



SAP RESEARCH

SYSTEMATIC THOUGHT LEADERSHIP FOR INNOVATIVE BUSINESS

**illustrations  
by Sebastian  
Cremers**

# The role of terminology management in creating ontologies

**Susan Thomas**  
**SAP AG, Research**  
**July 21, 2005**

## **Semantics, definitions, ontologies and terminology management**

- **The importance and role of definitions**

### **Terms**

### **Definitions – how to write them**

- **Emphasis on importance of purpose and context**

### **A bit of history and more on the relation to ontologies**

**Semantics** – ancient Greek for meaning σημαίνω – I signal, sign, show

**Semantics** has become a buzzword or even a fuzzword

**Example from a book about Eclipse:**

- “We’ll use the same mechanisms to navigate semantic errors (...) that we use to navigate compile errors.”
- (failing tests) – semantic error is less precise than “failing tests” – a fuzzword in this case

## Standard way to communicate meaning is by definitions

- Terminology management is the science of organizing and defining terms, usually to support technical documentation and translation.

## Development of ontologies for the semantic web start with definitions

- “the Semantic Web, as envisioned by Tim Berners-Lee and many others since, is a logical extension of the current Web that enables explicit representations of term meanings”<sup>1</sup>

# Standard stages to develop an ontology (by hand)<sup>1</sup>

SAP RESEARCH

1. Determine Scope - purpose
2. Consider reuse
3. Enumerate terms *Define Terms (this was left out of the book)*
  - Statement from member of EU project on semantic-web services: a major barrier to re-use is poor documentation
  - Need definitions so that users and ontology developers can understand the ontology
4. Define taxonomy
5. Define properties
6. Define facets
7. Define instances
8. Check for anomalies

1. from "A Semantic Web Primer", Grigoris Antoniou and Frank van Harmelen

**WordNet is a hand-crafted electronic lexical database of English**

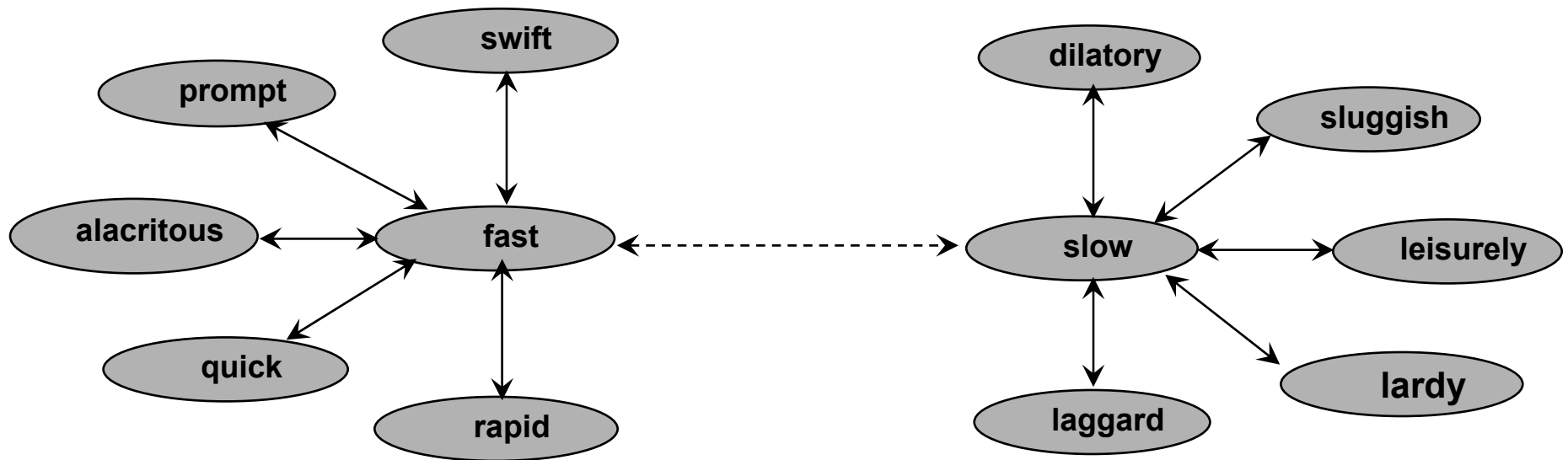
**Similar to simple ontology: network of lexical and semantic relationships**

**Meanings or concepts in WordNet are represented by sets of “synonyms” called synsets**

- {robin, redbreast}

**And organized in a taxonomy**

- {robin, redbreast}@->{bird}@->{animal, animate\_being}@->{organism, life\_form, living\_thing} (example from [Fe98] p. 25)



**Bipolar adjective structure (→ = similarity; ...→ = antonymy)**

1. from "WordNet" (1998) edited by Christiane Fellbaum, p. 51

**Initially WordNet contained no definitions of concepts/synsets at all.**

- **Creators of WordNet thought meanings of words could be easily inferred**
- **But extremely difficult without definitions and illustrative uses of words.**
- **So, they started adding them.**
- **As they themselves admit in [Fe98] p. Roman Numeral xx, they learned the hard way the importance of definitions.**



