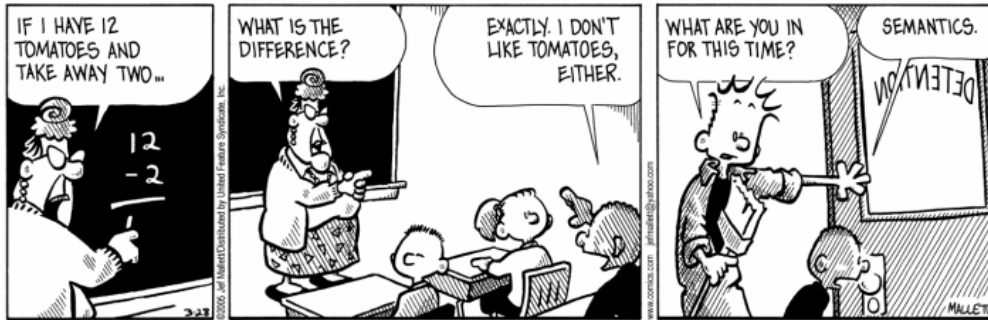




A guide to astrosemantics

Matthew J. Graham (Caltech)



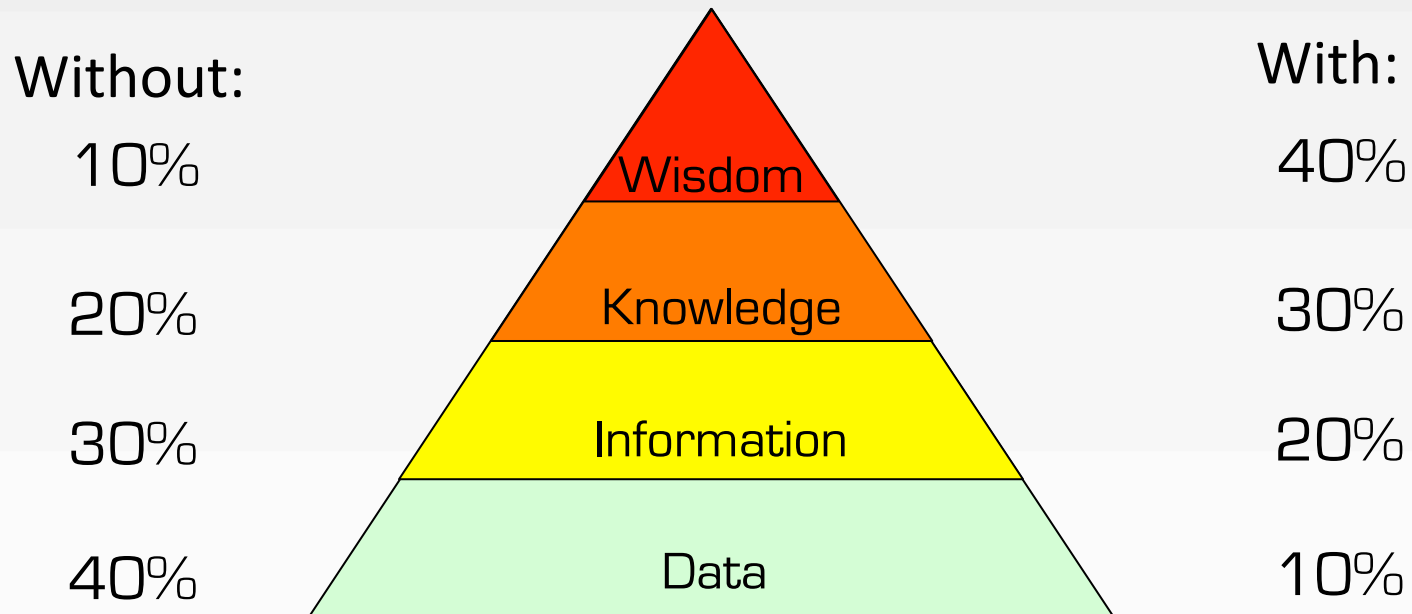


What?



What is astrosemantics?

