

Knowledge bases in astronomy

The example of the Astronomy IAU Thesaurus

N. Hernandez, J. Mothe (IRIT)

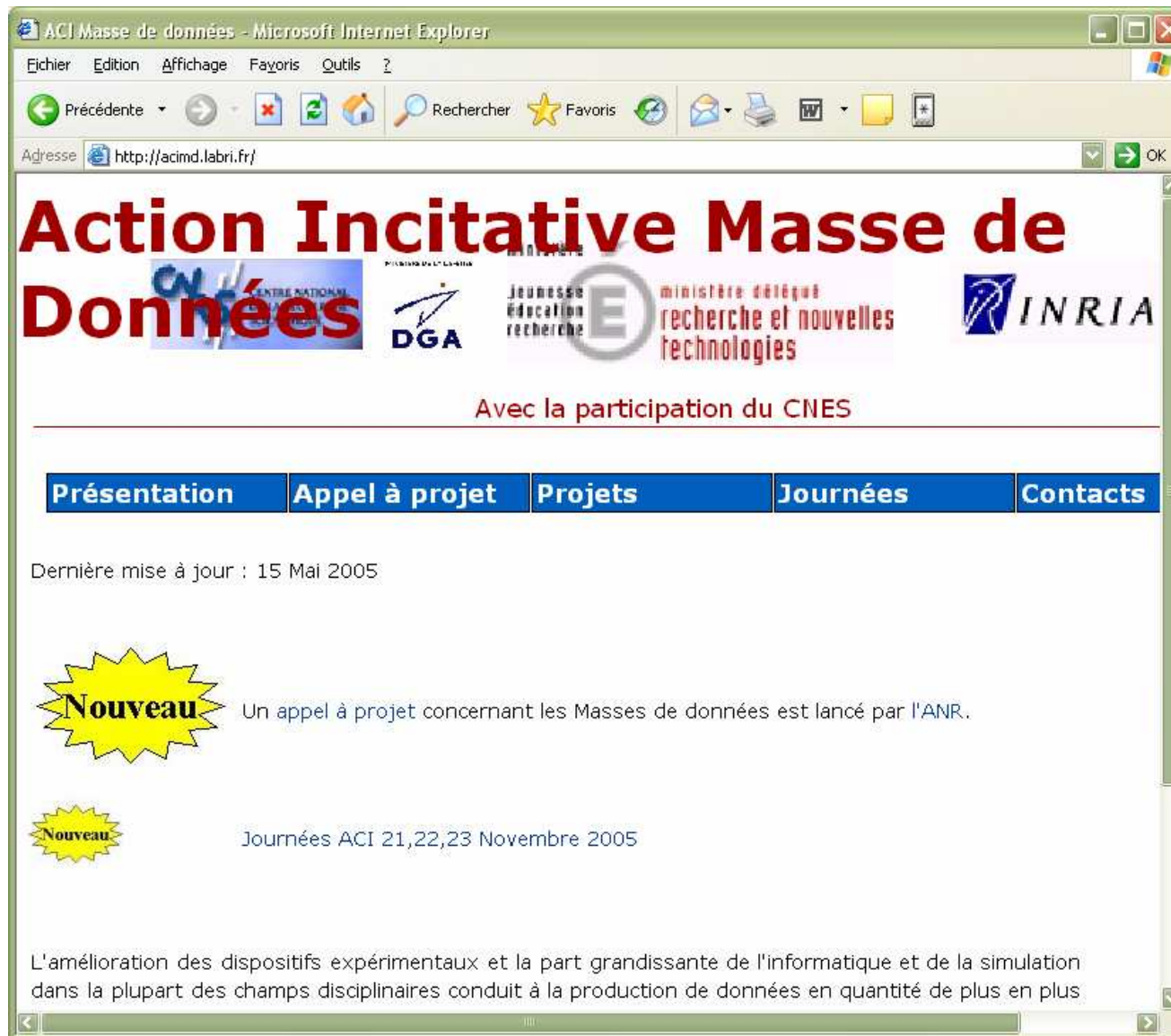
P. Dubois, S. Lesteven,

F. Genova, S. Derriere (CDS)

A. Preite Martinez (INAF)



- Collaboration between CDS and two French IT labs on R&D on ontologies in the frame of the *Massive Data in Astronomy* project
- With LORIA (Nancy): prototype ontology of UCDs
- With IRIT (Toulouse): work on the IAU Thesaurus. The Thesaurus was finalized in 1995 – one problem: how do we update it
- *Work in progress*



