

How to maintain intelligibility of scientific archives using ontology and text ?

Bernard Rothenburger (Inria/Irit)

with

Nathalie Aussenac-Gilles (Cnrs/Irit)

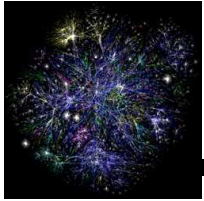
Nathalie Hernandez (Irit)

Nacim Chiki (Irit)



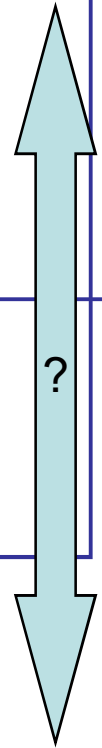
21 feb 2008

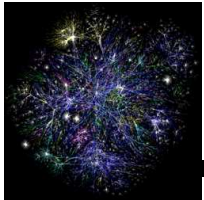
Practical Semantic Astronomy



Problem sketch

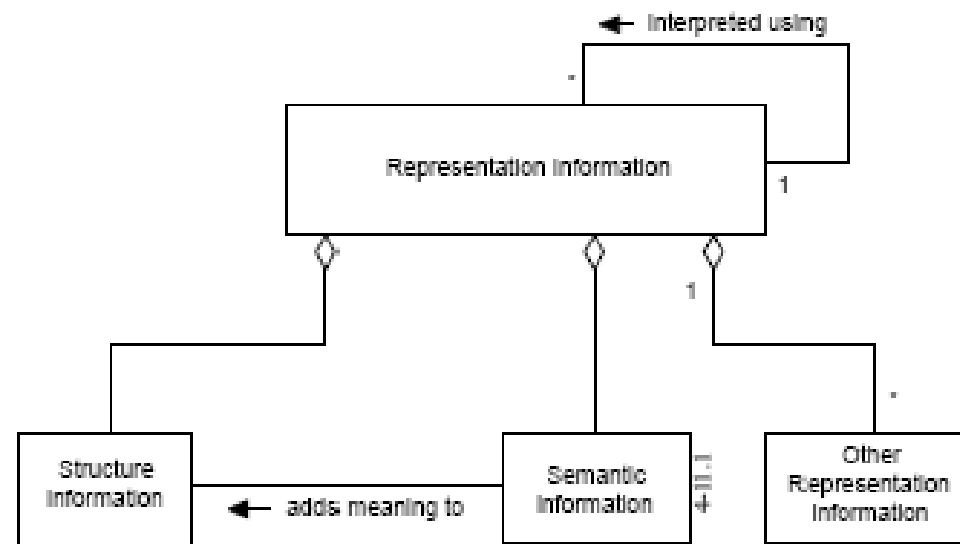
- 1993 : a space probe was designed.
(Albert finish his high school)
- 2000 : construction of the probe
(Albert does his Phd and joins the CNES)
- 2004 : probe is launched
- 2014 : the probe reach a comet, scientific experiments are carried on
(Albert contributes to the experiments and must understand technical documents written by others 21 years ago)
- 2016 : end of the life of the probe
(Albert' son finish his high school)
- 2036 : experiment results are always used by scientists
(Albert' son who is a biologist uses results of an experiment described for chemists 22 (43) years ago)
-Albert' grandson finish his high school ...





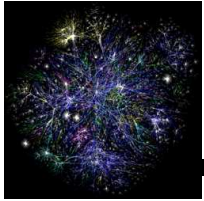
OAIS

- “An OAIS archive is one that intends to preserve information for access and use by a **Designated Community, ...**”(ccsds, ISO14721)
- “... ensure that the information to be preserved is **Independently Understandable...**”



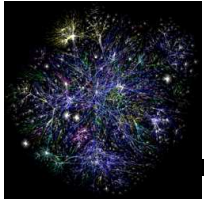
21 feb 2008

Figure 4-11: Representation Information Object



OAIS

- “The information being maintained has been deemed to need **Long Term Preservation**, ...**Long Term** is long enough to be **concerned** with the impacts of changing technologies, including support for new media and data formats, or with a **changing user community**.”
- “The Designated Community should be **monitored** to be sure the Content Information is still understandable to them. The Designated Community may lose its familiarity with some terminology, and the definition of the **community may be broadened to include other members with different backgrounds**. For example, a **periodic review** with participants representing the Designated Community could assist in this process.”

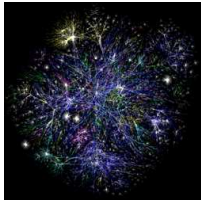


An example

Semantic misunderstanding:

- E.g. sunspot number and variations in solar radiation: over **90% of researchers outside the sub-field of solar radiation think: sunspot number is a measure of solar radiation**
- **In reality: a sunspot number is a measure (10 x #groups + individual spots) of the number of sunspots appearing on the visible solar surface. It is correlated to, but not a measure of, solar radiation.**
- **Why does this matter? Low sunspot number does NOT equal low solar radiation...**
- **How to 'explain' this to a computer?**

Peter Fox, James Benedict, Deborah McGuinness
2006 Semantic Technology Conference
San Jose, California March 6-9, 2006




hypothesis

A large part of d

PLASMA PHYSICS DATA CENTRE - Collection descriptio

http://cdpp2.cnes.fr/cdpp/showDataSetPage

PLASMA PHYSICS DATA CENTRE - Collection des



**CEN
PH**

You are here

Data

- Choose Data
 - By navigation
 - By keywords and criteria
 - By quicklooks
- My Selection
 - View selection
 - Reset selection
- My workspace

Services

- Documents
- Event tables
- Orbit viewer

User

- Authenticate

Descript

Ulysses :

Ulysses is a
Launched by
a gravity ass
pole of the
completed,
is now emb

The spacec

http://cdsads.u-strasbg.fr/cgi-bin/nph-iarticle_query?1992A%26A5...92..207W&data_type=PDF_H...

http://cdsads.u-strasbg.fr/cgi-bin/nph-iarticle_query?1992A%26A5...92..207W&data_type=...

http://cdsads.u-strasbg.fr/cgi-bin/nph...

