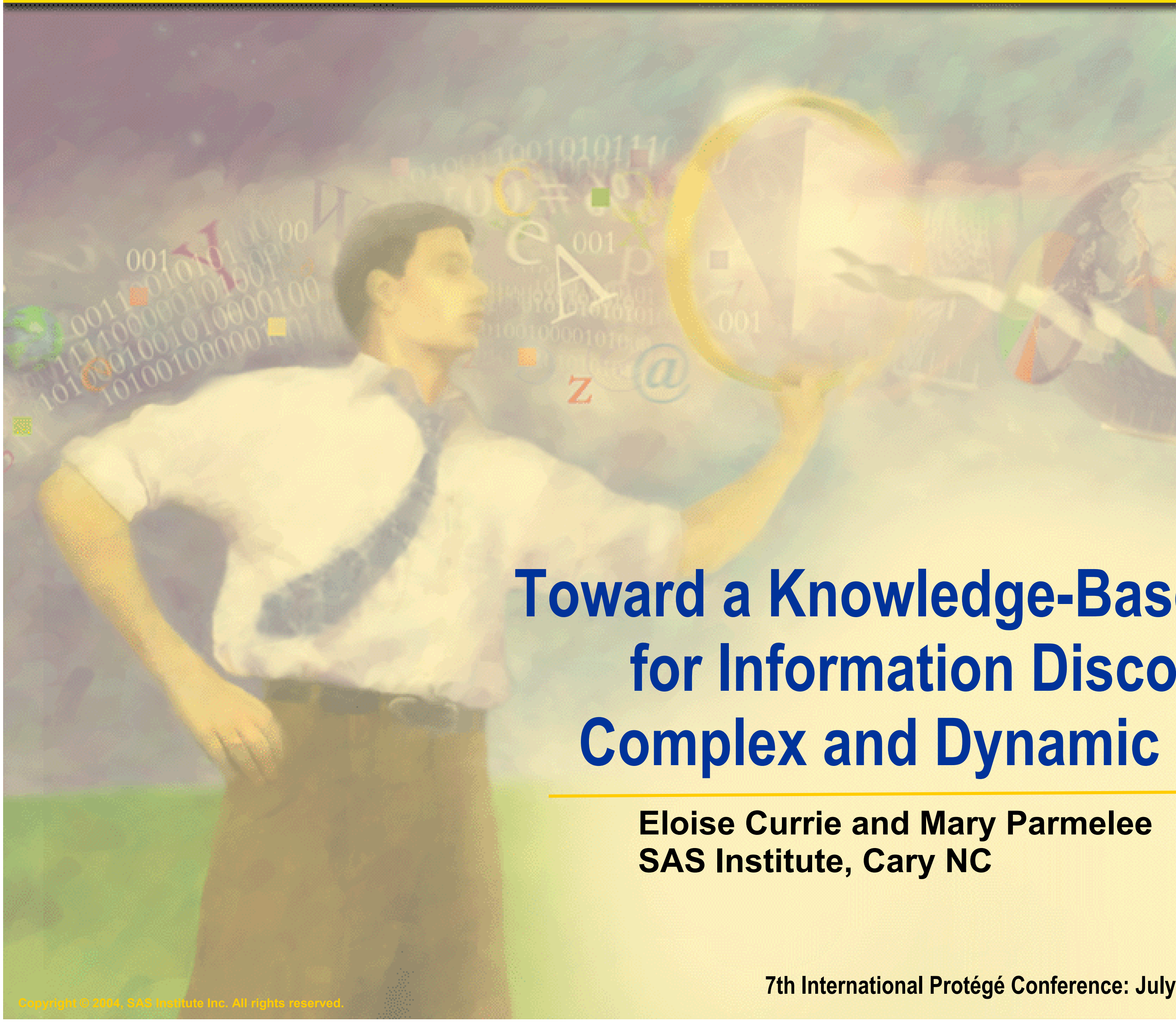
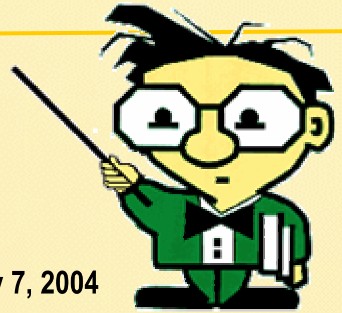


Toward a Knowledge-Based Solution for Information Discovery in Complex and Dynamic Domains

Eloise Currie and Mary Parmelee
SAS Institute, Cary NC



7th International Protégé Conference: July 7, 2004

About *SAS*: *The Power to Know®*

- ***SAS***: The Market Leader in Business Intelligence Software
- Founded: 1976
- World’s Largest Privately Held Software Company
- Worldwide Offices: 269
- Worldwide Employees: 9,238
- Worldwide Revenue 2003: $1.34 billion
- Reinvestment in R&D 2003: 26%
- ***SAS*** Solutions:
    - Used at more than 40,000 sites
    - Used by 96 of the top 100 of the 2003 Fortune Global 500

The Problem with Information

- From a Global Perspective
  - As the volume of online information grows, information retrieval (IR) has become a major challenge.
- How much is it changing?
  - In 1995, over 90% of corporate documents were in paper form. By 2005, less than 30% will remain in paper form.
  - Within the next 3 years, the world will produce as much data as has been produced since **THE DAWN OF TIME!**

The Problem with Information

From a SAS Perspective

1. Address Customer Pains:
   - **SAS** customers are satisfied with the quality of our documentation, but they have difficulty locating information.
2. Manage Growth and Complexity
   - How much is it changing?
   - SAS product growth: 1 > 170+
   - SAS user growth: 4 million users worldwide

Toward a Solution: Preliminary Research

- **Observation**
  - Online information and queries are communicated via natural language, which has two main properties:
    - semantics gives meaning in **context**
    - syntactics give structure and order
    - Yet, most IR systems match only syntactics
- **Objective**
  - Create an IR system that leverages the semantics of natural language.
- **Investigation**
  - Emerging technologies, initiatives and standards: Semantic Web, Ontologies, RDF
  - Consulted IR Experts (UNC-Chapel Hill)
  - Tools: Protégé, Jena Toolkit

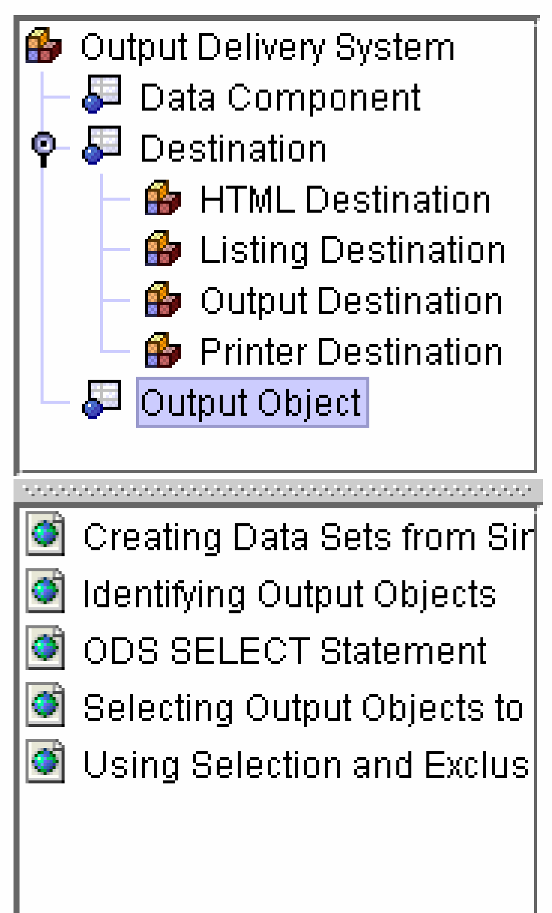
Toward a Solution: Progress to Date

2001: Proof of concept project

- **Tiny domain** (subject area): two pages of documentation
- **Rudimentary UI**
- **Deliverable:** development methodology and repeatable process

Browsable directory tree

Resources associated with a node in the tree



- Output Delivery System
  - Data Component
  - Destination
    - HTML Destination
    - Listing Destination
    - Output Destination
    - Printer Destination
  - Output Object

- Creating Data Sets from Sir
- Identifying Output Objects
- ODS SELECT Statement
- Selecting Output Objects to
- Using Selection and Exclus

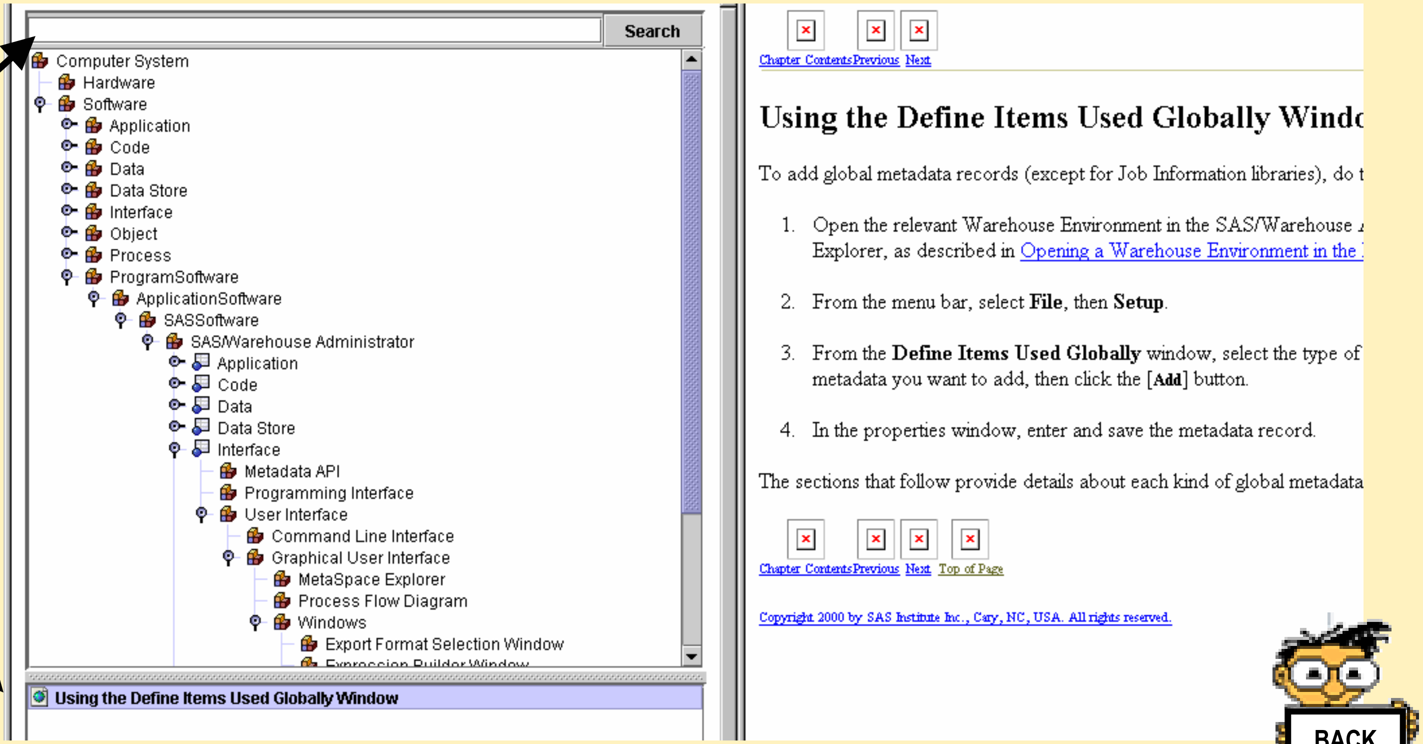
Toward a Solution: Progress to Date

2002: Pilot Project

- **Larger domain**: a complete product user's guide
- **More robust UI**
- **Deliverable**: A functioning mini system

Added search functionality

Search results pane



Using the Define Items Used Globally Windo

To add global metadata records (except for Job Information libraries), do t

1. Open the relevant Warehouse Environment in the SAS/Warehouse A Explorer, as described in Opening a Warehouse Environment in the

2. From the menu bar, select **File**, then **Setup**.

3. From the **Define Items Used Globally** window, select the type of metadata you want to add, then click the [**Add**] button.