The Astronomy Thesaurus - Microsoft Internet Explorer

ANU Home | Search ANU | Directories

ANU
THE AUSTRALIAN NATIONAL UNIVERSITY

Research School of Astronomy and Astrophysics
Mount Stromlo and Siding Spring Observatories



Research

The Astronomy Thesaurus

- Compiled by Robyn M. Shobbrook ([AAO](#)) & Robert R. Shobbrook (Syd. Uni.)
- Web interface by David Fullagar (RSAA)
- [Acknowledgements](#)
- [Help!](#) (A User's Manual)
- Click on a Flag (or language name) to begin

Note: this is a **thesaurus**, not a dictionary. You will not find definitions for the words within.

Index and Search

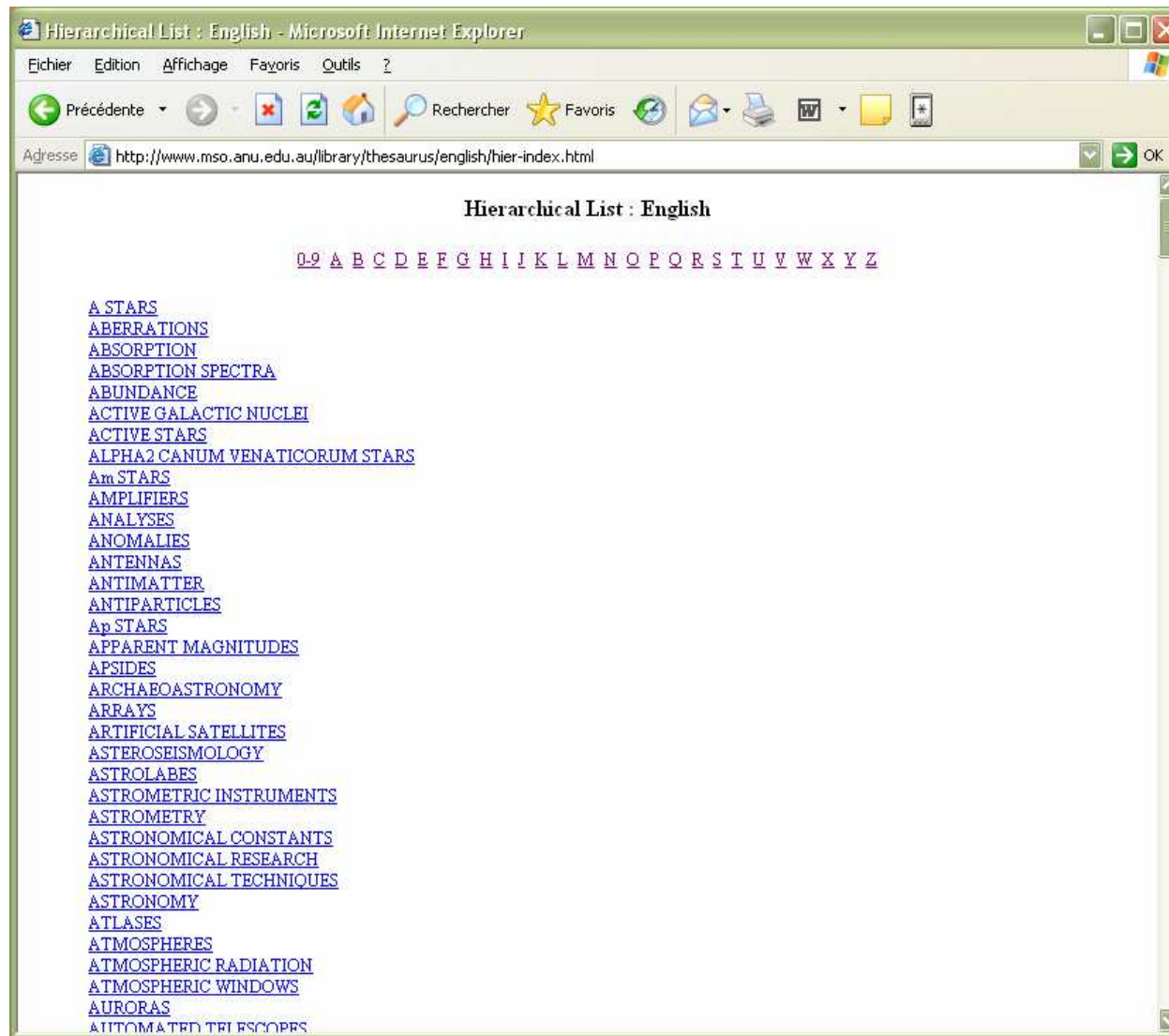



© 1992,1995 - International Astronomical Union
No part of this work may be published, changed or copied without the prior consent of the authors.

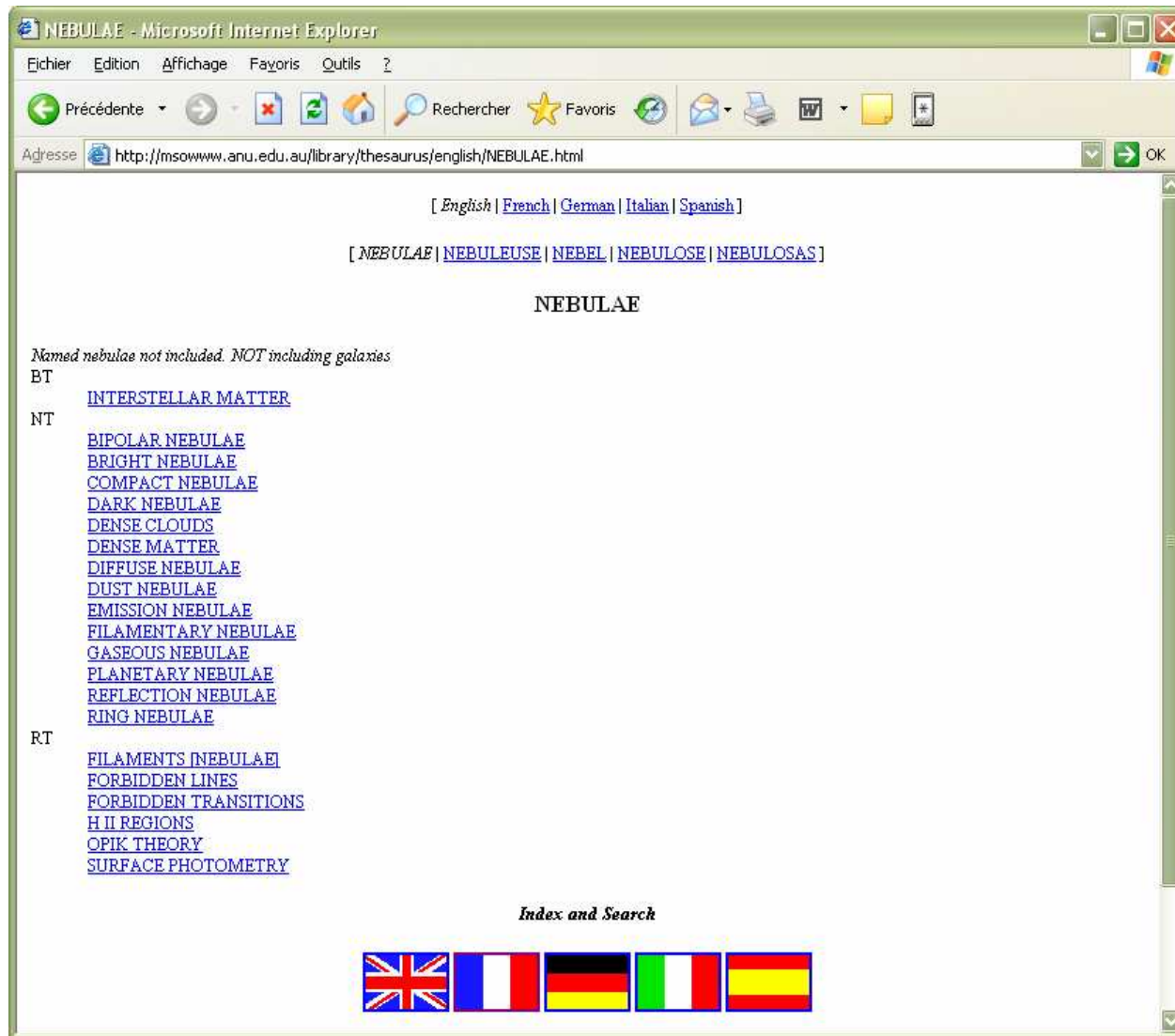
The Astronomy Thesaurus has not been updated since 1992/5 and there are not plans to do so in the foreseeable future.