## Business importance of terms – use same term for same concept

- To avoid misunderstandings that cost money, time, quality, reputation...
  - ◆ E.g., use of standard terms and signs in the chemical industry
- Use terminology database to support technical writing and translation, e.g., English to Spanish

## Similarity of process activities:

### Ontology Creation (use in programs)

1. Determine Scope - purpose
2. Consider reuse
3. Enumerate terms
4. Define taxonomy

### Terminology Management (use by people)

1. Determine boundaries of subject
2. With the help of experts locate artifacts
3. Extract terms from artifacts
4. Write definitions
5. Organize terms

**Interviews with experts**

**Handbooks, Textbooks, Authoritative texts**

**Standards, like ISO standards**

**Glossaries from professional or industry associations**

**Laws, Regulations**

**Dictionaries and Encyclopedias**

**Terminology Management is the science of terms and definitions**

**A term is a linguistic expression of a single specific concept from a particular subject field for a particular purpose**

**Emphasis on:**

- **Concept**
- **Subject field, context**
- **Purpose**

# Terms: examples

TERM FORM	EXAMPLE		TERM FORM	EXAMPLE
Single word	<i>machine</i>		Multi-word	<i>rear-wheel-drive vehicle</i>
Set phrase	<i>Hawks and doves</i>		Collocation	<i>file a patent (words can occur separated in text)</i>
Short form	<i>m (for meter)</i>		Boilerplate	<i>(a set text that describes how to handle a hazardous chemical)</i>

# Comparison of DICT and TERM entries

SAP RESEARCH

<b>LEXICAL ENTRY</b>	<b>TERM ENTRY</b>
<b>Pertains to one word</b>	<b>Pertains to one concept</b>
<b>Lists multiple senses of the word</b>	<b>Lists the terms assigned to the concept</b>
<b>Arranged alphabetically</b>	<b>Arranged in accordance with a concept system, e.g., a taxonomy</b>
<b>Defined based on general knowledge</b>	<b>Defined by careful relation to other concepts in the system</b>
	<b>Facilitates translation, since one concept can be related to the terms for it in multiple languages.</b>

**cardinal**, the noun, has one entry in the American Heritage® dictionary with five senses:

- a Roman Catholic high-church official,
- a color,
- a bird,
- a cloak
- a type of number.

In a terminology collection, which covered all five senses, there would be five entries, one for each word sense.



Example of synonyms, in this context

SAP RESEARCH



“Right now it’s only a notion, but I think I can get money to make it into a concept... and later turn it into an idea.”

**If a good one exists quote it. See dictionaries and ISO standards.**

**Otherwise choose a suitable method:**

- **Intensional**
- **Extensional**
- **Ostensive**
- **Theoretical**
- **Operational**
- **Ad-hoc**
- **Genus and difference, the classic method**

**Key: Determine the intension of the concept and include examples of how the term is used.**



The intension of a concept is its set of distinguishing characteristics

It is essentially a formula for testing if the concept is applicable to something.

The extension of a concept, on the other hand, is determined by its intension, and is the set of all those classes and objects that each have the distinguishing characteristics of the concept.

**Subject field:** Roman Catholic Church

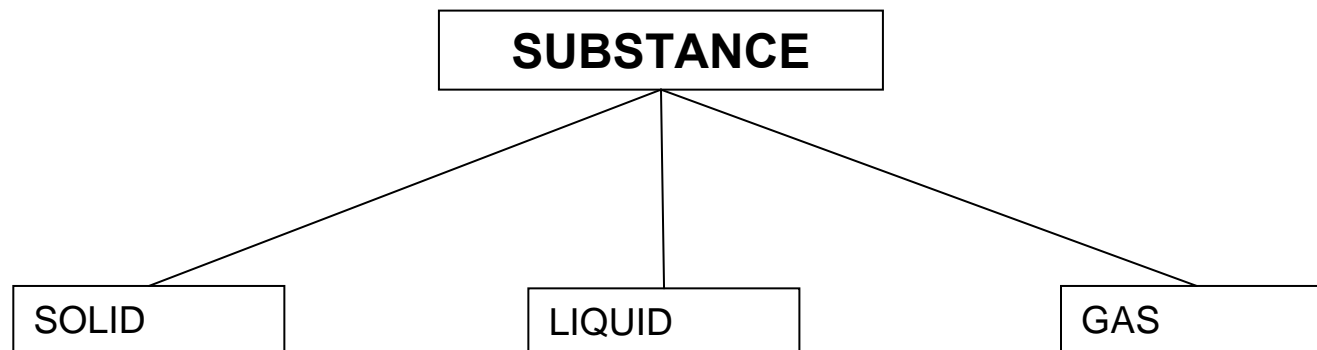
**INTENSION:**

is a high church-official;  
ranks just below the pope;  
and is appointed by the pope to  
membership in the College of Cardinals