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



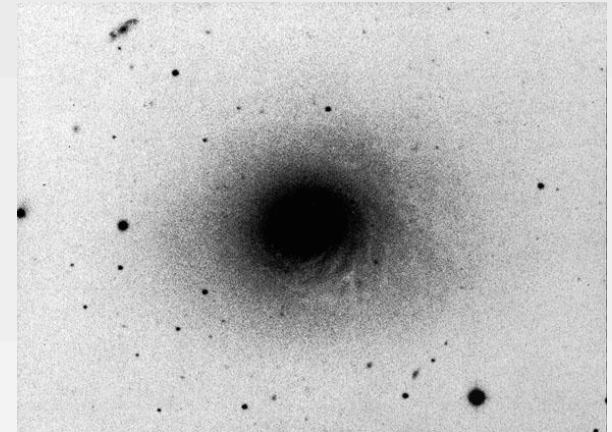


Why?



NGC 7377

Classification	Reference Code	Classification	Flag	Bandpass	Region	Notes
Hierarchy						
P	1995A&AS..110..371S	1 in a pair	...	Optical	Environment	see ESO 534- G 024
GR	1993A&AS...100...47G	1 Group member, radial-velocity confirmed	...	Optical	Environment	see GG 463
GR	1992A&AS...93..211F	1 Group member, radial-velocity confirmed	...	Optical	Environment	see NGC 7377 GROUP
Galaxy Morphology						
SA0 ⁺ (s)	1991RC3.9.C...0000d	1 SA0 ⁺ (s)	...	Optical	Galaxy	Code: [.LAS+..]
SO _{2/3} / Sa pec	1981RSA...C...0000S	1 SO _{2/3} / Sa pec	...	Optical	Galaxy	Code: [S 4 0 P]
Sa	1982ESOU..C...0000L	1 Sa	...	Optical	Galaxy	...
3	1975SoByu..47...43A	1 no nucleus	...	Optical	Galaxy	...
E	1996MNRAS.278.1025L	1 E	...	Optical	Galaxy	...
3	1975SoByu..47...43A	1 continuous brightening towards center	...	Optical	Galaxy	...
silhouetted dust lanes	1994CAG1..B...0000S	1 silhouetted dust lanes	...	Optical	Galaxy	...
multi-armed spiral	1994CAG1..B...0000S	1 multi-armed spiral	...	Optical	Galaxy	...
most unusual galaxy	1994CAG1..B...0000S	1 most unusual galaxy	...	Optical	Galaxy	...
intricate dust lanes pervade the disk	1994CAG1..B...0000S	1 intricate dust lanes pervade the disk	...	Optical	Galaxy	...
dust lanes silhouetted on disk	1994CAG1..B...0000S	1 dust lanes silhouetted on disk	...	Optical	Galaxy	...
dust lanes define the spiral structure	1994CAG1..B...0000S	1 dust lanes define the spiral structure	...	Optical	Galaxy	...
asymmetric	1994CAG1..B...0000S	1 asymmetric	...	Optical	Galaxy	...
Patchy dust	1988AJ....95..422E	1 patchy dust	...	Optical	Galaxy	...



Find me more by:

- Type
- Properties
- Distance
- ...

not just by tagging but
also by inference



Artificial docents

Welcome to the ground floor!

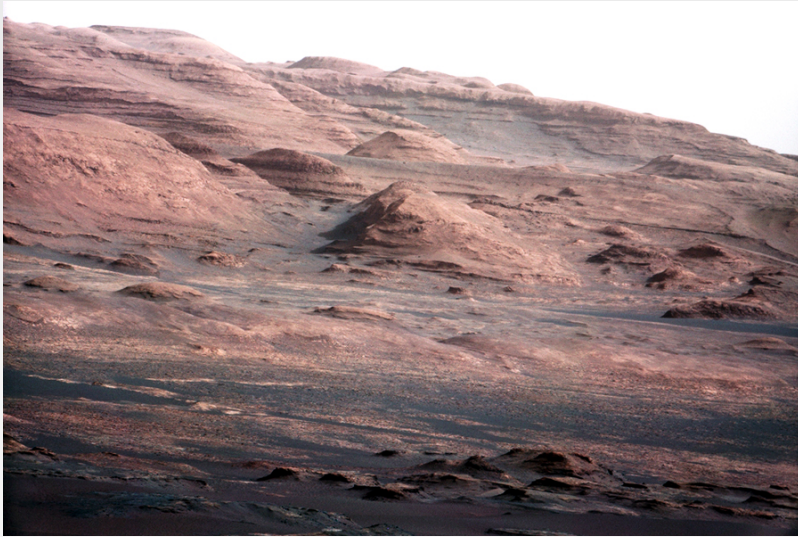
- What the robot says about the exhibits is not written in advance! The robot uses natural language generation technology to produce texts automatically from knowledge stored in logical form and small pieces of text. This allows the robot to tailor what it says, for example by including comparisons to previous exhibits. It also allows it to generate texts in multiple languages from the same knowledge representation. The robot currently speaks English and Greek.