4. In the properties window, enter and save the metadata record.

The sections that follow provide details about each kind of global metadata

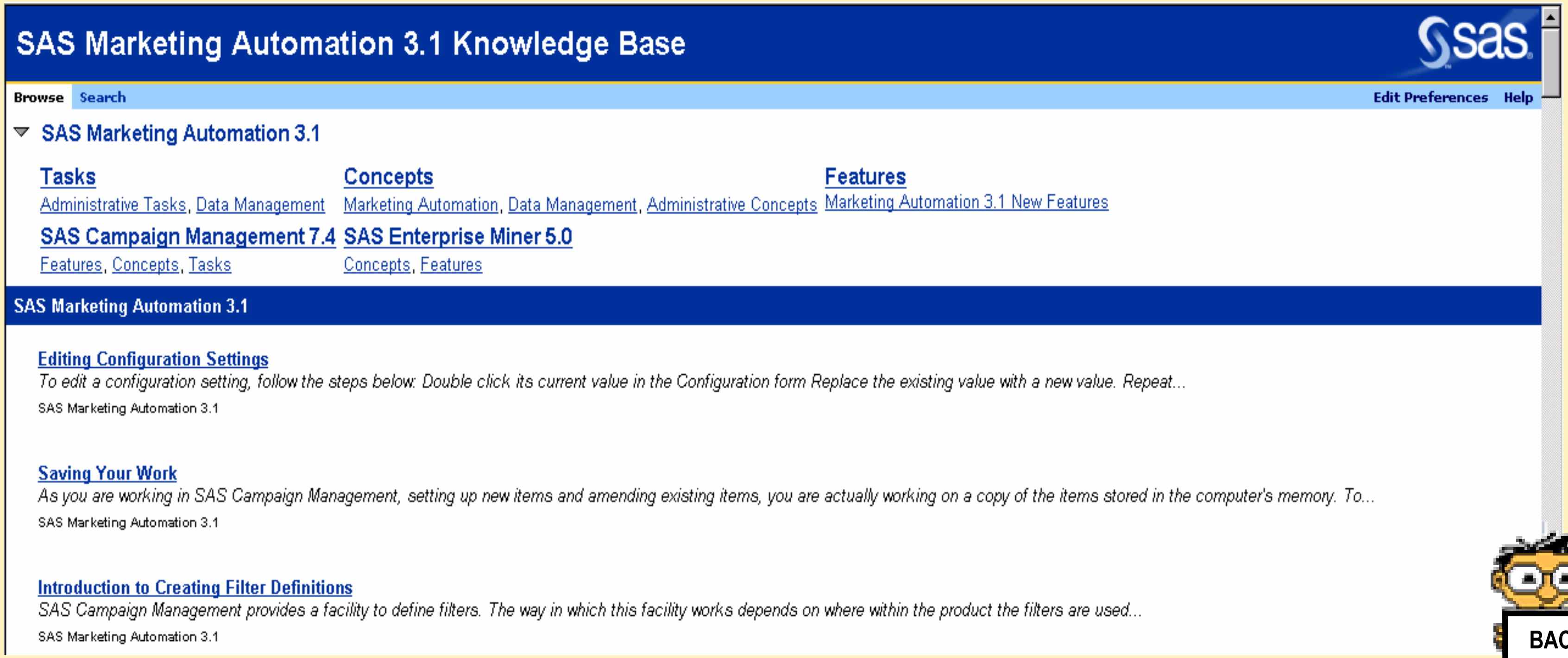
Copyright 2000 by SAS Institute Inc., Cary, NC, USA. All rights reserved.

BACK GROUND

Toward a Solution: Progress to Date

February 2004: Prototype Project

- **Large domain**: Includes several products
- **Intelligence layer:** Enables advanced search and reasoning capability
- **Advanced UI**: Delivers information in context while minimizing the complexity surfaced to the user
- **Deliverable:** Fully Functional Prototype


SAS Marketing Automation 3.1 Knowledge Base

▽ SAS Marketing Automation 3.1

Tasks
Administrative Tasks, Data Management

Concepts
Marketing Automation, Data Management, Administrative Concepts

Features
Marketing Automation 3.1 New Features

SAS Campaign Management 7.4
Features, Concepts, Tasks

SAS Enterprise Miner 5.0
Concepts, Features

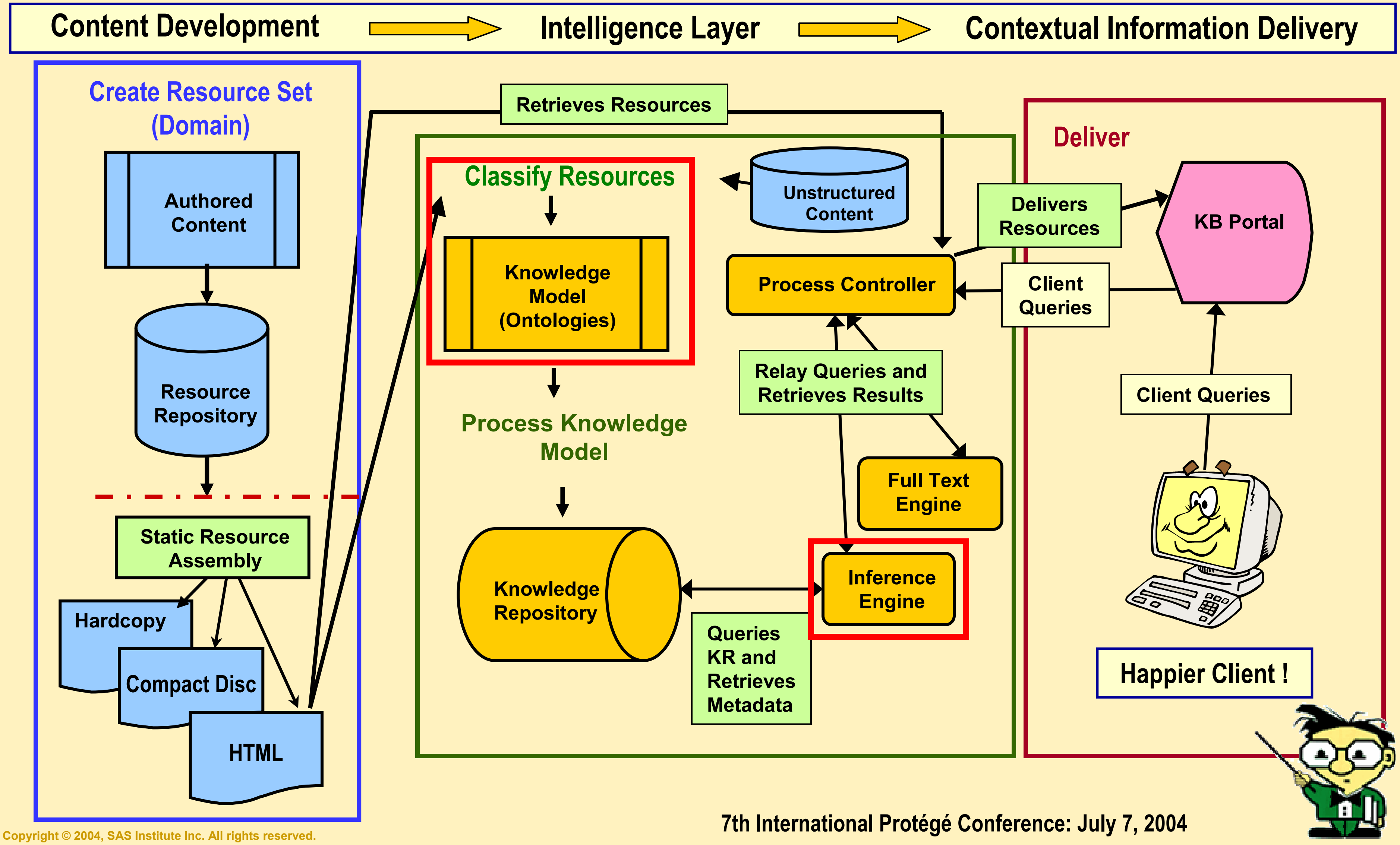
SAS Marketing Automation 3.1

Editing Configuration Settings
To edit a configuration setting, follow the steps below: Double click its current value in the Configuration form Replace the existing value with a new value. Repeat...
SAS Marketing Automation 3.1

Saving Your Work
As you are working in SAS Campaign Management, setting up new items and amending existing items, you are actually working on a copy of the items stored in the computer's memory. To...
SAS Marketing Automation 3.1

Introduction to Creating Filter Definitions
SAS Campaign Management provides a facility to define filters. The way in which this facility works depends on where within the product the filters are used...
SAS Marketing Automation 3.1

Knowledge Base System Architecture



7th International Protégé Conference: July 7, 2004

System Development Process

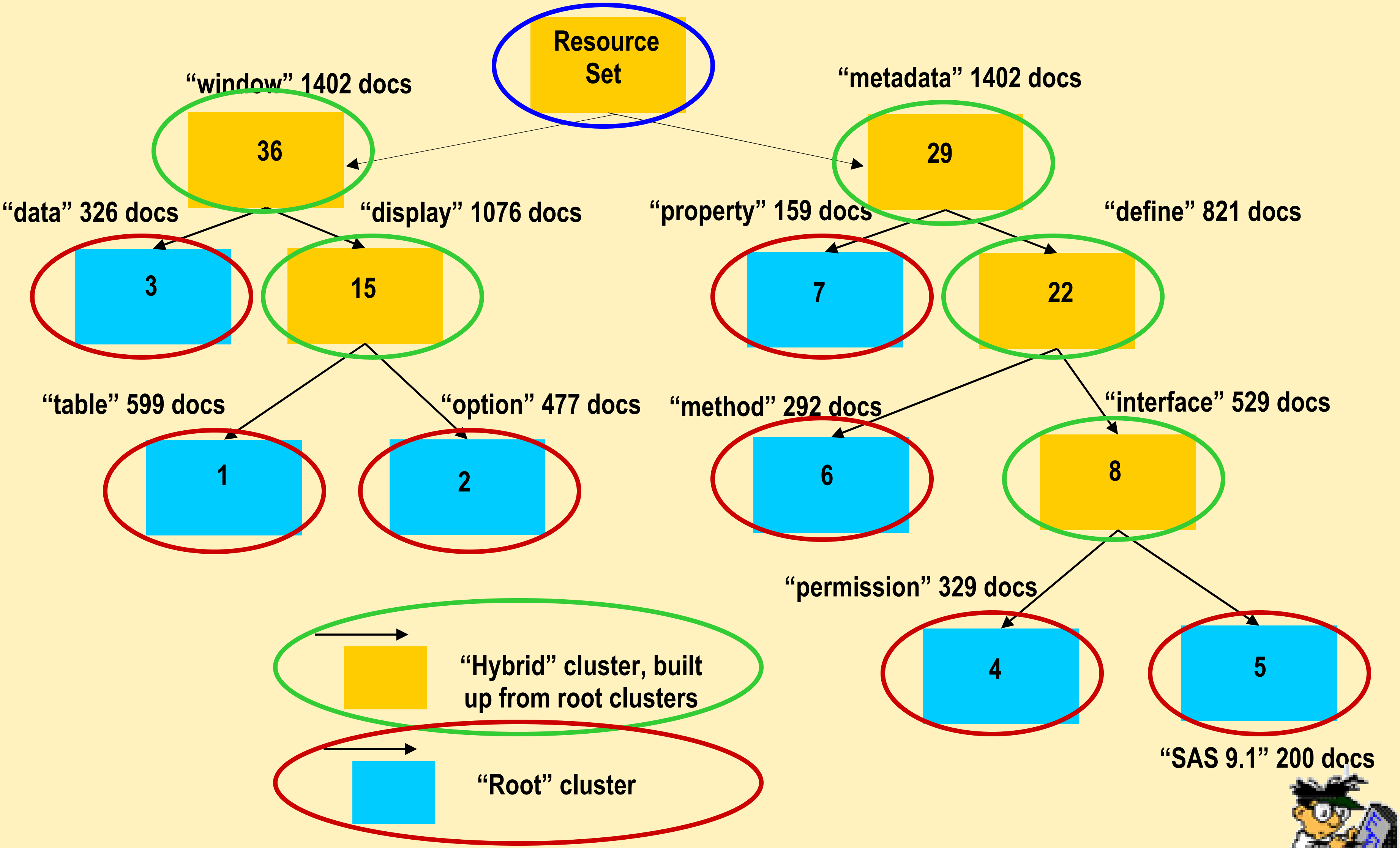
We begin with a document collection (a “resource set”)..