**EXTENSION:** all the members of the  
College of Cardinals

**As seen in this example a characteristic can be just about anything that can be said about something: an attribute, relationships with other concepts, or the function or behavior associated with the concept.**





**Subject-field hydromechanics: “a liquid is a substance which is incompressible, very dense and capable of flowing”**

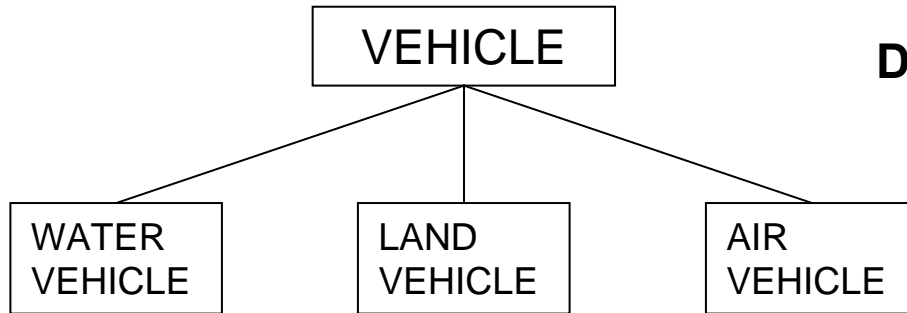
**Genus: substance**

**Subject-field thermodynamics: “a liquid is a substance in a condensed state, intermediate between a solid and a gas”**

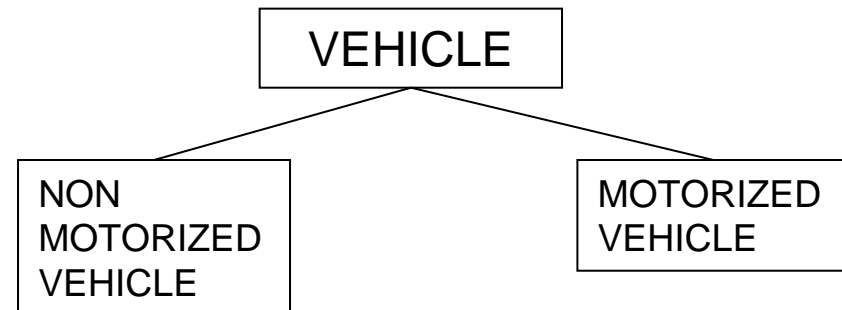
# Multiple Viewpoints of the Concept VEHICLE

