A guide to astrosemantics

Matthew J. Graham (Caltech)
What?
What is astrosemantics?

- The branch of astronomy which deals with machine-processible knowledge (via semantic technologies)
- The promise:

Without:
- Data: 40%
- Information: 30%
- Knowledge: 20%
- Wisdom: 10%

With:
- Data: 10%
- Information: 20%
- Knowledge: 30%
- Wisdom: 40%
Why?
### NGC 7377

#### Classification

<table>
<thead>
<tr>
<th>Classification</th>
<th>Reference Code</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>1995A&amp;AS_110_..371S</td>
<td>in a pair</td>
</tr>
<tr>
<td>GR</td>
<td>1993A&amp;AS_100_..47G</td>
<td>Group member, radial-velocity confirmed</td>
</tr>
<tr>
<td>GR</td>
<td>1992A&amp;AS_93_..211F</td>
<td>Group member, radial-velocity confirmed</td>
</tr>
</tbody>
</table>

#### Galaxy Morphology

<table>
<thead>
<tr>
<th>Morphology</th>
<th>Reference Code</th>
<th>Classification</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SA0*(s)</td>
<td>1991RCJ9.C..0000D</td>
<td>SA0*(s)</td>
<td>Optical, Galaxy, Code: [LA9+..]</td>
</tr>
<tr>
<td>SO2/3 / Sa pec</td>
<td>1981RS...C..0000S</td>
<td>SO2/3 / Sa pec</td>
<td>Optical, Galaxy, Code: [S 4 0 P]</td>
</tr>
<tr>
<td>Sa</td>
<td>1982ESOU.C..0000L</td>
<td>Sa</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>3</td>
<td>1975SoByu..47_..43A</td>
<td>no nucleus</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>E</td>
<td>1996MNRAS_278_1023L</td>
<td>E</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>3</td>
<td>1975SoByu..47_..43A</td>
<td>continuous brightening towards center</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Silhouetted dust lanes</td>
<td>1994CAG1.B..0000S</td>
<td>Silhouetted dust lanes</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Multi-armed spiral</td>
<td>1994CAG1.B..0000S</td>
<td>Multi-armed spiral</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Most unusual galaxy</td>
<td>1994CAG1.B..0000S</td>
<td>Most unusual galaxy</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Intricate dust lanes pervade the disk</td>
<td>1994CAG1.B..0000S</td>
<td>Intricate dust lanes pervade the disk</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Dust lanes silhouetted on disk</td>
<td>1994CAG1.B..0000S</td>
<td>Dust lanes silhouetted on disk</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Dust lanes define the spiral structure</td>
<td>1994CAG1.B..0000S</td>
<td>Dust lanes define the spiral structure</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Asymmetric</td>
<td>1994CAG1.B..0000S</td>
<td>Asymmetric</td>
<td>Optical, Galaxy</td>
</tr>
<tr>
<td>Patchy dust</td>
<td>1988AJ...95_..422E</td>
<td>Patchy dust</td>
<td>Optical, Galaxy</td>
</tr>
</tbody>
</table>

Find me more by:
- Type
- Properties
- Distance
- ...
not just by tagging but also by inference
This is a prochous. A prochous is a vessel that was used to pour liquids. It has one handle, a large body, and a nib-like neck. This particular prochous was created during the Classical period. The Classical period is characterised by the initial political supremacy of Athens (its “golden age”), which was followed by the expansion of the Greek world under the rule of Alexander the Great of Macedonia. It covers the time between 480 and 323 B.C. This prochous dates from the second quarter of the 5th century B.C. It depicts a Greek hunting a Persian. Currently it is exhibited in the Museum für Kunst in Hamburg.”
What if…

“Planetary scientists believe that the same geological processes that have shaped the Earth – volcanism, tectonism, water, ice, and impacts – are at work on Mars.”

“Jupiter acts like a kind of ‘cosmic vacuum cleaner’ for the inner Solar System. It has been speculated that without it, complex life might not have been able to develop on Earth.”
Conceptual hierarchies

Credit: L. Eyer & N. Mowlavi (10/2007)
Incorporating knowledge

- Ontological SOM
  - NGC catalog
    - 7840 objects, 5886 are annotated in NED (more info in Wikipedia)
    - 2118 different terms

- Markov Logic Networks
  - A framework combining statistical and first-order logical reasoning
  - Rules associated with weight expressing probability
  - Infer most likely world or marginal probabilities
How?
The quantum of knowledge

simple

subject – predicate – object
Extracting knowledge from the literature

• Who?
  Babylonian Astronomical Diary

• What?
  The comet which previously had appeared in the east in the path of Anu in the area of Pleiades and Taurus

• Where?
  to the west [...] and passed along in the path of Ea in the region of Sagittarius, 1 cubit in front of Jupiter, 3 cubits high toward the north [...] 

• When?
  Month VIII, SE 148 (lunar month beg. 21 October 164 BC)

• How?
  By eye

• Why?
  Celestial divination
Mining to create concept schemes

- ATels
  - Hierarchically cluster vector representations
- ADS Labs
- Paragraph-level index for arXiv (idea courtesy David Hogg)
- NED annotations
Knowledge repositories

- 11013 VOEvents from 8 different event streams
  - 420760 parameters in all: `<param name="..." value="..."/>
- Different stores:
  - MySQL:
    - with a params table
    - with a separate table for each stream
  - MongoDB (NoSQL) – id, param name, value
  - StarDog (triple store)
  - eXist (native XML)
- Find me all events with a param called “Event_Flux” and a value < 40.0:
  - StarDog: 0.016s
  - MySQL (streams): 0.071s
  - Mongo: 0.195s
  - MySQL (params): 0.247s
  - eXist: 2.128s
Codifying knowledge

- Controlled vocabularies
  - VO Theory (http://votheory.obspm.fr), etc.

- Taxonomies
  - GCVS, etc.

- Thesauri:
  - IAUT, etc.

- Ontologies (concepts, relations, taxonomy, axioms)
  - CDS ObjectType ontology
  - Data Mining ontology
  - UMD astronomy ontology
  - NASA SWEET
Combining everything

- SKA Information Intensive Framework provides semantic services to manage data
  - Ingest – receives data in appropriate formats from instruments
  - Classify – uses an ontology-driven solution to classify data
  - Problem solvers – make inferences across information to solve astronomical problems (e.g. transient detection)
  - Search – access mechanisms for retrieving data/information/knowledge