Registry session 2

Table discovery and mining

Sebastien Derriere, IVOA Beijing Interoperability meeting, may 2007

Registry graininess

Coarse grained

Fine grained

Use cases

- Use the Registry for Table discovery:
  - Find catalogues containing measures of trigonometric parallaxes
  - Find catalogues with more than 100 sources and containing J or K-band magnitudes
  - Extract, RA, DEC, pos. error, V-band mag and redshift from catalogues A, B and C

Coarse-grained registry

- Relies on general catalogue description, keywords (<subject>, <description>)
- Need 2-steps approach to query on detailed metadata (irrealistic for 10^4 resources)
- Catalogue mining can be biased by incomplete descriptions
- Favors large surveys (e.g. search for 2MASS, SDSS...)

Fine-grained registry

- Queries possible on every metadata element (possible to write Xquery)
- Direct access to many DataCollections' descriptions at registry level
- Serendipitous catalogue discovery
- More visibility of small DataCollections

Tools with fine-grained registries

- Registry
  - Containing full Catalogue description: <table><columns><ucd>
  - Support for Xquery
- Examples :
  - CDS: UCD registry query tool (VOTech)
  - AstroGrid: VOExplorer

CDS registry query tool

- Search tables by logical expression on UCD contents: UCD1 and (UCD2 or UCD3)
- Configurable access to remote Registry
  - XMLDB API, Xquery
- Successfully used by B. Vollmer to extend the work done on SPECFIND on radio catalogues

Data Extraction Tool

- Developed in VOTech project
- Takes list of ivo IDs as an input
- User defines a template for output columns
- Column metadata retrieved from registry to define an extraction schema for various tables
- Extract various catalogues in homogeneous format

Conclusions

- It takes time to fill in detailed metadata
  - but true for most metadata !
- More risks of metadata changes -> more curation
- It makes metadata volume larger !
- It allows advanced discovery/query
- Detailed metadata useful for:
  - large data providers
  - advanced users
  - tools that can automate boring tasks