1. Use SAS® Text Miner to create a hierarchy of resource clusters
2. Use a custom Protégé plugin to generate a **Domain ontology** that categorizes resources based on content
3. Use a custom Protégé plugin to extract resource information and generate a **Resource Manager ontology**
4. Merge Domain and Resource ontologies into a **Master ontology**
5. Use a custom Protégé plugin to map Resource instances to Domain instance slots
6. Use Protégé to develop the merged ontology into a production Master ontology
7. Use a custom Protégé plugin to reverse map Domain Instances to Resource Instance Slots

Step 1: Use SAS® Text Miner to Create a Hierarchy of Resource Clusters

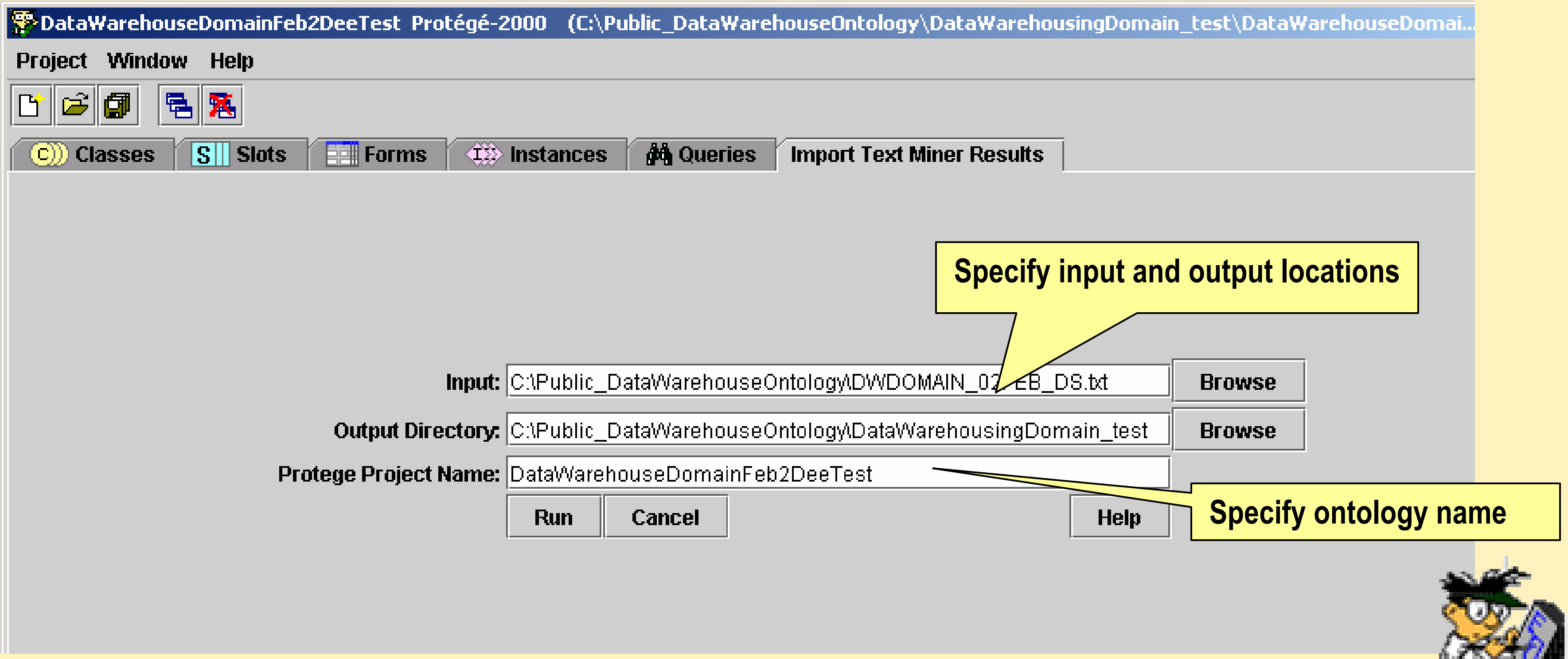
Steps	Result
Preprocessing	Create a SAS data set from the document collection.
Text parsing	Generate quantitative representation of the content
Transformation	Consolidate quantitative representation
Document analysis	Cluster documents by concept

Step 1:Hierarchical Output



Step 2: Use a Custom Protégé Plugin to Generate a Domain Ontology in Protégé

- Saves Text Miner hierarchy as a Protégé ontology
- Preserves URI's of documents in a Resource ID slot


DataWarehouseDomainFeb2DeeTest Protégé-2000 (C:\Public_DataWarehouseOntology\DataWarehousingDomain_test\DataWarehouseDomai...
Project
Window
Help
Classes
Slots
Forms
Instances
Queries
Import Text Miner Results
Browse
Browse
Run
Cancel
Help

Specify input and output locations

Input: C:\Public_DataWarehouseOntology\DWDOMAIN_02 EB_DS.txt

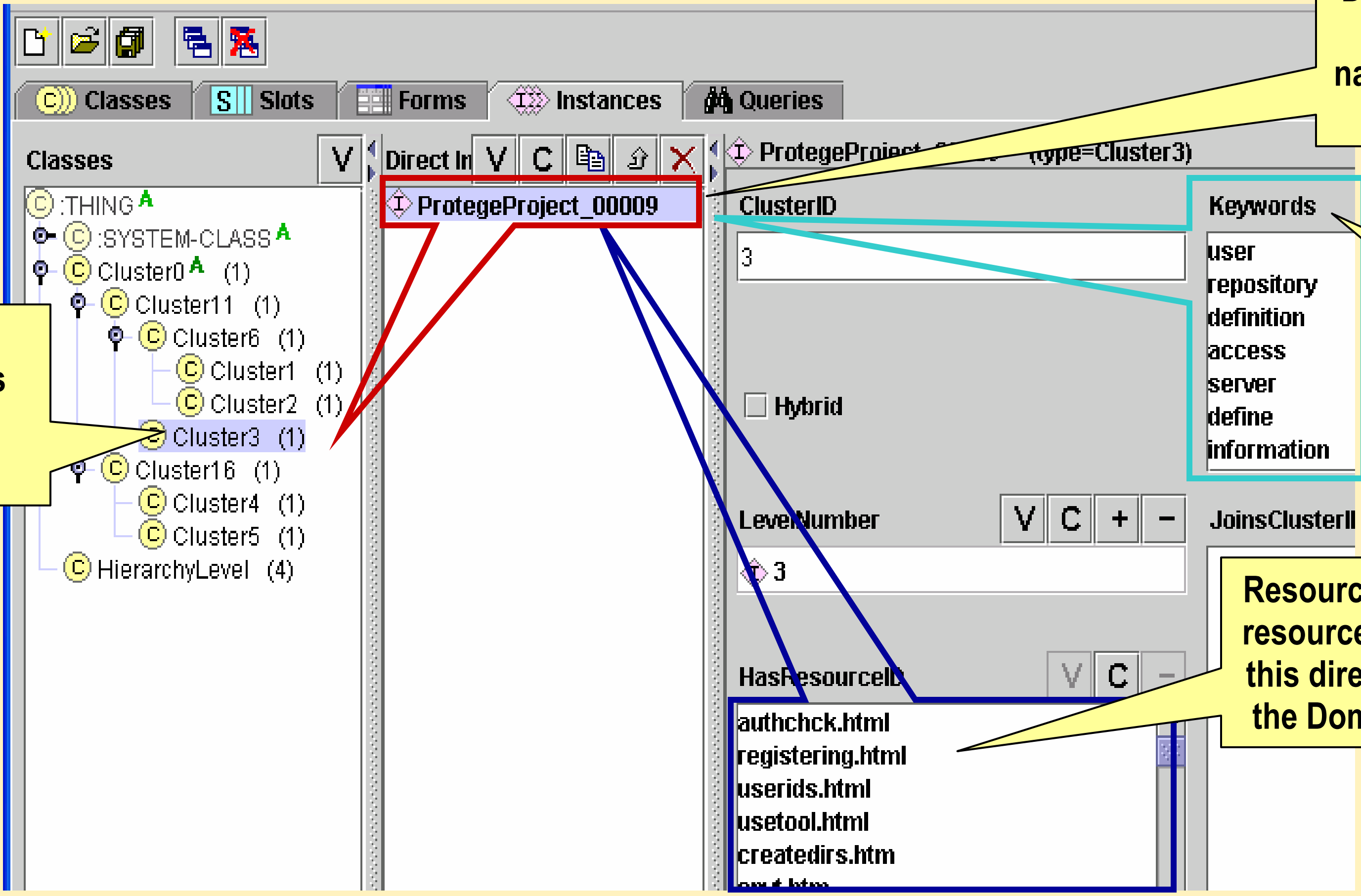
Output Directory: C:\Public_DataWarehouseOntology\DataWarehousingDomain_test

Protege Project Name: DataWarehouseDomainFeb2DeeTest

Specify ontology name

Step 2: Domain Ontology in Protégé

- Each direct instance is a Text Miner document cluster
- Generic instance and class names will be edited



:THING
:SYSTEM-CLASS
Cluster0 (1)
Cluster11 (1)
Cluster6 (1)
Cluster1 (1)
Cluster2 (1)
Cluster3 (1)
Cluster16 (1)
Cluster4 (1)
Cluster5 (1)
HierarchyLevel (4)

ProtegeProject_00009

ProtegeProject (type=Cluster3)

ClusterID: 3

☐ Hybrid

Keywords: user, repository, definition, access, server, define, information

LevelNumber: 3

JoinsClusterI

HasResourceID:
authchck.html
registering.html
userids.html
usetool.html
createdirs.htm