SAP RESEARCH

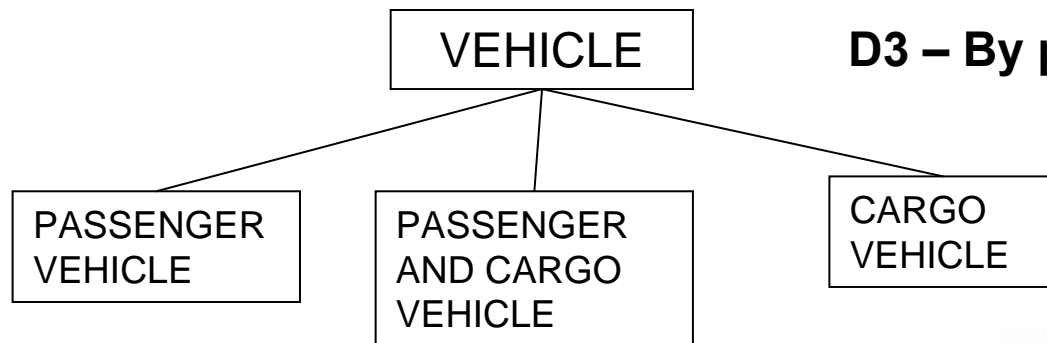
## D1 – By medium of transportation



## D2 – By method of propulsion

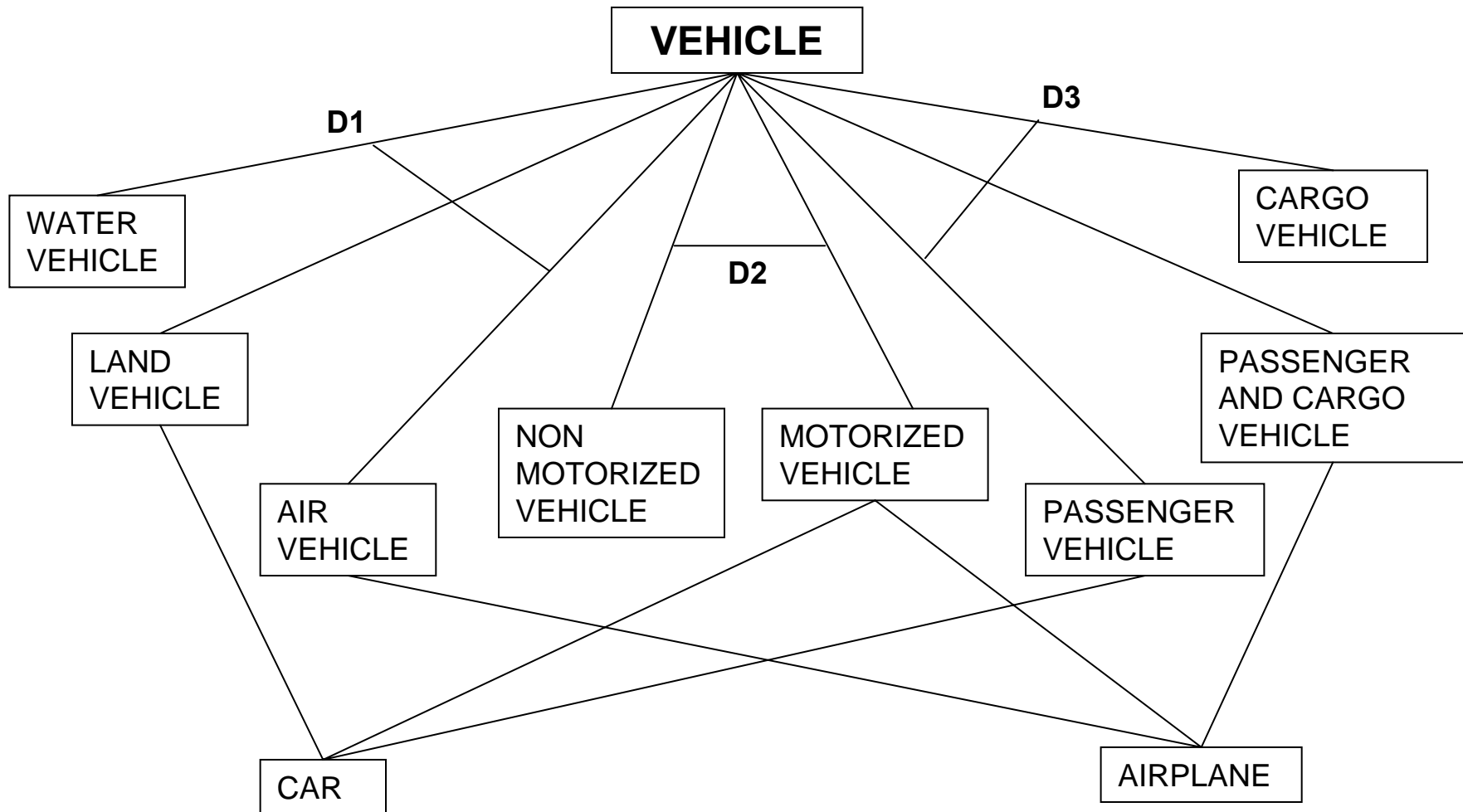


## D3 – By principal type of load carried



# Basic multidimensional view of the concept VEHICLE

SAP RESEARCH



a car is a motorized vehicle  
for transporting passengers by land

## DIMENSIONS:

D1 – By medium of transportation

D2 – By method of propulsion

D3 – By principal type of load carried

***heat* is “a form of ENERGY, viz. the kinetic and potential energy of the invisible molecules of bodies” OED 2nd edition**

***heat* is “an elastic material fluid of extreme subtlety attracted and absorbed by all bodies” (OED, same entry as previous definition)**

**INCOTerms example:**

**"ICC introduced the first version of Incoterms - short for "International Commercial Terms" - in 1936. Since then, ICC expert lawyers and trade practitioners have updated them six times to keep pace with the development of international trade."**

**Next: some examples of definitions and difficulties**

**The UN/CEFACT Core Component Technical Specification (CCTS) is a basis for defining business documents from re-usable components.**

**Universal Business Language (UBL) implements CCTS.**

- <http://docs.oasis-open.org/ubl/cd-UBL-1.0/>

**Core Component Types:**

- Amount, BinaryObject, Code, DateTime, Identifier, Indicator, Measure, Numeric, Quantity, Text

**Some problems with core types as evidenced by UBL V1.0**

- Use of numeric for interest rates
- Confusion of identifier and code
- Confusion of measure and quantity

**Numeric: “Numeric information that is assigned or is determined by calculation, counting, or sequencing. It does not require a unit of quantity or unit of measure.”**

**unit-free measure (OED definition)**

**“ A variable which is a pure number, independent of the units in which variables such as price or quantity are measured. Examples of unit-free measures are percentages, market shares, and elasticities. Interest rates and growth rates are not unit-free measures, as while they are independent of the units in which prices and quantities are measured, they do depend on the units used to measure time .”**

**Unit of Measure (from Rec 20 of UN/ECE) (1000 UoMs w/o defn.)**

**"Particular quantity, defined and adopted by convention, with which other quantities of the same kind are compared in order to express their magnitudes relative to that quantity."**