Fichier >> Google G

Go

Bookmarks >>

Settings >

1 / 13

65%

mission

ASTRONOMY & ASTROPHYSICS
SUPPLEMENT SERIES

Astron. Astrophys. Suppl. Ser. **92**, 207-219 (1992)

The Ulysses Mission

K.-P. Wenzel¹, R.G. Marsden¹, D.E. Page² and E.J. Smith³

¹ Space Science Department of ESA, ESTEC, P.O. Box 299, 2200 AG Noordwijk, The Netherlands

² ESA, Ulysses Science Coordinator in residence at Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

³ Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

Received March 28, 1991; accepted September 4, 1991

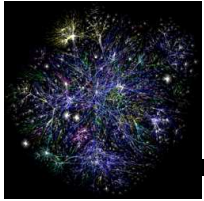
Abstract. — The Ulysses mission is unique in the history of the exploration of our solar system by spacecraft. The path followed by Ulysses will enable us, for the first time, to explore the heliosphere within a few astronomical units of the Sun over the full range of heliographic latitudes, thereby providing the first characterisation of the uncharted third heliospheric dimension. Advanced scientific instrumentation carried on board the spacecraft is designed to measure the properties of the heliospheric magnetic field, the solar wind, the Sun/wind interface, solar radio bursts and plasma waves, solar energetic particles and galactic cosmic rays, solar X-rays, and interplanetary/interstellar neutral gas and dust. Ulysses will also be used to detect cosmic gamma-ray bursts and search for gravitational waves. The mission, a collaboration between ESA and NASA, was launched in October 1990 and employs a Jupiter gravity-assist to achieve the trajectory extending to high solar latitudes. Ulysses will spend a total of 234 days, equivalent to about 8 solar rotations, at latitudes in excess of 70°. The purpose of this paper is to describe the characteristics of the Ulysses mission in order to establish a framework within which to better understand the objectives and goals of the scientific investigations which are described in subsequent papers.

Key words: interplanetary medium — solar system: general — space vehicles — Sun (the): general.

1. Introduction.

Despite the general acceptance that the heliosphere - the vast region of space dominated by the radial outflow of

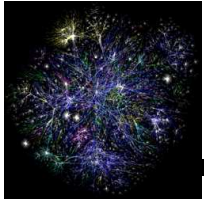
tion and Lyman alpha scattering, charged particle gradients between Voyager 1 and the ecliptic, a warped current sheet varying with sunspot cycle, the inability to explain cosmic ray modulation by measurements made in



hypothesis

A large part of description of data (cf OAIS)
have a textual form

Data intelligibility mainly depends on this textual
descriptions



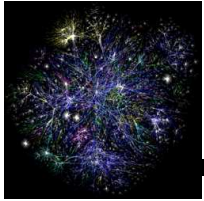
strategy

Based on Knowledge engineering tools

Preventive : audit > diagnostic > correction

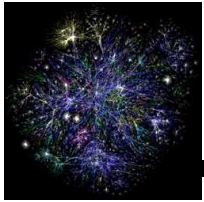
two frameworks :

- new community apparition (I)
- are knowledge used for data description these of the different (even new) communities (II)



(I) Community mining and link analysis

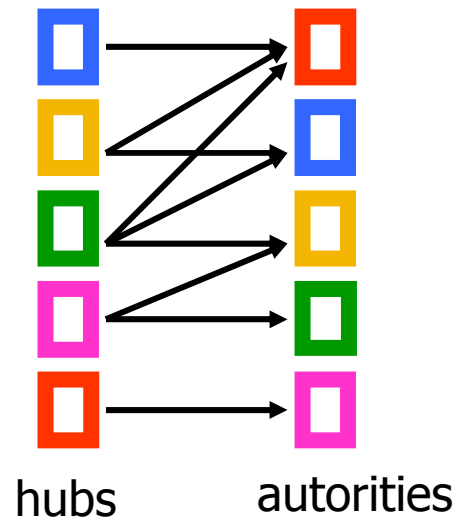
- **Epistemic community** : group of agents sharing a common set of subjects, concepts, problems; sharing a common aim of knowledge creation
(Haas (1992), Cowan et al. (2000), Dupouet et al. (2001)).
- A **(web) community**, is a [group](#) of people that primarily interact via web communication media such as [email](#), [Usenet](#), [forum](#), ...rather than face to face
- A **(web) community**, is a set of pages web pages (documents) having more [links to members](#) web pages (documents) than to web pages (documents) outside of the community
- **Explicit** community vs **implicit** community

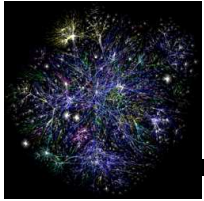


(I) new community apparition: from HITS to NHITS

- Basic idea : **authorities**
(cf PageRank) but also **hubs**
- Depending on an initial query (expl :
“virtual observatory”, “space plasma”,
“space plasma” and magntosphehre)
- Output : a set of authorities and hubs
(pages or documents) which
represent the whole community.

$$h_i = \sum_{j:i \rightarrow j} a_j \quad a_i = \sum_{j:j \rightarrow i} h_j$$





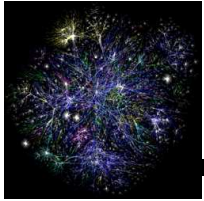
(I) new community apparition from HITS to NHITS

- Idea: Use of non principal eigenvectors in order to identify others community
- Output : sets of pages (documents) corresponding to the authorities of each community
- Limitation: « interpretability » (vectors have positive and negative ends, positive or negative may (or may not) be a community)
- Idea: We use Non Negative Factorisation (NMF)
- Output: sets of pages (documents) corresponding to the main communities with

membership (the certainty that this document belongs to the community)

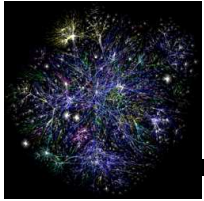
representativeness (the authority of the document for the community)

overlapping community possible



(I) NHITS: experiment

- We get 5572 articles from ADS using the query : “space plasma”,
- There where 41879 citations
- We built an adjacency matrix
- We identify the 10 first communities



(I) NHITS: experiment

Community 1 (**Polar cap, [theta] aurora, IMF,...**)

0.031454 The **theta aurora**

0.024609 **Polar cap arcs** and the plasma sheet

0.022961 **Polar** views of the earth's **aurora** with Dynamics Explorer

0.018686 Origin of the plasma in a **cross-polar cap** auroral feature (**theta aurora**)

0.017095 Extremely high latitude auroras

0.01332 Interplanetary magnetic field dependency of stable Sun-aligned **polar cap arcs**

0.012685 **IMF** B(y)-dependent plasma flow and Birkeland currents in the dayside magnetosphere. II - A global model for northward and southward **IMF**

0.010855 Electric and magnetic field characteristics of discrete arcs in the **polar cap**

0.010677 B(y)-dependent convection patterns during northward interplanetary magnetic field

0.010644 Simultaneous optical observations of transpolar arcs in the two **polar caps**

0.010473 A classification of polar **cap auroral arcs**

0.010437 DMSP optical and electron measurements in the vicinity of **polar cap arcs**

0.0091752 A comparison of a model for the **theta aurora** with observations from Polar, Wind, and SuperDARN

0.0089724 Transpolar **auroras**, their particle precipitation, and **IMF** B(y) component

0.0086518 **IMF** By-dependent plasma flow and Birkeland currents in the dayside magnetosphere. I - Dynamics Explorer observations

0.0084544 Large-scale Birkeland currents in the dayside polar region during strongly northward **IMF** - A new Birkeland current system

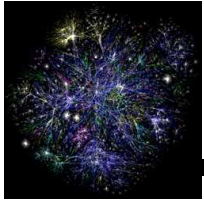
0.0083218 Evolution of the global **aurora** during positive **IMF** B_z and varying **IMF** B_y conditions