Direct Instances have generic names generated by Protégé

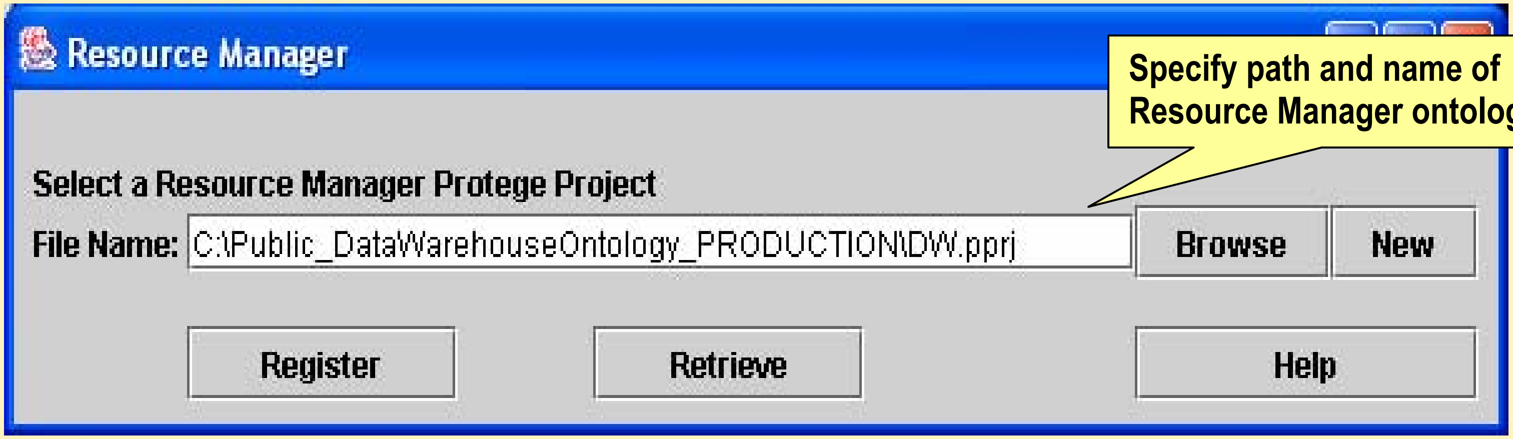
Classes have generic names generated by Text Miner

Significant cluster terms identified by Text Miner

Resource IDs of those resources assigned to this direct instance of the Domain ontology

Step 3: Use a Custom Protégé Plugin to Generate a Resource Manager Ontology in Protégé

- Extracts resource information
- Preserves URI’s of documents in a Resource ID slot
- Saves resource information in a Protégé ontology



Select a Resource Manager Protege Project

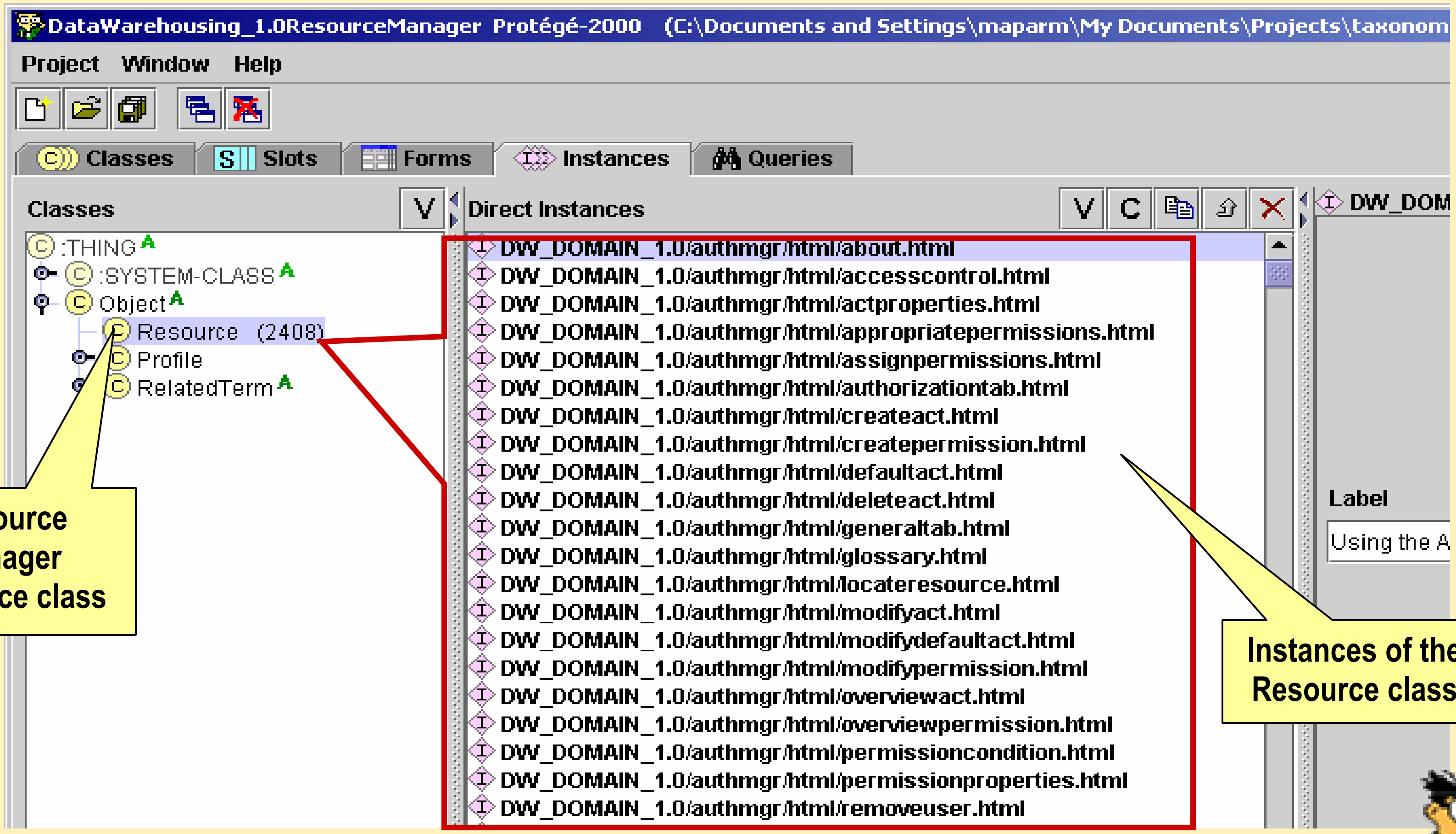
File Name: C:\Public_DataWarehouseOntology_PRODUCTION\DW.pprj

Specify path and name of Resource Manager ontology

7th International Protégé Conference: July 7, 2004

Step 3: Resource Manager Ontology

Each Direct Instance is a document in the Resource class



DataWarehousing_1.0ResourceManager Protégé-2000 (C:\Documents and Settings\maparm\My Documents\Projects\taxonom

Classes

- :THING
  - :SYSTEM-CLASS
  - Object
    - Resource (2408)
    - Profile
    - RelatedTerm

Direct Instances

DW_DOMAIN_1.0/authmgr/html/about.html
DW_DOMAIN_1.0/authmgr/html/accesscontrol.html
DW_DOMAIN_1.0/authmgr/html/actproperties.html
DW_DOMAIN_1.0/authmgr/html/appropriatepermissions.html
DW_DOMAIN_1.0/authmgr/html/assignpermissions.html
DW_DOMAIN_1.0/authmgr/html/authorizationtab.html
DW_DOMAIN_1.0/authmgr/html/createact.html
DW_DOMAIN_1.0/authmgr/html/createpermission.html
DW_DOMAIN_1.0/authmgr/html/defaultact.html
DW_DOMAIN_1.0/authmgr/html/deleteact.html
DW_DOMAIN_1.0/authmgr/html/generaltab.html
DW_DOMAIN_1.0/authmgr/html/glossary.html
DW_DOMAIN_1.0/authmgr/html/locateresource.html
DW_DOMAIN_1.0/authmgr/html/modifyact.html
DW_DOMAIN_1.0/authmgr/html/modifydefaultact.html
DW_DOMAIN_1.0/authmgr/html/modifypermission.html
DW_DOMAIN_1.0/authmgr/html/overviewact.html
DW_DOMAIN_1.0/authmgr/html/overviewpermission.html
DW_DOMAIN_1.0/authmgr/html/permissioncondition.html
DW_DOMAIN_1.0/authmgr/html/permissionproperties.html
DW_DOMAIN_1.0/authmgr/html/removeuser.html

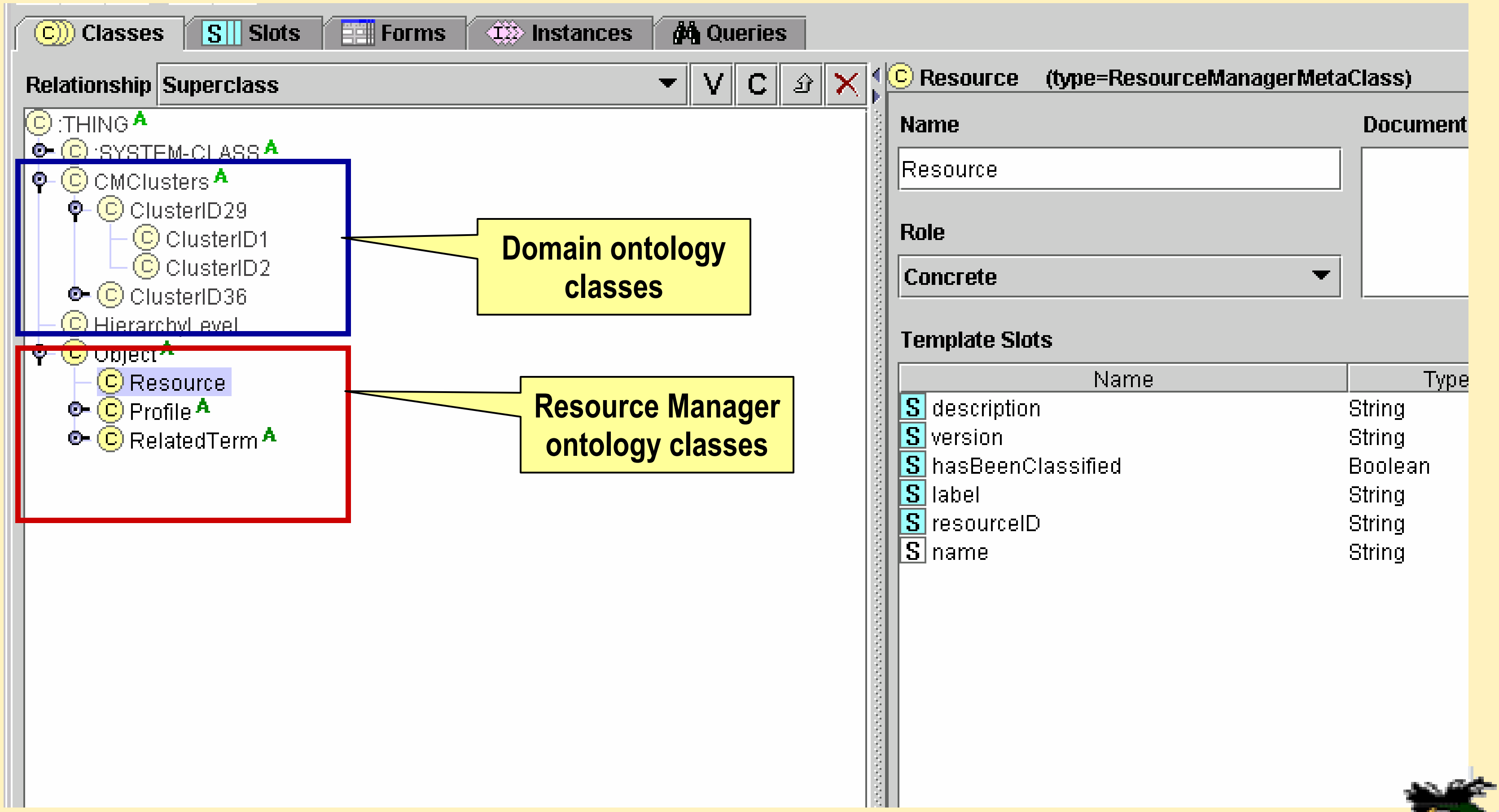
DW_DOM

Resource Manager Resource class

Instances of the Resource class

7th International Protégé Conference: July 7, 2004

Step 4: Merge the Domain and Resource Ontologies into a Master Ontology



Relationship: Superclass

Resource (type=ResourceManagerMetaClass)