**Compare how UBL and TBG17 specify the country in an address.**

- **UBL: Address.Country and Country.Identification.Code (a code)**

**Compared to TBG17**

- **TBG17: Address.Country.Identifier (an identifier)**

**In both cases, the definitions are similar, showing that the same concept is intended.**

- **UBL: "provides the country part of an address using a code. ISO3166 alpha codes are recommended"**
- **TBG17: "To specify any identifier of a country such that specified in ISO 3166 and UN/ECE Rec 3."**

**Identifier: A character string to identify and distinguish uniquely, one instance of an object in an identification scheme from all other objects in the same scheme together with relevant supplementary information.**

**Code: A character string (letters, figures, or symbols) that for brevity and/or language independence may be used to represent or replace a definitive value or text of an attribute together with relevant supplementary information**

**Measure and quantity are two other CC types that are confused.**

**An example from UBL is as follows:**

- **Invoice Line. Invoiced\_Quantity. Quantity (BIE)**

**Instance:**

- **<cbc:InvoicedQuantity  
quantityUnitCode="feet">30</cbc:InvoicedQuantity>**

**Looking at the CC definitions of Quantity and Measure, this appears to be a Measure.**

- **Measure: A numeric value determined by measuring an object along with the specified unit of measure.**
- **Quantity: A counted number of non-monetary units possibly including fractions.**

**Terms and definitions play a central role in semantics**

- **Examples also key; problems in understanding traceable to both**

**Purpose and context play an extremely important role in definitions**

**Writing definitions is an art not a science**

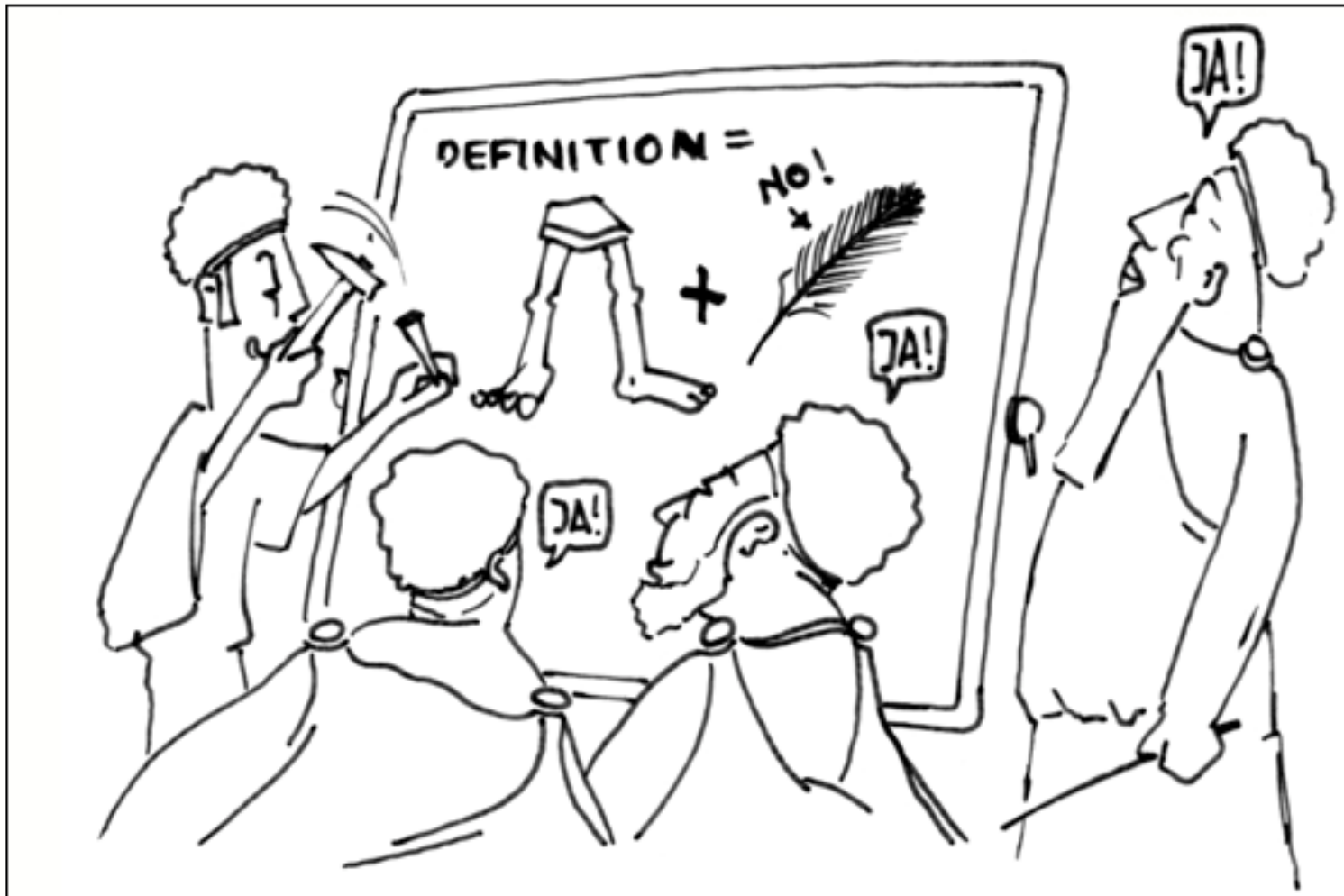
**It is and was difficult as shown next.**