0.0072503 Creation of **Theta-Auroras** - the Isolation of Plasma Sheet Fragments in the **Polar CAP**

0.0069344 A simple model for **polar cap** convection patterns and generation of **theta auroras**

21 Feb 2008

Practical Semantic Astronomy



(I) NHITS: l'expérimentation

community 4 (**dusty [acoustic] plasma...**)

0.047304 **Dust**-acoustic waves in dusty plasmas

0.039622 **Dusty** plasmas in the solar system

0.026847 Cosmic **Dusty** Plasmas

0.024998 **Dust** ion-acoustic wave

0.024247 The electrostatics of a **dusty** plasma

0.021562 Ion plasma waves in **dusty** plasmas - Halley's comet

0.019689 Low-frequency electrostatic waves in **dusty** plasmas

0.018189 Ion- and dust-acoustic instabilities in **dusty** plasmas

0.015564 A new damping effect for the **dust**-acoustic wave

0.01547 Waves and Instabilities in **Dusty** Space Plasmas

0.012191 Waves in **Dusty** Space Plasmas

0.011748 Linear and nonlinear **dust** drift waves

0.010704 A kinetic model for **dust** acoustic waves applied to planetary rings

0.010231 Collective effects due to charge-fluctuation dynamics in a **dusty** plasma

0.0086977 **Dusty** plasmas.

0.0082804 Large amplitude ion-acoustic solitons in a **dusty** plasma

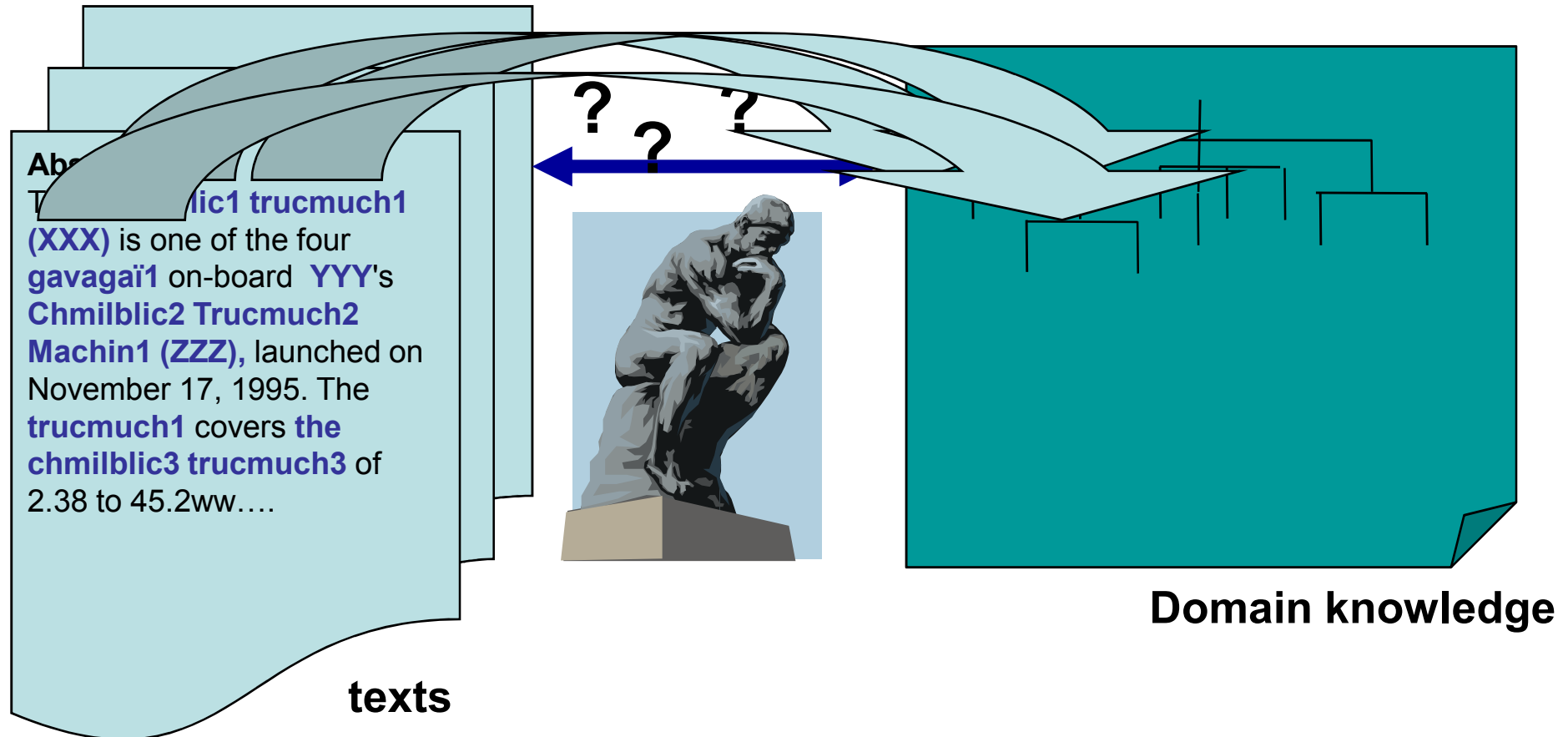
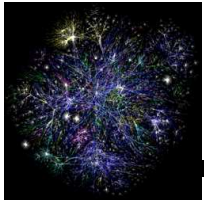
0.0075366 Dispersion properties of **dusty** plasmas

0.0072596 A survey of **dusty** plasma physics

0.0068176 **Dust** charges, cloud potential, and instabilities in a dust cloud embedded in a plasma