Name: Resource

Role: Concrete

Template Slots

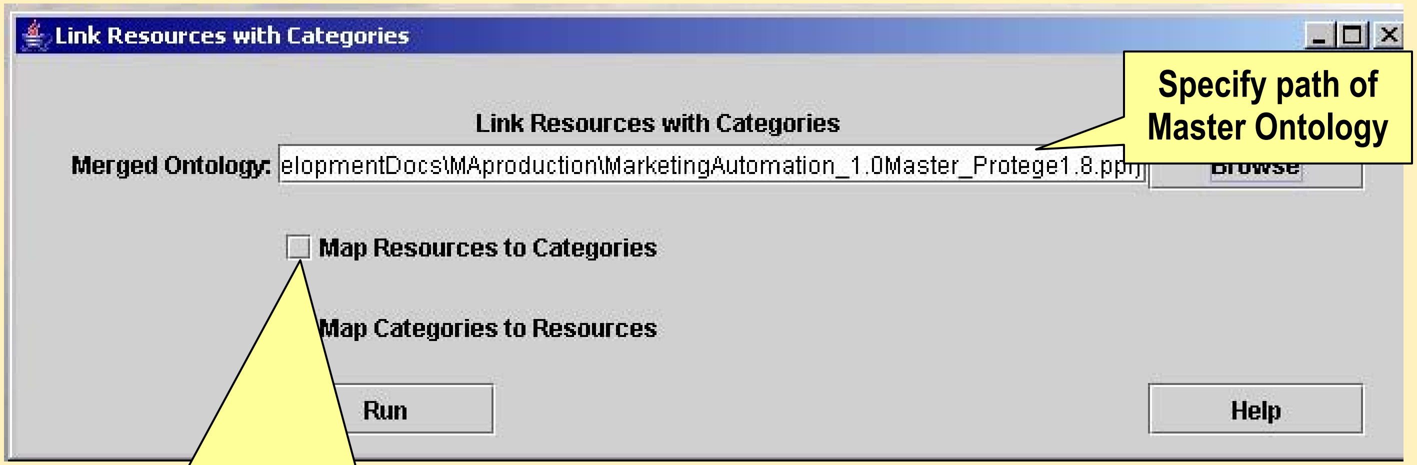
Name	Type
description	String
version	String
hasBeenClassified	Boolean
label	String
resourceID	String
name	String

Domain ontology classes

Resource Manager ontology classes

Step 5: Use a Custom Protégé Plugin to Map Resource Instances to Domain Instance Slots

- Maps by Resource ID
- Populates resource instance slot



Link Resources with Categories

Link Resources with Categories

Merged Ontology: elopmentDocs\MAproduction\MarketingAutomation_1.0Master_Protege1.8.ppr

□ Map Resources to Categories

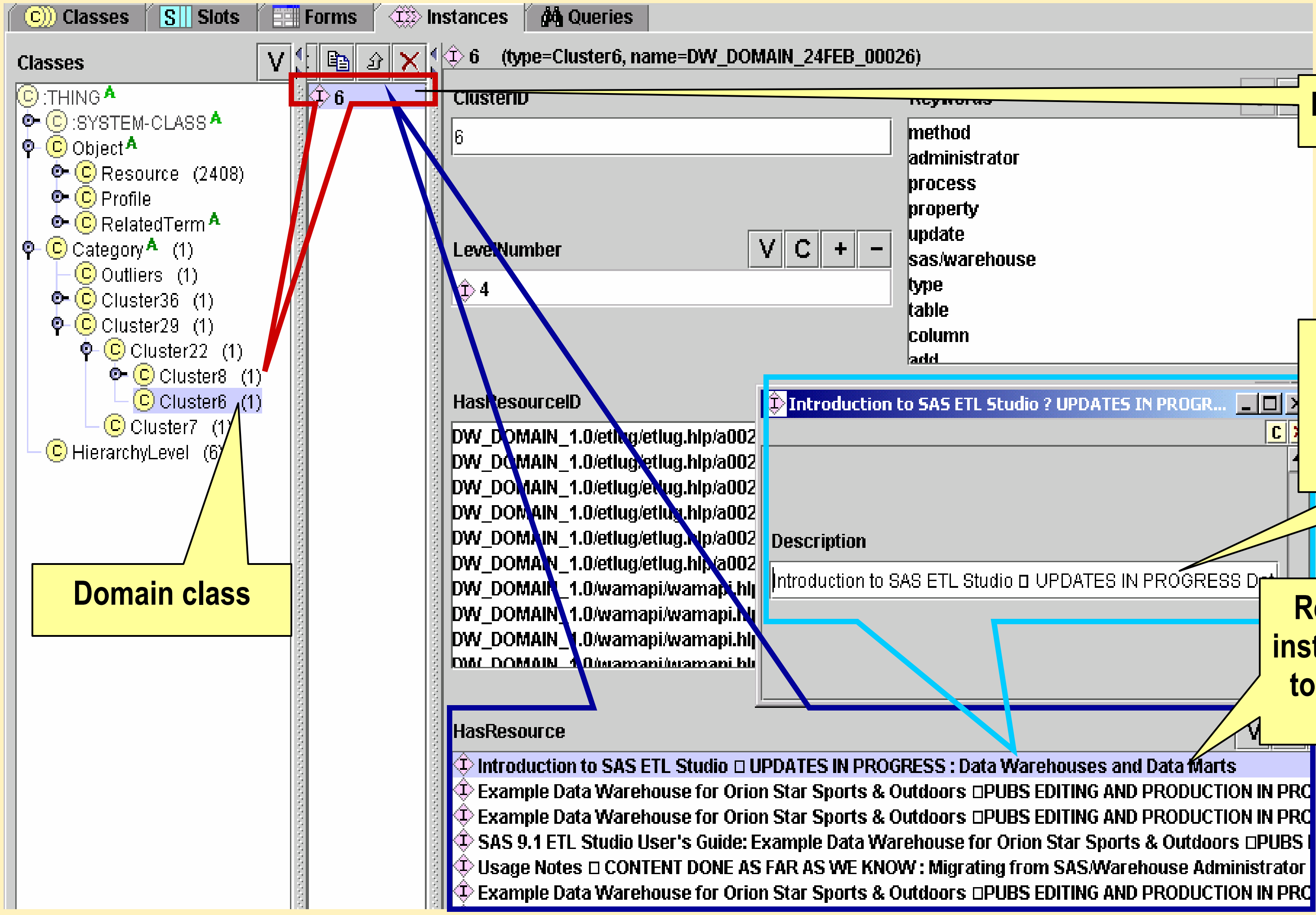
Map Categories to Resources

Specify path of Master Ontology

Select "Map Resources to Categories"

7th International Protégé Conference: July 7, 2004

Step 5: Resource Instances Mapped to Domain Instance Slots



6 (type=Cluster6, name=DW_DOMAIN_24FEB_00026)

ClusterID: 6

LevelNumber: 4

Keywords:
method
administrator
process
property
update
sas/warehouse
type
table
column
add

HasResourceID:
DW_DOMAIN_1.0/etlug/etlug.hlp/a002
DW_DOMAIN_1.0/etlug/etlug.hlp/a002
DW_DOMAIN_1.0/etlug/etlug.hlp/a002
DW_DOMAIN_1.0/etlug/etlug.hlp/a002
DW_DOMAIN_1.0/etlug/etlug.hlp/a002
DW_DOMAIN_1.0/etlug/etlug.hlp/a002
DW_DOMAIN_1.0/wamapi/wamapi.hl
DW_DOMAIN_1.0/wamapi/wamapi.hl
DW_DOMAIN_1.0/wamapi/wamapi.hl
DW_DOMAIN_1.0/wamapi/wamapi.hl

Introduction to SAS ETL Studio ? UPDATES IN PROGR...

Description: Introduction to SAS ETL Studio □ UPDATES IN PROGRESS D

HasResource:
Introduction to SAS ETL Studio □ UPDATES IN PROGRESS : Data Warehouses and Data Marts
Example Data Warehouse for Orion Star Sports & Outdoors □PUBS EDITING AND PRODUCTION IN PRO
Example Data Warehouse for Orion Star Sports & Outdoors □PUBS EDITING AND PRODUCTION IN PRO
SAS 9.1 ETL Studio User's Guide: Example Data Warehouse for Orion Star Sports & Outdoors □PUBS
Usage Notes □ CONTENT DONE AS FAR AS WE KNOW : Migrating from SAS/Warehouse Administrator
Example Data Warehouse for Orion Star Sports & Outdoors □PUBS EDITING AND PRODUCTION IN PRO