# Definition is not easy

SAP RESEARCH

**An example: not long after the days of Plato in the Academy at Athens.....**











**Necessity and sufficiency provide another way to think of intension and throw some light on the previous academic definition of man.**

- **Characteristics that are necessary to a class define a super-class of that class.**
  - ◆ For example, featherless biped is a super-class of man.
- **Characteristics that are sufficient to a class define a sub-class of that class.**
  - ◆ Investing 150,000 pounds in the country is sufficient to qualify for residency in the UK
  - ◆ people who qualify under this rule are a sub-class of people who qualify for residency.

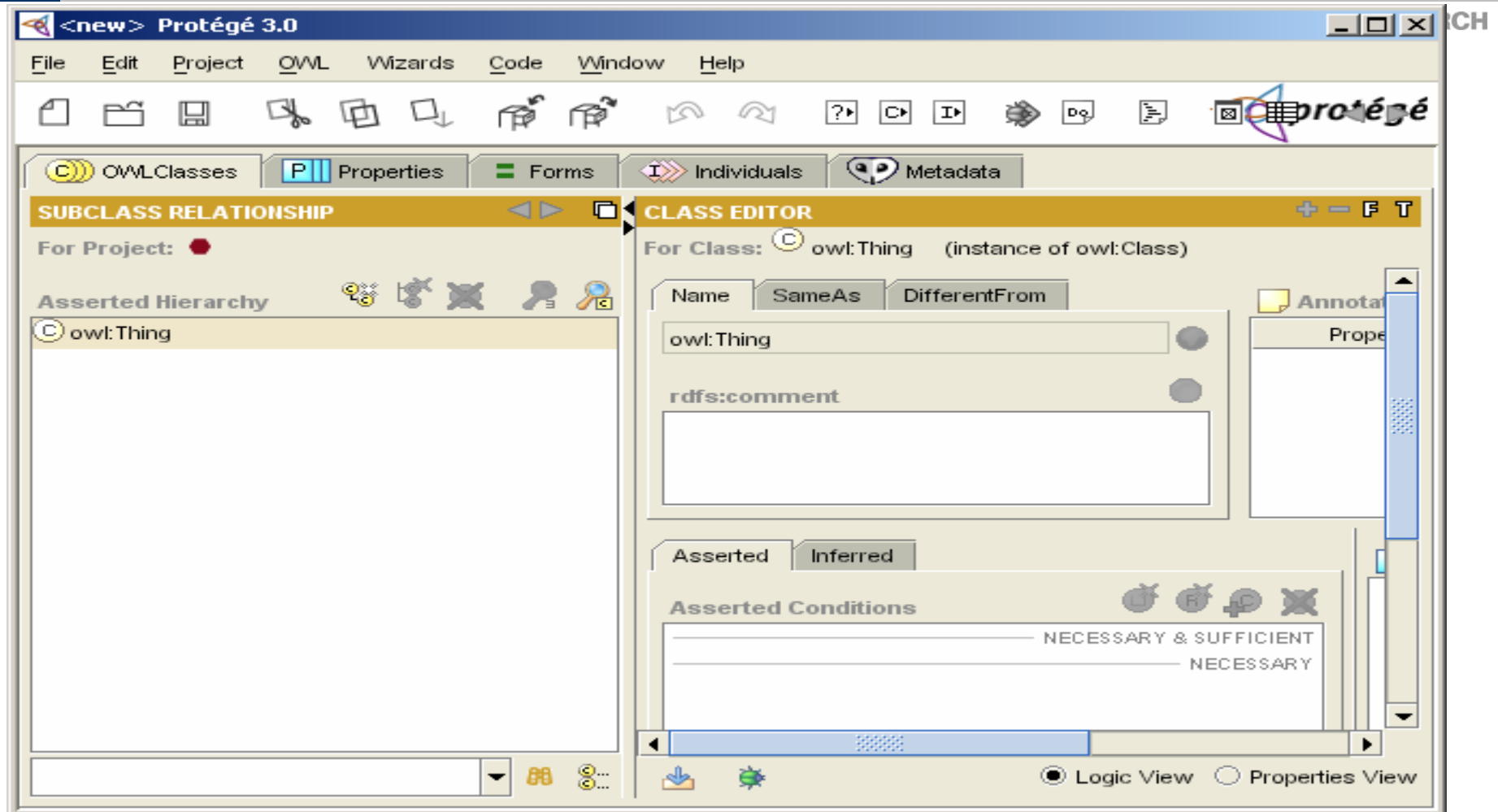
**The set of characteristics that is both necessary and sufficient defines a class by intension.**

- **necessary conditions only defined a super-class of man, a super-class that also included plucked chickens.**

**Corresponds to asserted conditions in the Protégé OWL editor**

- **Necessary and sufficient – defined class, classifiers work**
- **Necessary – primitive class**

# Protégé window



**Corresponds to Asserted Conditions in the Protégé OWL editor**  
**Necessary – primitive class**  
**Necessary and sufficient – defined class, classifiers work**

**Book: Definition, Richard Robinson, 1968, Oxford Press**

**Paper on definitions in BTW2005 proceedings:**

- **“The Importance of being Earnest about Definitions”**

Where did the title of the paper come from?