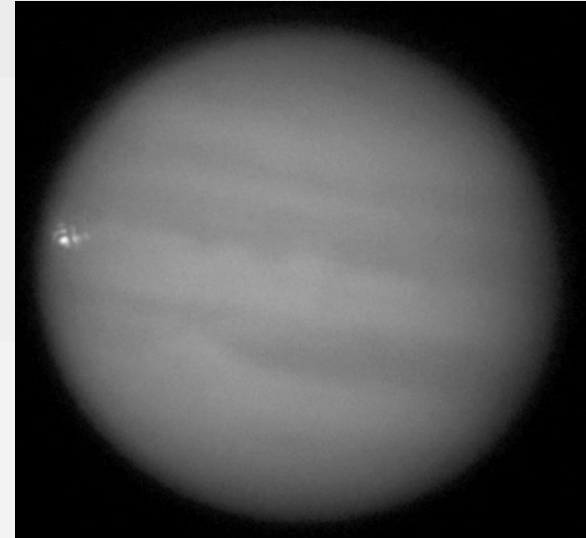
“This is a prochous. A prochous is a vessel that was used to pour liquids. It has one handle, a large body, and a nib-like neck. This particular prochous was created during the Classical period. The Classical period is characterised by the initial political supremacy of Athens (its “golden age”), which 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...



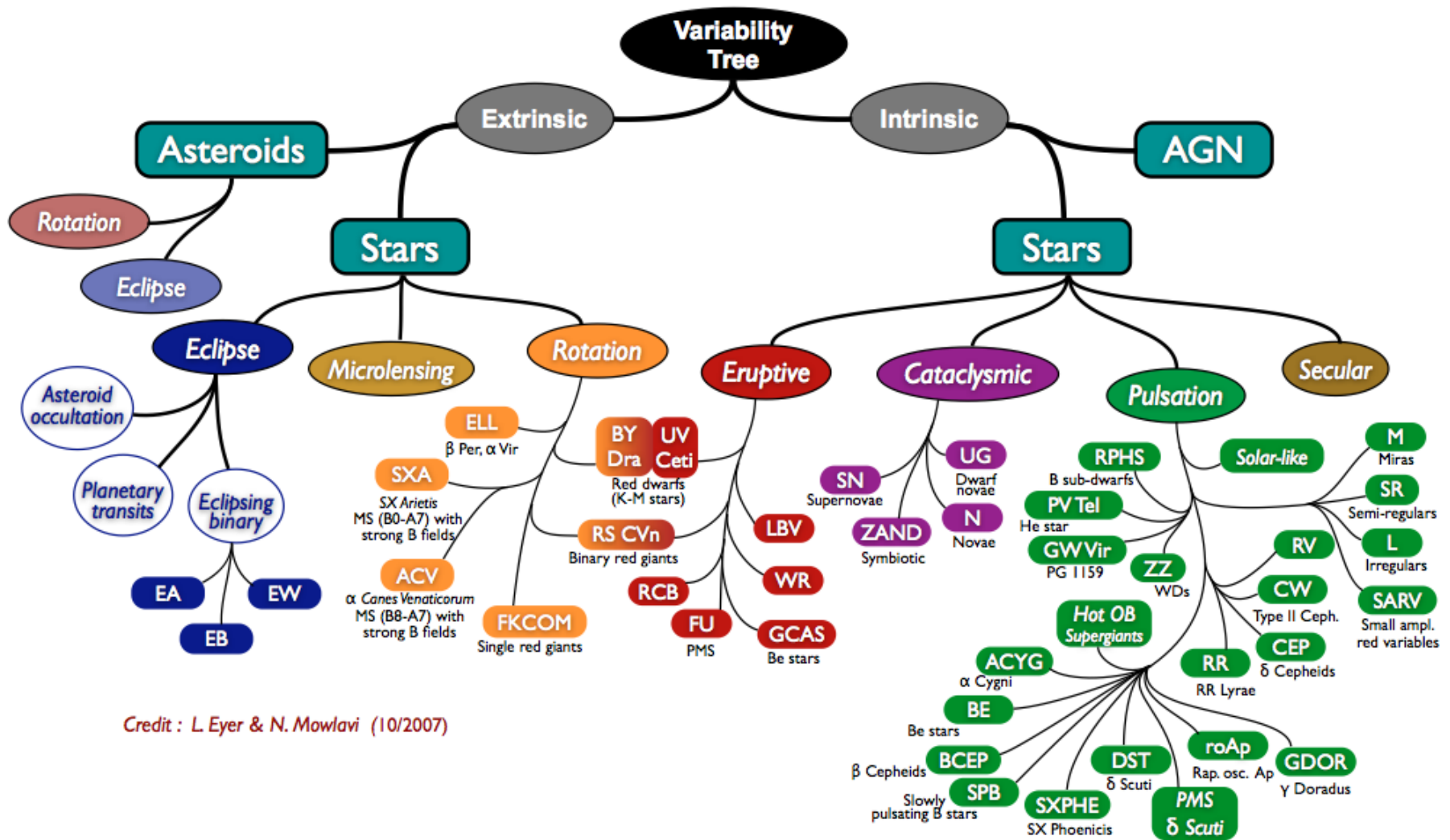
“Planetary scientists believe that the same geological processes that have shaped the Earth – volcanism, tectonism, water, ice, and impacts – are at work on Mars”