Domain instance

Resource Manager Description slot facilitates ontology refinement

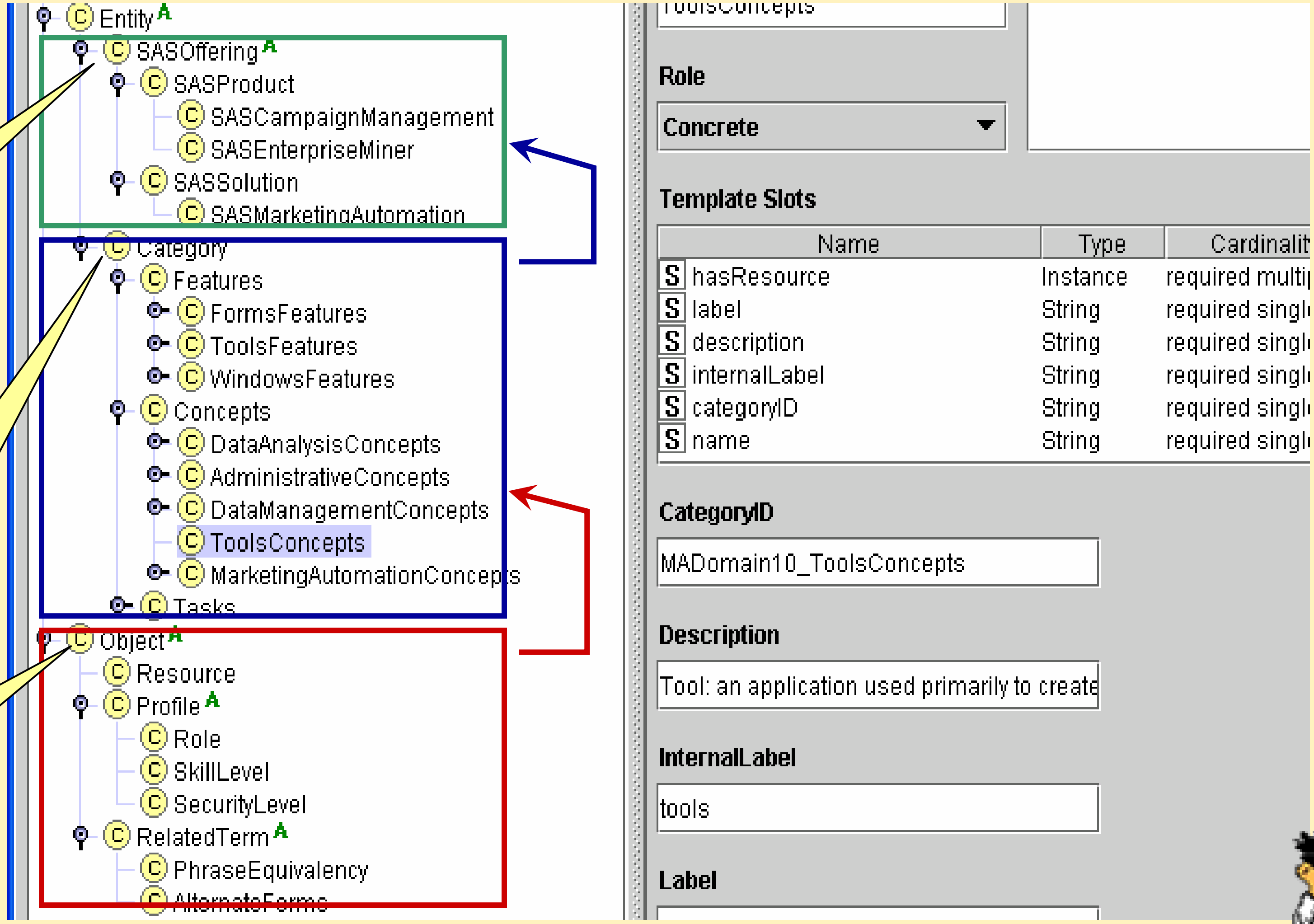
Domain class

Resource Manager instances are mapped to Domain instance slots

Step 6: Refine and Expand the Merged Ontology into a Production Master Ontology

- Add Related Terms to enable fuzzy matching of misspellings, synonymous phrases and alternative word forms
- Refine the Domain Hierarchy
  - Add the SAS Offerings Model
    - SAS products and solutions
  - Develop full ontology from Text Miner hierarchy
  - Add metaclasses to assign URIs at the class level

Step 6: Production Master Ontology



- Entity
  - SASOffering
    - SASProduct
      - SASCampaignManagement
      - SASEnterpriseMiner
    - SASSolution
      - SASMarketingAutomation
  - Category
    - Features
      - FormsFeatures
      - ToolsFeatures
      - WindowsFeatures
    - Concepts
      - DataAnalysisConcepts
      - AdministrativeConcepts
      - DataManagementConcepts
      - ToolsConcepts
      - MarketingAutomationConcepts
    - Tasks
  - Object
    - Resource
    - Profile
      - Role
      - SkillLevel
      - SecurityLevel
    - RelatedTerm
      - PhraseEquivalency
      - AlternateForms

ToolsConcepts

Role

Concrete

Template Slots

	Name	Type	Cardinalit
S	hasResource	Instance	required multi
S	label	String	required singl
S	description	String	required singl
S	internalLabel	String	required singl
S	categoryID	String	required singl
S	name	String	required singl

CategoryID

MADomain10_ToolsConcepts

Description

Tool: an application used primarily to create

InternalLabel

tools

Label

SAS Offerings Model

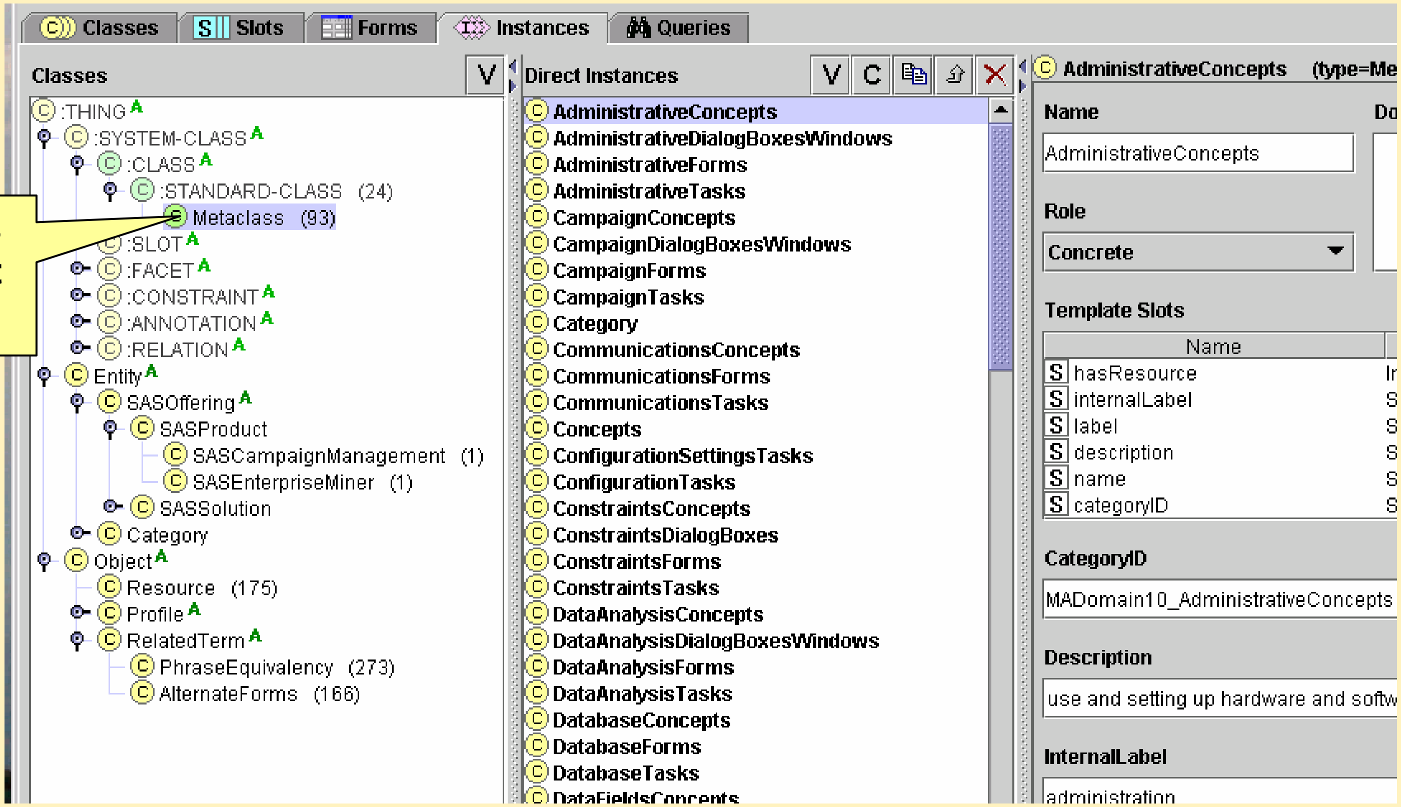
Domain Ontology Classes with edited, meaningful names

Resource Manager ontology classes

7th International Protégé Conference: July 7, 2004

Step 6: Add Metaclasses

Add Metaclasses to assign URIs at the class level



Classes

- :THING
- :SYSTEM-CLASS
  - :CLASS
    - :STANDARD-CLASS (24)
      - Metaclass (93)
  - :SLOT
  - :FACET
  - :CONSTRAINT
  - :ANNOTATION
  - :RELATION
- Entity
  - SASOffering
    - SASProduct
      - SASCampaignManagement (1)
      - SASEnterpriseMiner (1)
    - SASSolution
  - Category
- Object
  - Resource (175)
  - Profile
  - RelatedTerm
    - PhraseEquivalency (273)
    - AlternateForms (166)

Direct Instances

AdministrativeConcepts
AdministrativeDialogBoxesWindows
AdministrativeForms
AdministrativeTasks
CampaignConcepts
CampaignDialogBoxesWindows
CampaignForms
CampaignTasks
Category
CommunicationsConcepts
CommunicationsForms
CommunicationsTasks
Concepts
ConfigurationSettingsTasks
ConfigurationTasks
ConstraintsConcepts
ConstraintsDialogBoxes
ConstraintsForms
ConstraintsTasks
DataAnalysisConcepts
DataAnalysisDialogBoxesWindows
DataAnalysisForms
DataAnalysisTasks
DatabaseConcepts
DatabaseForms
DatabaseTasks
DataFieldsConcepts

AdministrativeConcepts (type=Me

Name: AdministrativeConcepts

Do

Role: Concrete

Template Slots

Name	
S hasResource	Ir
S internalLabel	S
S label	S
S description	S
S name	S
S categoryID	S

CategoryID: MADomain10_AdministrativeConcepts

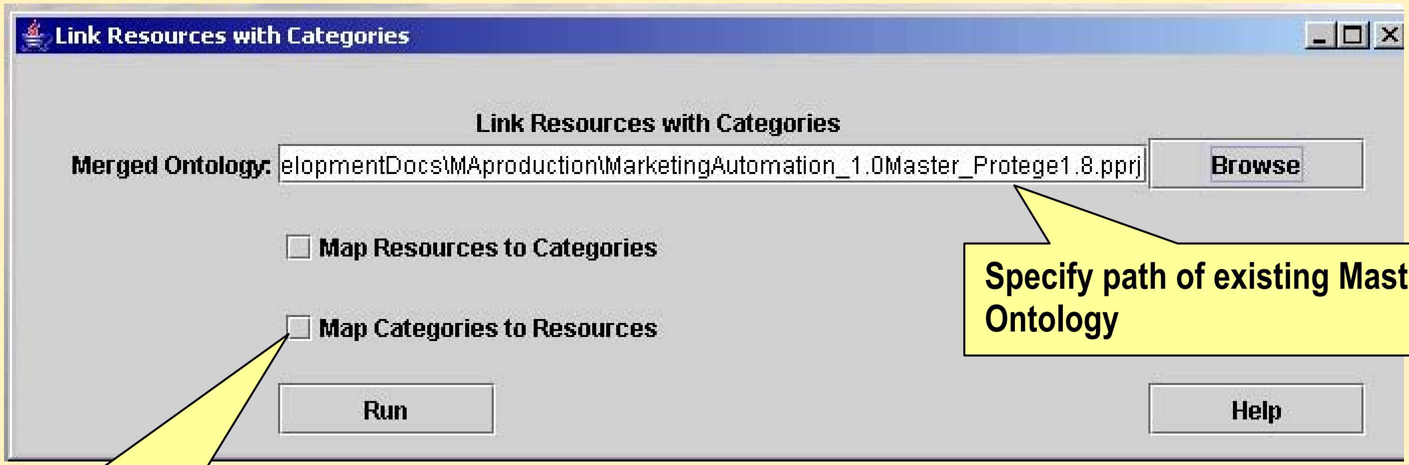
Description: use and setting up hardware and softw

InternalLabel: administration

7th International Protégé Conference: July 7, 2004

Step 7: Use a Custom Protégé Plugin to Map Domain Instances to Resource Instance Slots

Reverse mapping of Resource to Domain instance (Step 5)



Link Resources with Categories

Link Resources with Categories

Merged Ontology: elopmentDocs\MAproduction\MarketingAutomation_1.0Master_Protege1.8.pprj [Browse]