Search RSAA

<http://www.mso.anu.edu.au/library/thesaurus/>



Relations



Hierarchy

Hierarchical List : NEBULAE - Microsoft Internet Explorer

Eichier Edition Affichage Favoris Outils ?

Précédente Rechercher Favoris

Adresse <http://msowwww.anu.edu.au/library/thesaurus/english/hier-NEBULAE.html> OK

[[English](#) | [French](#) | [German](#) | [Italian](#) | [Spanish](#)]

[[NEBULAE](#) | [NEBULEUSE](#) | [NEBEL](#) | [NEBULOSE](#) | [NEBULOSAS](#)]

Hierarchical List : [NEBULAE](#)

- [BIPOLAR NEBULAE](#)
- [BRIGHT NEBULAE](#)
- [COMPACT NEBULAE](#)
- [DARK NEBULAE](#)
- [BOK GLOBULES](#)
- [DENSE CLOUDS](#)
- [DENSE MATTER](#)
- [DIFFUSE NEBULAE](#)
- [DUST NEBULAE](#)
- [EMISSION NEBULAE](#)
- [H II REGIONS](#)
- [COMPACT H II REGIONS](#)
- [HERBIG HARO OBJECTS](#)
- [PLANETARY NEBULAE](#)
- [FILAMENTARY NEBULAE](#)
- [GASEOUS NEBULAE](#)
- [EMISSION NEBULAE](#)
- [H II REGIONS](#)
- [COMPACT H II REGIONS](#)
- [HERBIG HARO OBJECTS](#)
- [PLANETARY NEBULAE](#)
- [PLANETARY NEBULAE](#)
- [REFLECTION NEBULAE](#)
- [RING NEBULAE](#)

Index and Search

Astronomy Thesaurus Introduction - Microsoft Internet Explorer

Eichier Edition Affichage Favoris Outils ?

Précédente Recherche Favoris

Adresse <http://www.mso.anu.edu.au/library/thesaurus/introduction.html> OK

see under 'Use' below.

SCOPE NOTE (SN) or message.
 These notes are in normal text, except that references to other PTs are in upper case. The notes have been used for the following purposes:

- a) To instruct in the use of a term for indexing and retrieval, e.g. ENERGY - (SN) Combine with other terms
- b) To define and to limit use, e.g. GALACTIC - (SN) Used as an adjective only for Milky Way properties.
- c) To explain a principle of division (different meanings for the same word), e.g. ABERRATION - (SN) Of starlight; ABERRATIONS - (SN) Of optical images.

