 2.0 licensed
- direct fork of Elasticsearch 7.10.2 (from AWS)
- JSON API

```
curl -XGET "http://localhost:9200/_search?pretty"  \
    -H 'Content-Type: application/json' \
    -d '{ "query": { "simple_query_string": { "query": "binary stars __created:>2020", "fields": ["title","_created"] } } }'
```

**TODO**: make  better !
→ Opensearch combined with **MOC** search
(using mocset/moc-cli tools (F-X. Pineau) that allows moc in order>6)

# User-interface

This interface has been created by the Mistral AI using a simple prompt