□ Map Resources to Categories

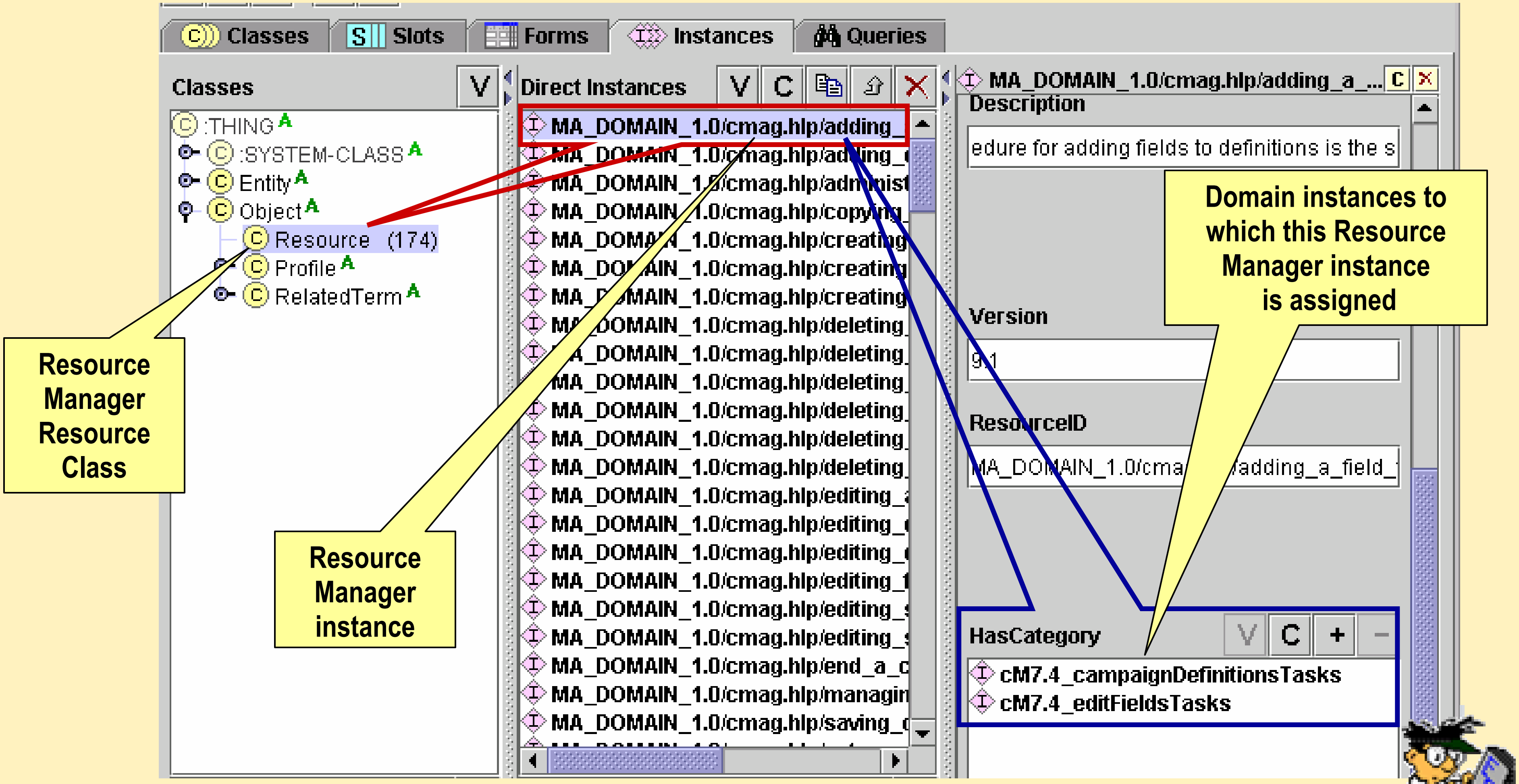
□ Map Categories to Resources

[Run] [Help]

Specify path of existing Master Ontology

Select “Map Categories to Resources”

Step 7: Domain Instances are Mapped to Resource Instance Slots



Classes

- :THING
  - :SYSTEM-CLASS
  - Entity
  - Object
    - Resource (174)
    - Profile
    - RelatedTerm

Direct Instances

- MA_DOMAIN_1.0/cmag.hlp/adding_
- MA_DOMAIN_1.0/cmag.hlp/adding_
- MA_DOMAIN_1.0/cmag.hlp/administ
- MA_DOMAIN_1.0/cmag.hlp/copying_
- MA_DOMAIN_1.0/cmag.hlp/creating
- MA_DOMAIN_1.0/cmag.hlp/creating
- MA_DOMAIN_1.0/cmag.hlp/creating
- MA_DOMAIN_1.0/cmag.hlp/deleting_
- MA_DOMAIN_1.0/cmag.hlp/deleting_
- MA_DOMAIN_1.0/cmag.hlp/deleting_
- MA_DOMAIN_1.0/cmag.hlp/deleting_
- MA_DOMAIN_1.0/cmag.hlp/deleting_
- MA_DOMAIN_1.0/cmag.hlp/deleting_
- MA_DOMAIN_1.0/cmag.hlp/editing_a
- MA_DOMAIN_1.0/cmag.hlp/editing_
- MA_DOMAIN_1.0/cmag.hlp/editing_
- MA_DOMAIN_1.0/cmag.hlp/editing_f
- MA_DOMAIN_1.0/cmag.hlp/editing_s
- MA_DOMAIN_1.0/cmag.hlp/editing_s
- MA_DOMAIN_1.0/cmag.hlp/end_a_c
- MA_DOMAIN_1.0/cmag.hlp/managin
- MA_DOMAIN_1.0/cmag.hlp/saving_c

MA_DOMAIN_1.0/cmag.hlp/adding_a_...

Description: edure for adding fields to definitions is the s

Version: 9.1

ResourceID: MA_DOMAIN_1.0/cmag.hlp/adding_a_field_

HasCategory:
- cM7.4_campaignDefinitionsTasks
- cM7.4_editFieldsTasks

Resource Manager Resource Class

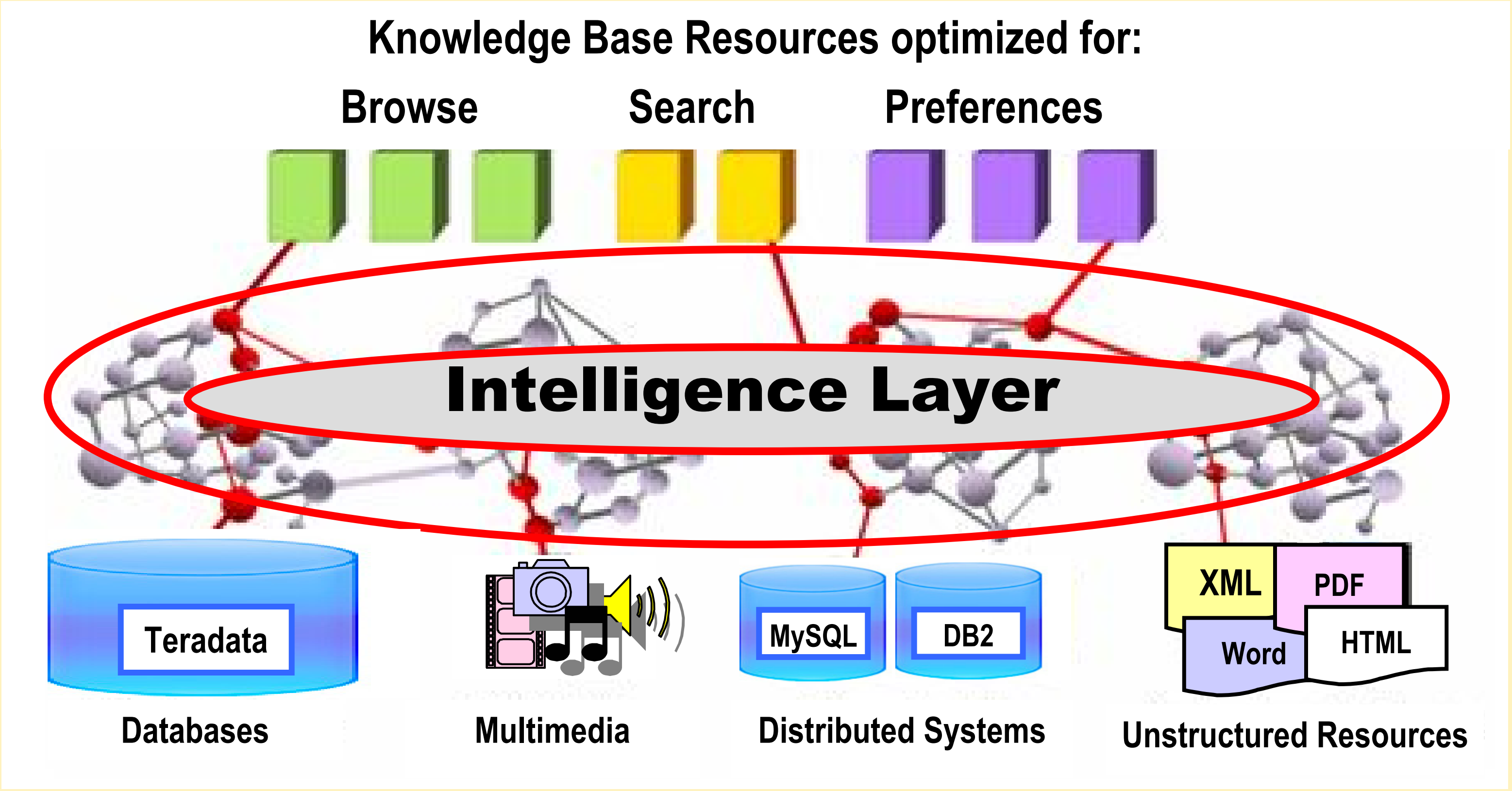
Resource Manager instance

Domain instances to which this Resource Manager instance is assigned

7th International Protégé Conference: July 7, 2004

The Power to Know.