0.0063939 Charging of **dust** grains in a plasma

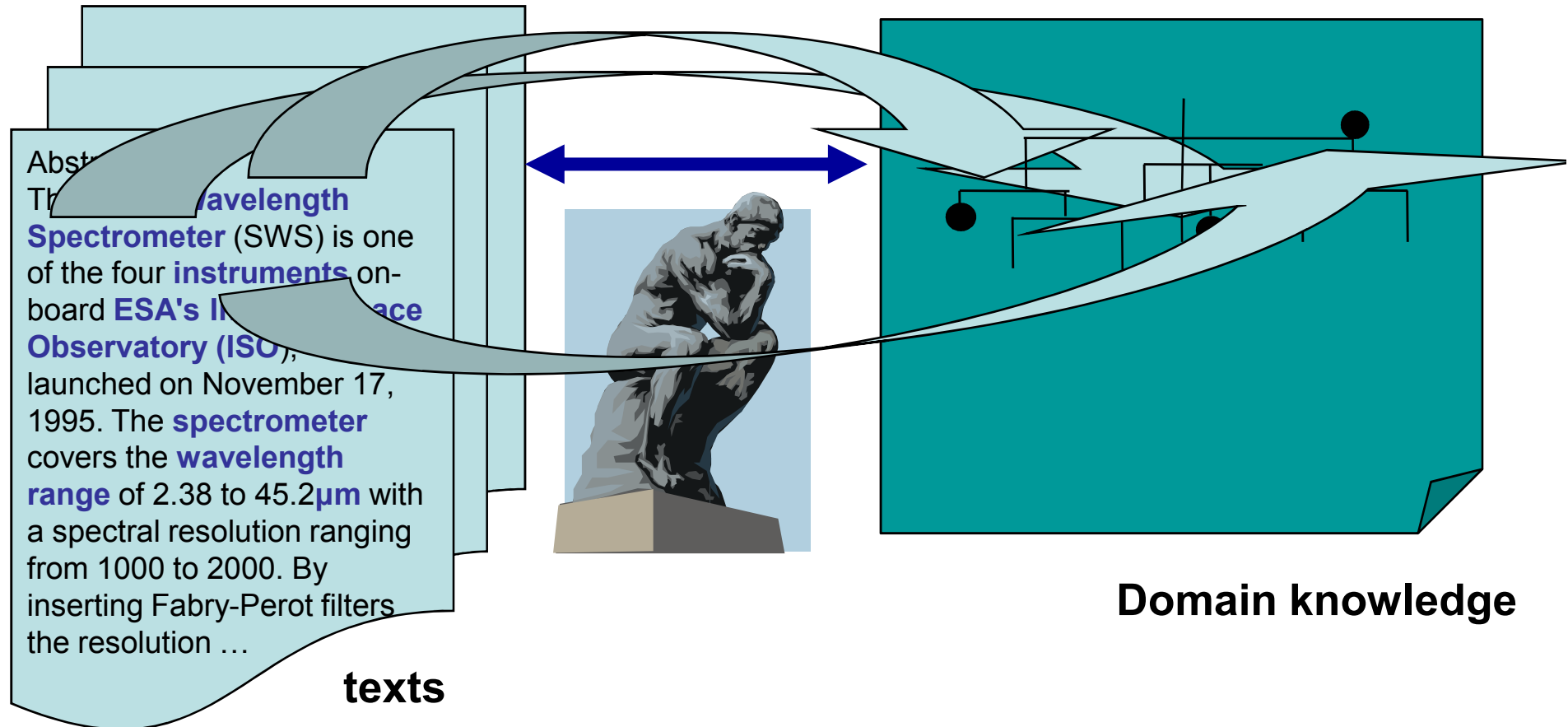
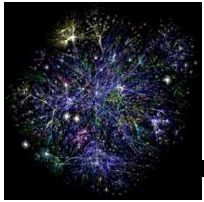
(II) Archives Intelligibility : hypothesis



21 feb 2008

Practical Semantic Astronomy

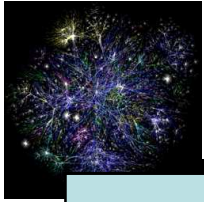
(II) Archives Intelligibility : hypothesis



21 feb 2008

Practical Semantic Astronomy

(II) Archives Intelligibility : hypothesis



Abstract

The **Short-Wavelength Spectrometer (SWS)** is one of the four instruments on-board ESA's Infrared Space Observatory (ISO), launched on November 17, 1995. **The spectrometer covers the wavelength** range of 2.38 to 45.2 μm with a spectral resolution ranging from 1000 to 2000. By inserting Fabry-Perot filters the resolution can be enhanced by a factor 20 for the wavelength range from 11.4 to 44.5 μm . An overview is given of the instrument, its in-orbit calibration, performance, observing modes and off-line processing software.



wavelengths

DEF Distance in the direction of propagation of a periodic wave between two successive points at which the phase is the same (at the same time).

GS wavelengths

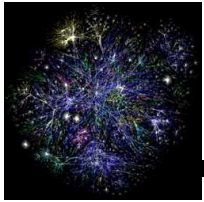
- . de Broglie wavelengths
- RT antinodes
 - harmonics
 - infrared radiation
 - laser modes
 - laser outputs
 - longitudinal waves
 - maser outputs
 - millimeter waves
 - nodes (standing waves)
 - spectral emission
 - standing waves
- Stokes law of radiation
- submillimeter waves
- whispering gallery modes

21 feb 2008

Practical Semantic Astro

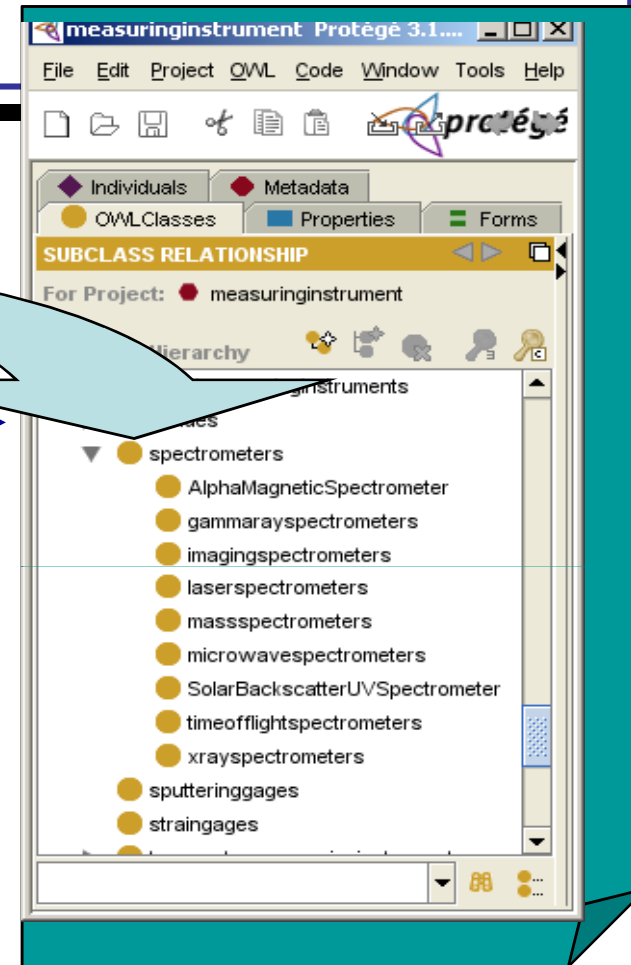
Terminological and ontological resources