SAP RESEARCH

Oscar Wilde (author of the play, The Importance of being Earnest)

*"To be really medieval one should have no body."*

SAP RESEARCH



*"To be really modern one should have no soul."*

SAP RESEARCH



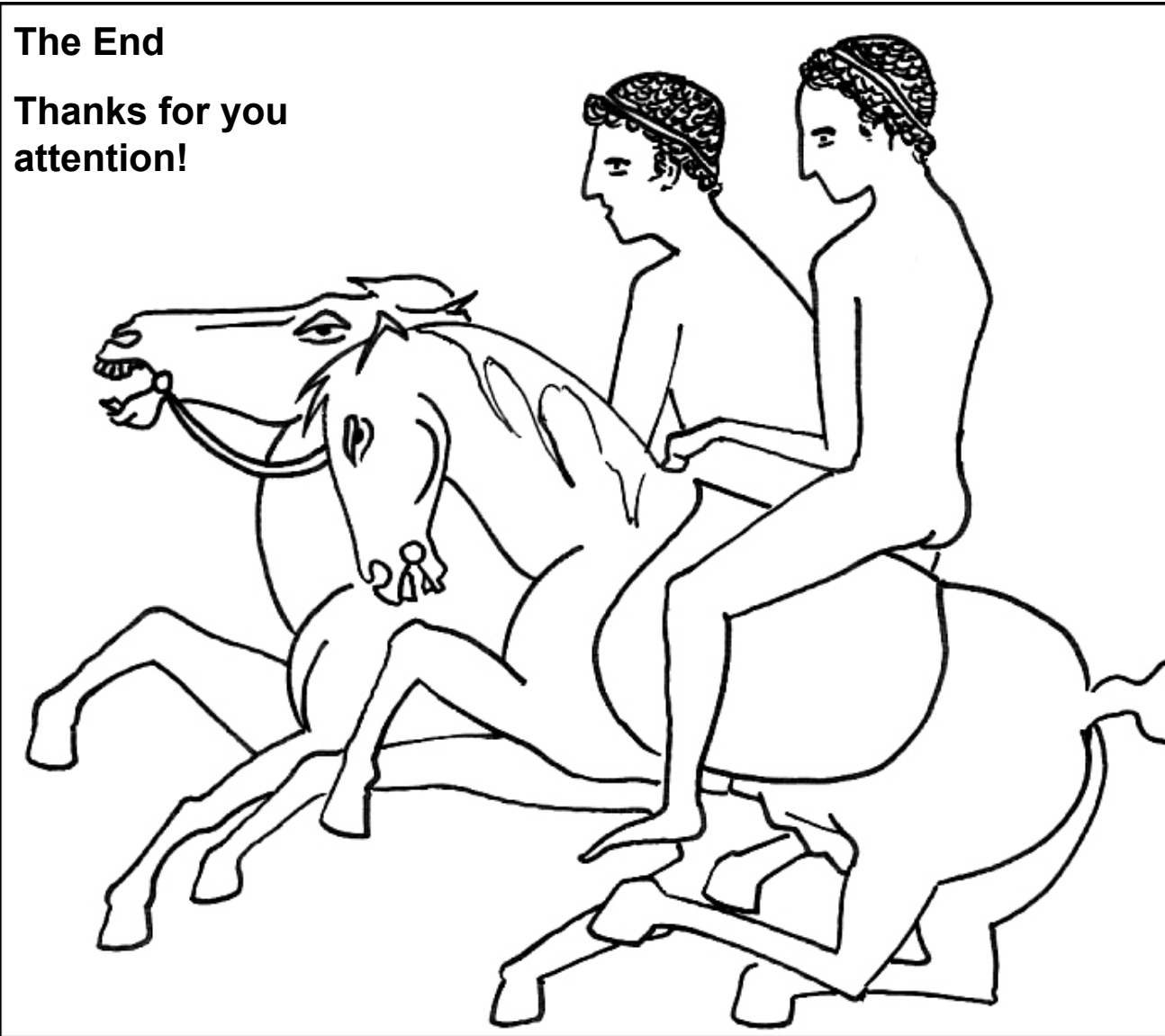
SC  
'05

*"To be really Greek one should have no clothes."*

SAP RESEARCH

**The End**

**Thanks for you  
attention!**



- No part of this publication may be reproduced or transmitted in any form or for any purpose without the express permission of SAP AG. The information contained herein may be changed without prior notice.
- Some software products marketed by SAP AG and its distributors contain proprietary software components of other software vendors.
- Microsoft, Windows, Outlook, and PowerPoint are registered trademarks of Microsoft Corporation.
- IBM, DB2, DB2 Universal Database, OS/2, Parallel Sysplex, MVS/ESA, AIX, S/390, AS/400, OS/390, OS/400, iSeries, pSeries, xSeries, zSeries, z/OS, AFP, Intelligent Miner, WebSphere, Netfinity, Tivoli, and Informix are trademarks or registered trademarks of IBM Corporation in the United States and/or other countries.
- Oracle is a registered trademark of Oracle Corporation.
- UNIX, X/Open, OSF/1, and Motif are registered trademarks of the Open Group.
- Citrix, ICA, Program Neighborhood, MetaFrame, WinFrame, VideoFrame, and MultiWin are trademarks or registered trademarks of Citrix Systems, Inc.
- HTML, XML, XHTML and W3C are trademarks or registered trademarks of W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.
- Java is a registered trademark of Sun Microsystems, Inc.
- JavaScript is a registered trademark of Sun Microsystems, Inc., used under license for technology invented and implemented by Netscape.
- MaxDB is a trademark of MySQL AB, Sweden.
- SAP, R/3, mySAP, mySAP.com, xApps, xApp, SAP NetWeaver and other SAP products and services mentioned herein as well as their respective logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries all over the world. All other product and service names mentioned are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.
- These materials are subject to change without notice. These materials are provided by SAP AG and its affiliated companies ("SAP Group") for informational purposes only, without representation or warranty of any kind, and SAP Group shall not be liable for errors or omissions with respect to the materials. The only warranties for SAP Group products and services are those that are set forth in the express warranty statements accompanying such products and services, if any. Nothing herein should be construed as constituting an additional warranty.