U: USE.
 Directs users to a preferred term. The term following U will always be in UPPER CASE. The primary term is in bold type - as for all PTs - but in lower case. The user should turn to the given (U) upper case term to find the BTs, NTs and RTs (see below).

Note that the lower case terms, since they are often perfectly acceptable working alternatives to the upper case terms, may well be used as search terms for references to older publications or data. They are included because they were in fairly common use at some time, or by some authors. A typical example is BETA CEPHEI STARS. The currently accepted term is as just stated, but before about 1980 they were often called Beta Canis Majoris stars; many publications concerning these stars will not be found using only the first term.

UF: USE FOR.
 The inverse of U(se). Terms - always in lower case - under the (upper case) primary term are those for which the primary term is to be substituted.

BT: BROADER TERM.
 Indicates to which class or genus the term belongs. A term may belong to more than one class and thus may have more than one broader term, e.g.

B STARS are both EARLY TYPE STARS and YOUNG STARS.

NT: NARROWER TERM.
 Terms which are sub-categories of the main term.

RT: RELATED TERM.
 Directs a user to a selection of other terms which are closely related or associated with the PT but are not sub-categories of it. The RT list will often indicate other terms which might be used instead of, or in conjunction with, the PT.

The U/UF and the BT/NT relationships are reciprocal. In the LEXICON software, when 'TERM1' is entered as an NT for 'TERM2', then 'TERM2' is automatically listed as a BT for 'TERM1'. Similarly, the RT relationship is reciprocal; if TERM1 is entered as an RT for TERM2, then TERM2 is automatically listed amongst the RTs of TERM1.

From a thesaurus to an ontology (1)

- Thesauri gather « terms » and relationships between them:
 - « used for » ;
 - « more generic / more specific » ;
 - « is-related to »
 - Time and expertise required to build them
- Ontologies gather « concepts » (and their labels = terms) and relationships between them
 - « is-a »
 - Semantics between relationships (« is part of » « measures » « phenomenon of » , etc.)
 - Formal (OWL) :
 - validation (no contradiction, etc..)
 - and used to infer new knowledge
- Both are built at one point ; update needed

From a thesaurus to an ontology (2)

- Transform existing thesauri into ontology
 - The knowledge it contains
 - Specifying it if necessary
 - Additional knowledge extracted from publications
- On going work (MDA-IRIT)
 - Methodology to formalize existing knowledge in OWL and formal ontology
 - Disambiguate some relationships
 - « is a » « related to »
 - Update with new terms and new relationships

From a thesaurus to an ontology (3)

- IAU:
 - « used for », « preferred » « more generic / more specific » : direct translation
 - **New concepts:** syntactic analysis of a corpus and weighting terms (concept labels) importance in the corpus
 - Corpus: A&A 1995, A&A 2002

column density ; high resolution ; globular cluster ; white dwarf ; soft X-ray ; time scale ; orbital period ; power law ; absorption line ; line emission ; active region

From IAU Thesaurus to an ontology (4)

- « generic/specific »: too many entries at the higher level of the Thesaurus.

Solution: Abstract level from a generic resource
(Wordnet) validated by astronomers

Property : a basic or essential attribute shared by all members of a class

Phenomenon : any state or process known through the senses rather than by intuition or reasoning

Event : something that happens at a given place and time

Instrumentation : an artifact (or system of artifacts) that is instrumental in accomplishing some end

Object : a tangible and visible entity

Natural object : an object occurring naturally; not made by man

Etc



From IAU Thesaurus to an ontology (5)

- « is-related to » : ambiguous

Solution: New relationships on the based of term usage in the corpus

Part of/ has Part ; Observes/Studied by ; Measured ; Is a property of ; etc..

images tube (Instrumentation) RT photocathode (Artefact) : **RT => "has part"**
cerenkov counters (Instrumentation) RT cosmic rays (Phenomenon) : **=> "observes"**
surface_brightness RT temperature : **RT => "influences"**