(II) Archives Intelligibility : hypothesis



Abstract

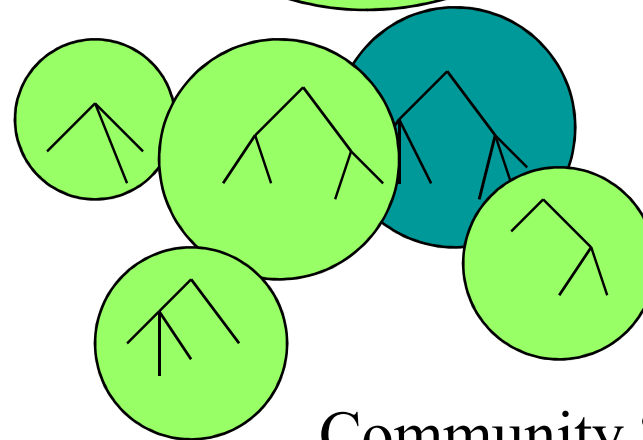
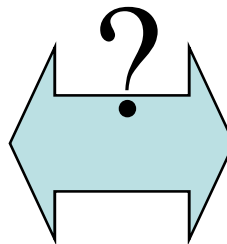
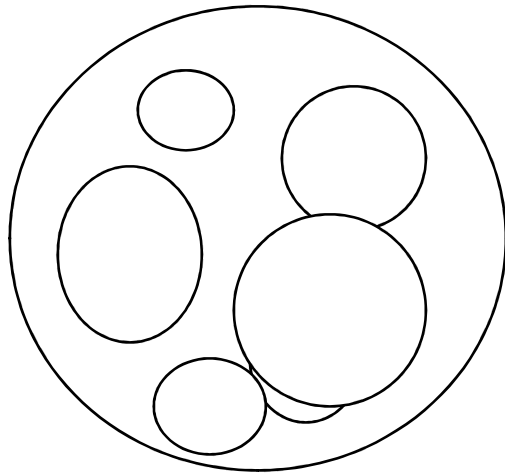
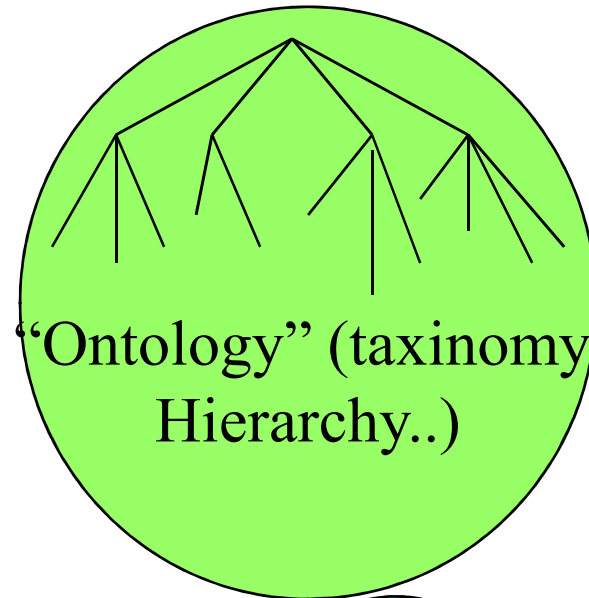
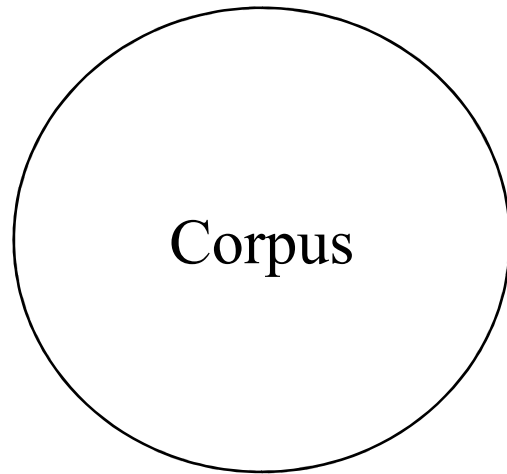
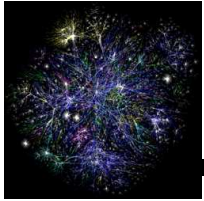
The **Short wavelength Spectrometer (SWS)** is one of the four instruments on-board ESA's Infrared Space Observatory (ISO), launched on November 17, 1995. **The spectrometer covers the wavelength** range of 2.38 to 45.2 μm with a spectral resolution ranging from 1000 to 2000. By inserting Fabry-Perot filters the resolution can be enhanced by a factor 20 for the wavelength range from 11.4 to 44.5 μm . An overview is given of the instrument, its in-orbit calibration, performance, observing modes and off-line processing software.