- Weitergabe und Vervielfältigung dieser Publikation oder von Teilen daraus sind, zu welchem Zweck und in welcher Form auch immer, ohne die ausdrückliche schriftliche Genehmigung durch SAP AG nicht gestattet. In dieser Publikation enthaltene Informationen können ohne vorherige Ankündigung geändert werden.
- Die von SAP AG oder deren Vertriebsfirmen angebotenen Softwareprodukte können Softwarekomponenten auch anderer Softwarehersteller enthalten.
- Microsoft, Windows, Outlook, und PowerPoint sind eingetragene Marken der Microsoft Corporation.
- IBM, DB2, DB2 Universal Database, OS/2, Parallel Sysplex, MVS/ESA, AIX, S/390, AS/400, OS/390, OS/400, iSeries, pSeries, xSeries, zSeries, z/OS, AFP, Intelligent Miner, WebSphere, Netfinity, Tivoli, und Informix sind Marken oder eingetragene Marken der IBM Corporation in den USA und/oder anderen Ländern.
- Oracle ist eine eingetragene Marke der Oracle Corporation.
- UNIX, X/Open, OSF/1, und Motif sind eingetragene Marken der Open Group.
- Citrix, ICA, Program Neighborhood, MetaFrame, WinFrame, VideoFrame, und MultiWin sind Marken oder eingetragene Marken von Citrix Systems, Inc.
- HTML, XML, XHTML und W3C sind Marken oder eingetragene Marken des W3C®, World Wide Web Consortium, Massachusetts Institute of Technology.
- Java ist eine eingetragene Marke von Sun Microsystems, Inc.
- JavaScript ist eine eingetragene Marke der Sun Microsystems, Inc., verwendet unter der Lizenz der von Netscape entwickelten und implementierten Technologie.
- MaxDB ist eine Marke von MySQL AB, Schweden.
- SAP, R/3, mySAP, mySAP.com, xApps, xApp, SAP NetWeaver und weitere im Text erwähnte SAP-Produkte und -Dienstleistungen sowie die entsprechenden Logos sind Marken oder eingetragene Marken der SAP AG in Deutschland und anderen Ländern weltweit. Alle anderen Namen von Produkten und Dienstleistungen sind Marken der jeweiligen Firmen. Die Angaben im Text sind unverbindlich und dienen lediglich zu Informationszwecken. Produkte können länderspezifische Unterschiede aufweisen.
- In dieser Publikation enthaltene Informationen können ohne vorherige Ankündigung geändert werden. Die vorliegenden Angaben werden von SAP AG und ihren Konzernunternehmen („SAP-Konzern“) bereitgestellt und dienen ausschließlich Informationszwecken. Der SAP-Konzern übernimmt keinerlei Haftung oder Garantie für Fehler oder Unvollständigkeiten in dieser Publikation. Der SAP-Konzern steht lediglich für Produkte und Dienstleistungen nach der Maßgabe ein, die in der Vereinbarung über die jeweiligen Produkte und Dienstleistungen ausdrücklich geregelt ist. Aus den in dieser Publikation enthaltenen Informationen ergibt sich keine weiterführende Haftung.