“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)

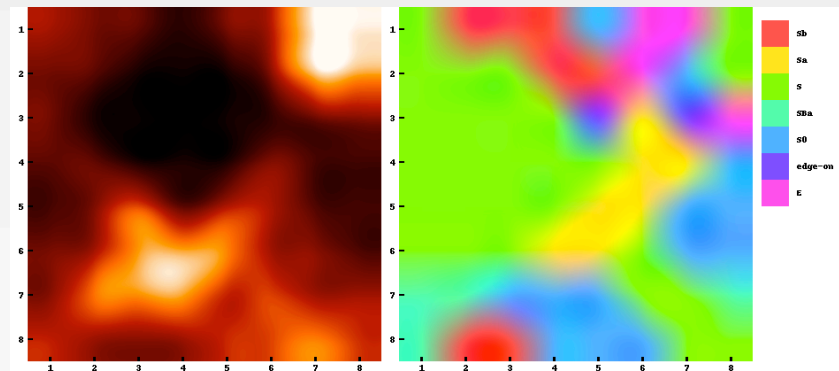
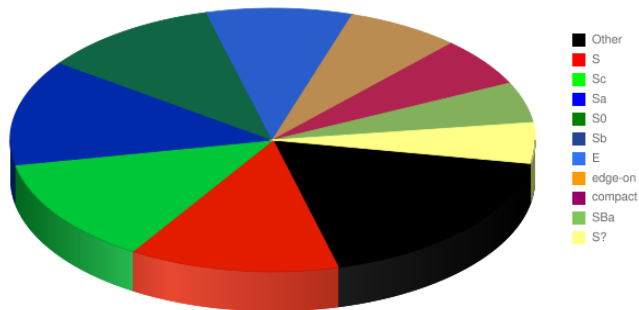


Incorporating knowledge

- 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

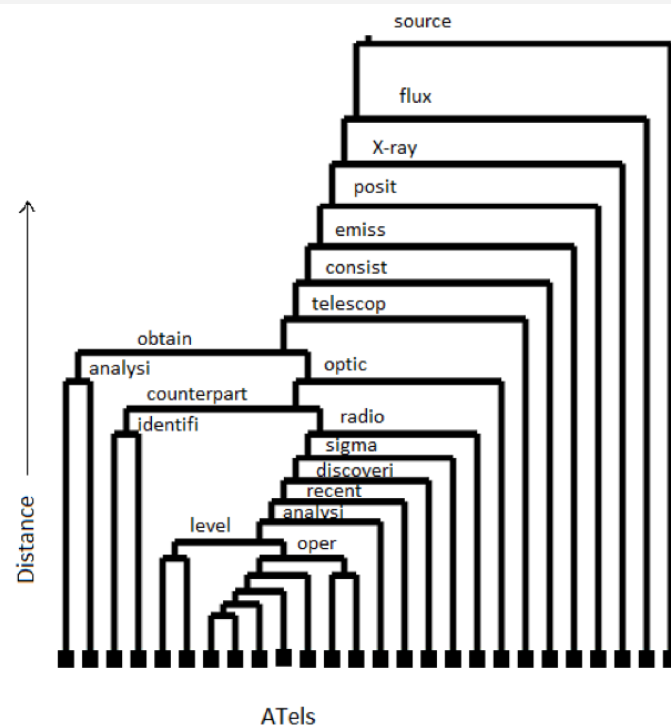


[...  ]                                



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