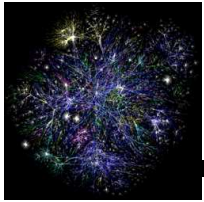


Terminological and ontological resources

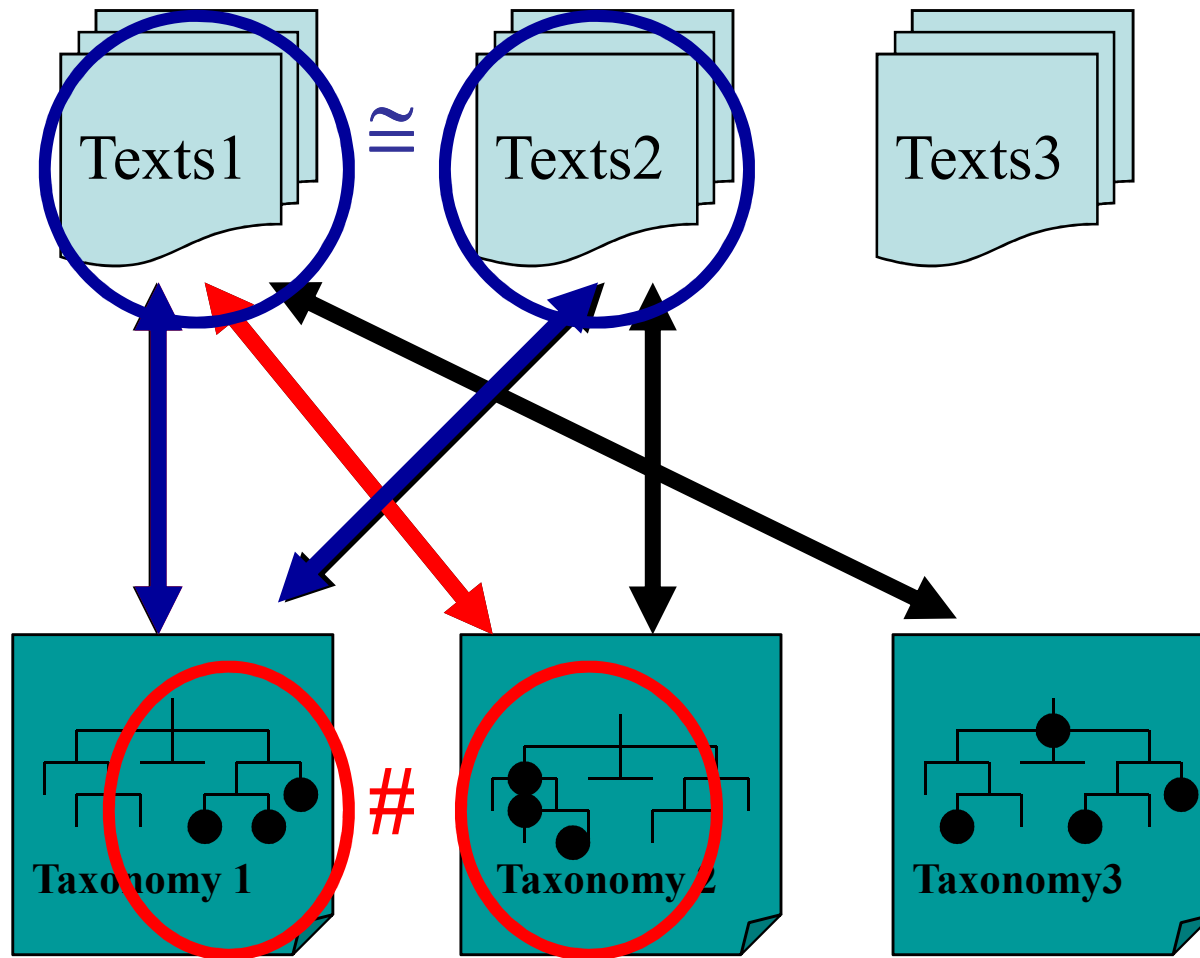
Practical Semantic Astronomy

(II) Archives Intelligibility : principle



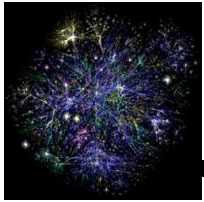


'II) Archives Intelligibility : principle: hitting the hierarchy



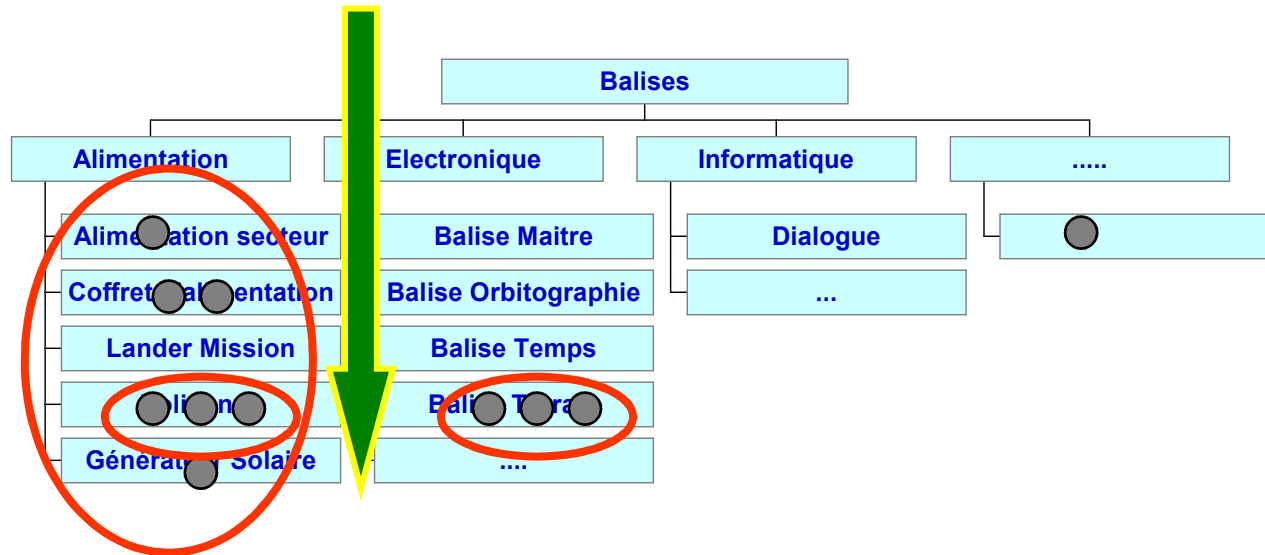
21 feb 2008

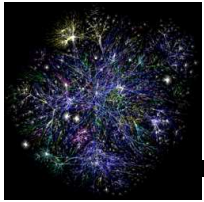
Practical Semantic Astronomy



(II) Archives Intelligibility : principle

Analysis profile : contribution, **spécificity**, **concentration**

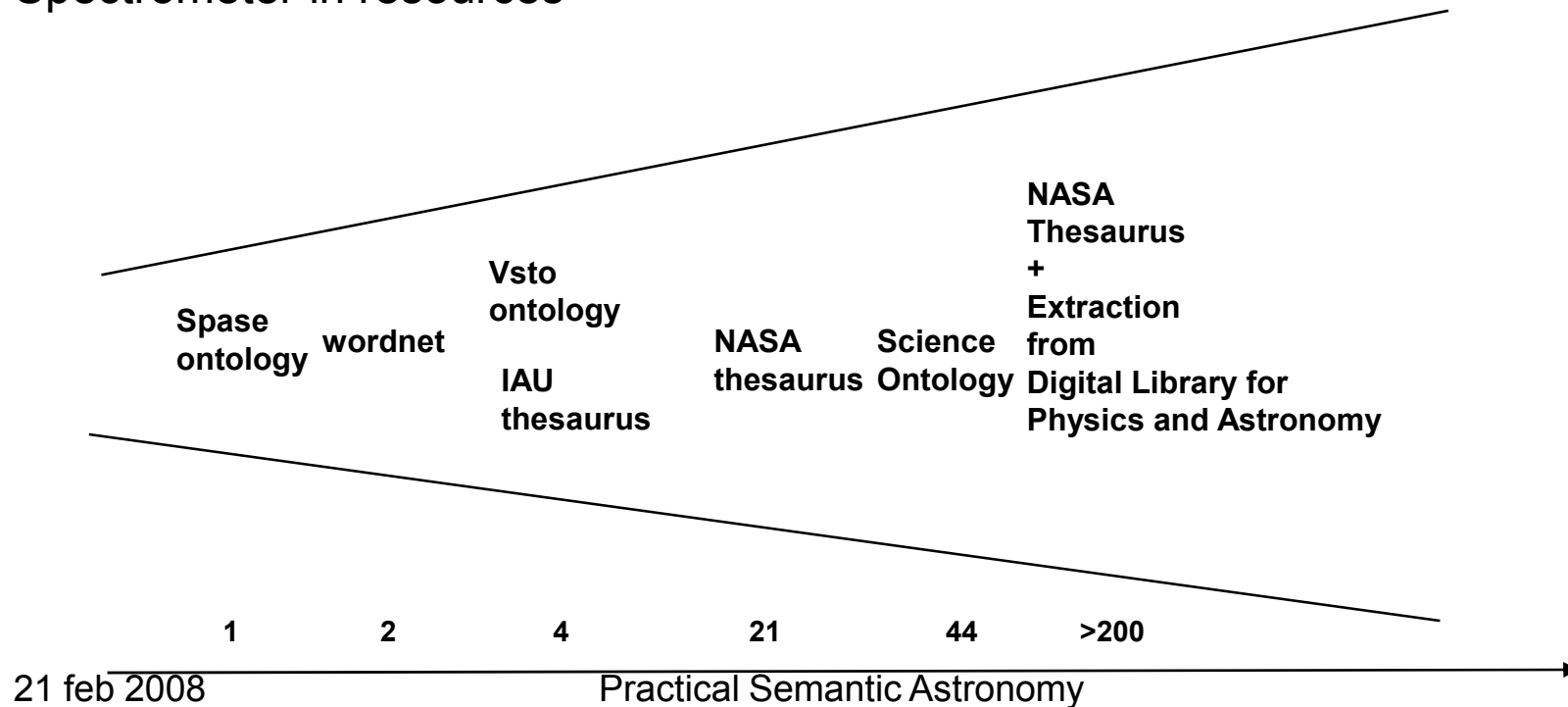




(II) Archives Intelligibility : principle: choosing the “ontology”

Improve coverage
Release consistency

Spectrometer in resources





(II) Archives Intelligibility : principle: choosing the “ontology”

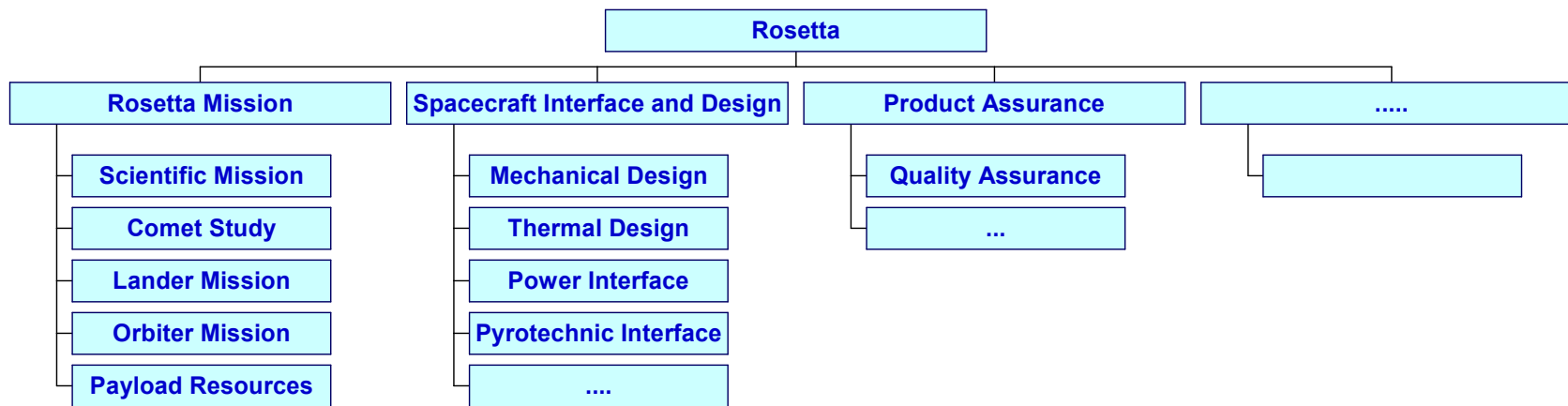
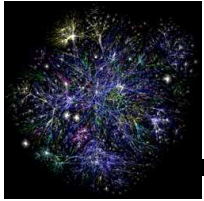
measuring instruments

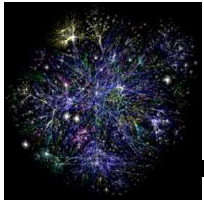
. spectrometers

- . . achromatic spectrometer
- . . Alpha Magnetic Spectrometer
- . . Crystal Spectrometer
- . . . High Resolution Crystal Spectrometer
- . . . single crystal spectrometer
- . . Ebert spectrometers
- . . electron spectrometer
- . . . multidetector electron spectrometer
- . . Fabry-Perot spectrometers
- . . Fourier Transform Spectrometer
- . . . Fourier-transform microwave spectrometer
- . . . high resolution Fourier Spectrometer
- . . . pulsed beam microwave Fourier transform spectrometer
- . . gamma ray spectrometers
- . . . gamma-ray and neutron spectrometers
- . . . high sensitivity beta-gamma spectrometer
- . . . Transient Gamma Ray Spectrometer
- . . Grating Spectrometer *
- . . . Low-Energy Transmission Grating Spectrometer
- . . . High Energy Transmission Grating Spectrometer
- . . . High and Medium Energy Transmission Grating Spectrometer
- . . . objective grating spectrometer
- . . imaging spectrometers
- . . . Gas Imaging Spectrometer
- . . . Low-Resolution Imaging Spectrometer
- . . infrared spectrometers
- . . . Mid-Infrared Spectrometer
- . . . near infrared imaging spectrometer
- . . . near-infrared mapping spectrometer
- . . . filter wheel infrared spectrometers

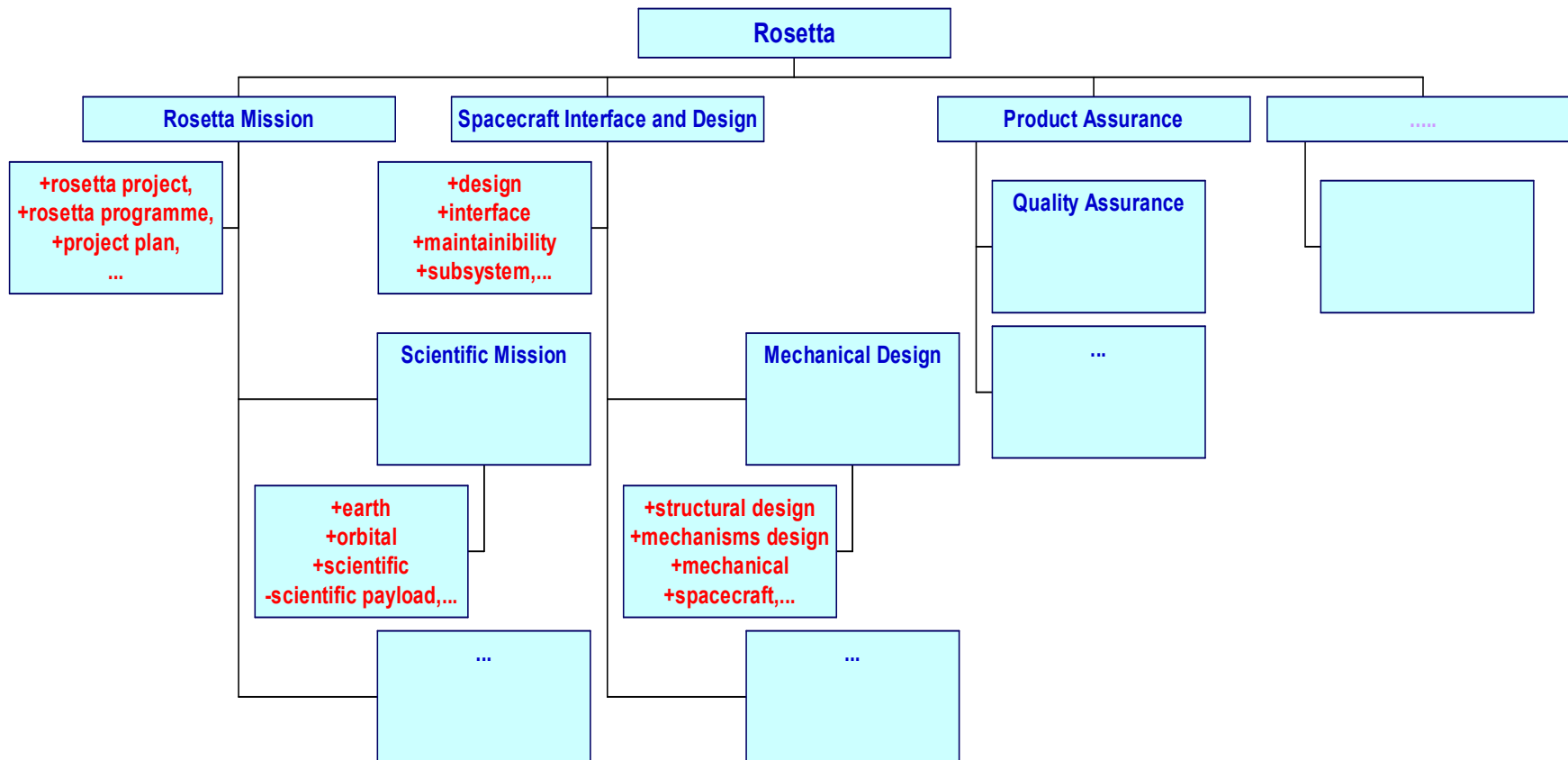
- . . ion spectrometer
- . . grating spectrometer
- . . laser spectrometers
- . . . diode laser spectrometer
- in situ sensor tunable diode laser spectrometer
- . . mass spectrometers
- . . . geomagnetic mass spectrometer
- . . . static mass spectrometer
- . . . on-line mass spectrometer
- . . . retarding ion mass spectrometer
- . . microwave spectrometers
- . . multichannel spectrometer
- . . . two-channel spectrometer
- . . . three-channel spectrometer
- . . neutron spectrometers
- . . plasma spectrometer
- . . Spectroheliographs
- . . Solar Backscatter UV Spectrometer
- . . solar spectrometers
- . . . Computer-Controlled Solar Radio Spectrometer
- . . time of flight spectrometers
- . . . high-resolution Mass Time-of-Flight (MTOF) spectrometer
- . . ultraviolet spectrometers (UV spectrometer)
- . . . Vacuum-Ultraviolet Spectrometer
- . . . limb scanning ultraviolet spectrometer
- . . . high dispersion spectrographs
- . . . Total Ozone Mapping Spectrometer
- . . visible spectrometer
- . . x ray spectrometers
- . . . Double X-Ray Spectrometer
- double crystal x-ray spectrometer
- hard X-ray burst spectrometer

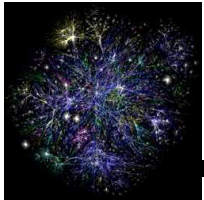
(II) Archives Intelligibility : principle: reference ontology



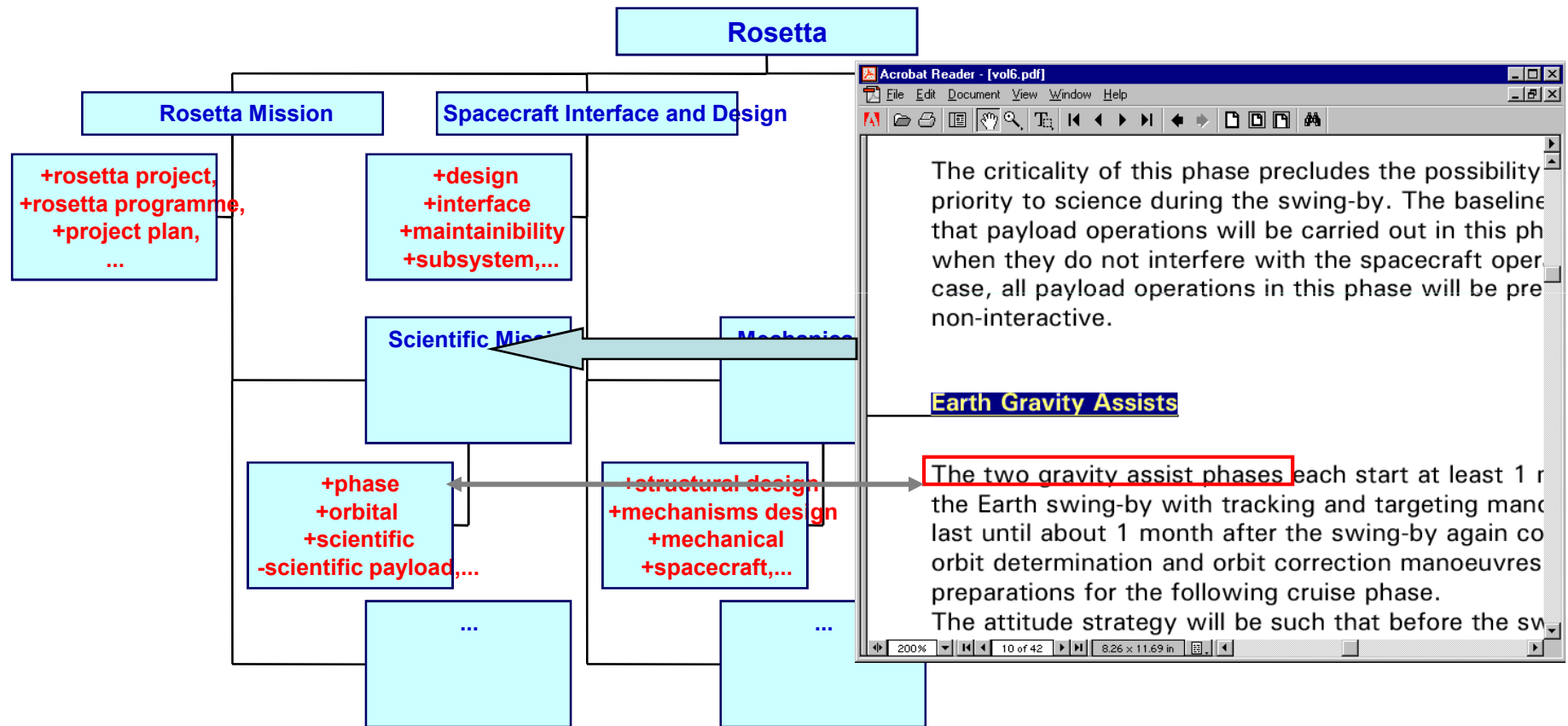


(II) Archives Intelligibility : principle: analyzing ontology



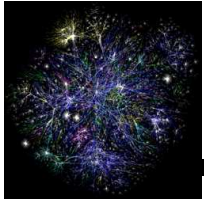


(II) Archives Intelligibility : principle: loaded ontology



21 feb 2008

Practical Semantic Astronomy



Archives Intelligibility : conclusion

Two step process

Identifying new community (quite good)

Description intelligibility (in test for this case)

To do:

Better metrics

Combining, comparing links and contents

More experiment

Fundamental issue:

Social network vs semantic network