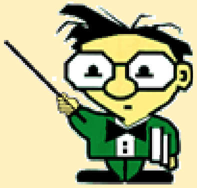
Ontologies Define the Intelligence Layer


Knowledge Base Resources optimized for:

Browse Search Preferences

Intelligence Layer

Databases Multimedia Distributed Systems Unstructured Resources

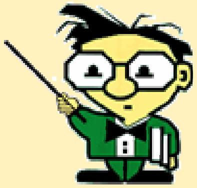


7th International Protégé Conference: July 7, 2004

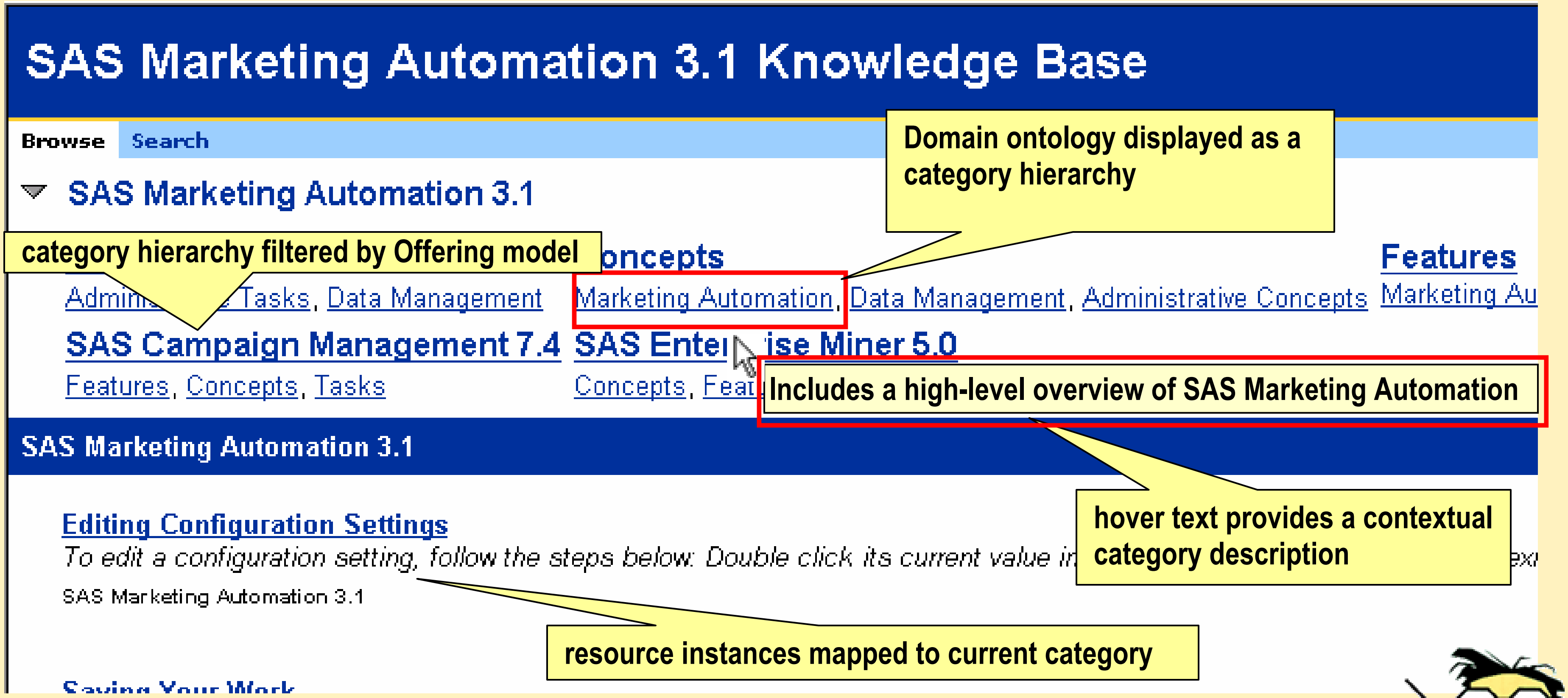
Knowledge Base Prototype

Delivers information in context using

- Browsable categories
- Categorized search results
- Hover text descriptions
- Category bread crumb trails
- Category and full text search
- Fuzzy matching



Browse View: Browsable Directory and Hover Text Contextual Cues



SAS Marketing Automation 3.1 Knowledge Base

SAS Marketing Automation 3.1

Admin… Tasks, Data Management

…oncepts

Marketing Automation, Data Management, Administrative Concepts

Features

Marketing Au…

SAS Campaign Management 7.4

Features, Concepts, Tasks

SAS Ente… ise Miner 5.0

Concepts, Feat…

Includes a high-level overview of SAS Marketing Automation

SAS Marketing Automation 3.1

Editing Configuration Settings

To edit a configuration setting, follow the steps below: Double click its current value i… ex…

SAS Marketing Automation 3.1

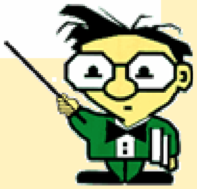
Saving Your Work

Domain ontology displayed as a category hierarchy

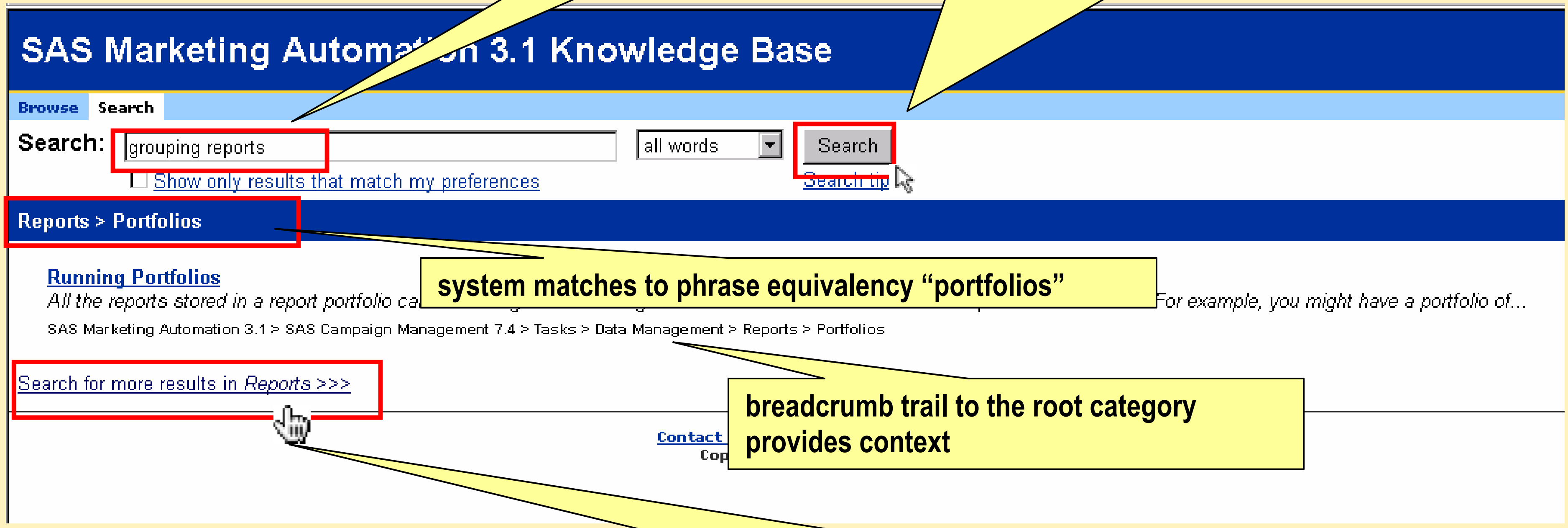
category hierarchy filtered by Offering model

hover text provides a contextual category description

resource instances mapped to current category



Search Results View: Search Expansion Fuzzy Match Synonymous Phrase



SAS Marketing Automa... 3.1 Knowledge Base

Search: grouping reports | all words | Search

□ Show only results that match my preferences

Search tip

Reports > Portfolios

Running Portfolios

All the reports stored in a report portfolio ca... For example, you might have a portfolio of...

SAS Marketing Automation 3.1 > SAS Campaign Management 7.4 > Tasks > Data Management > Reports > Portfolios

Search for more results in *Reports* >>>

Contact

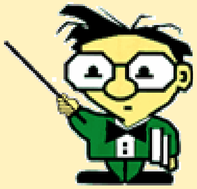
Cop...

enter phrase “grouping reports” and push the search button

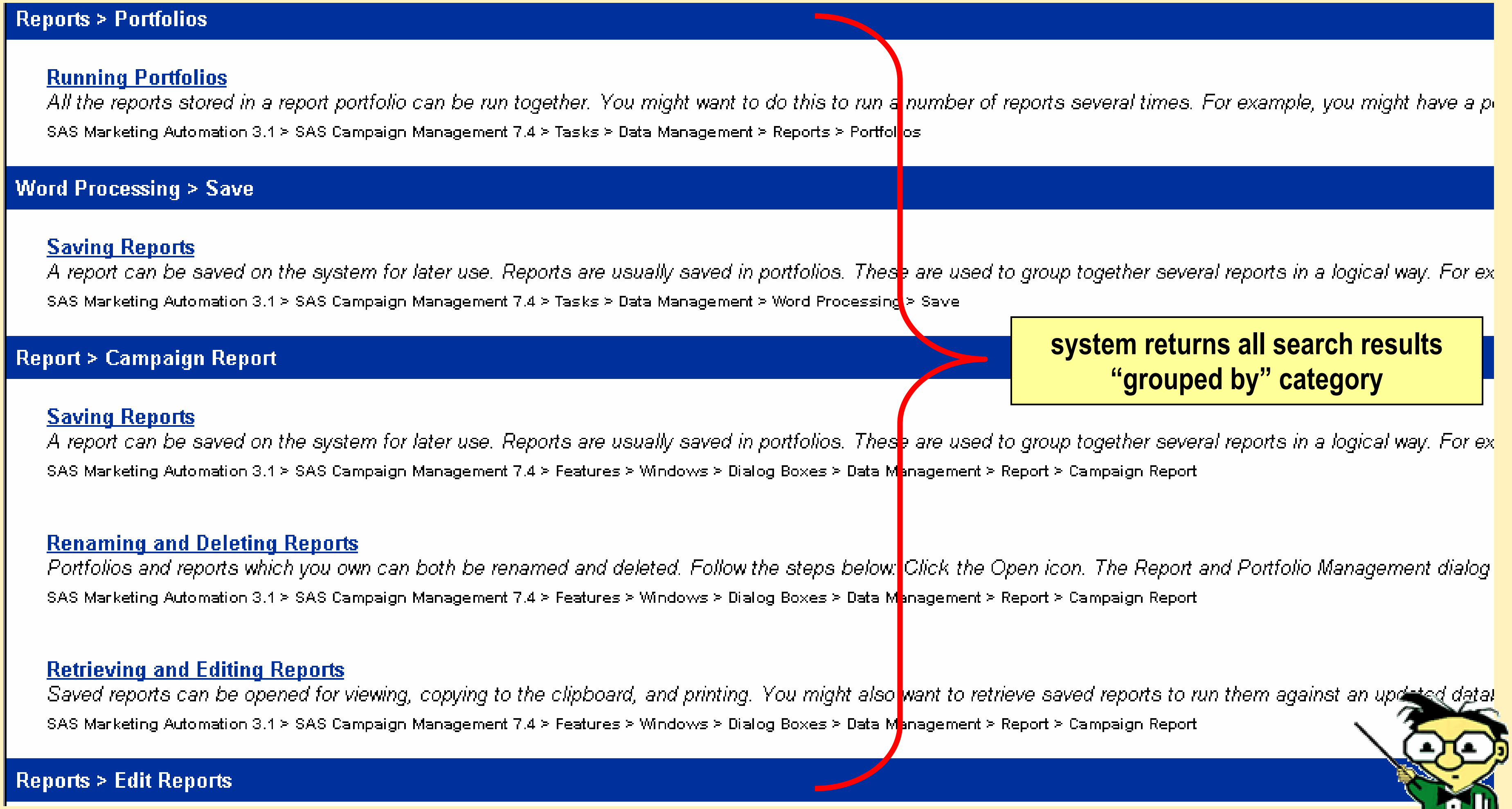
system matches to phrase equivalency “portfolios”

breadcrumb trail to the root category provides context

Optional full text search filtered by the current category



Search Results View: Contextual Cues
Grouped Results of Full Text Search



Reports > Portfolios

Running Portfolios

All the reports stored in a report portfolio can be run together. You might want to do this to run a number of reports several times. For example, you might have a p

SAS Marketing Automation 3.1 > SAS Campaign Management 7.4 > Tasks > Data Management > Reports > Portfolios

Word Processing > Save

Saving Reports

A report can be saved on the system for later use. Reports are usually saved in portfolios. These are used to group together several reports in a logical way. For ex

SAS Marketing Automation 3.1 > SAS Campaign Management 7.4 > Tasks > Data Management > Word Processing > Save

Report > Campaign Report

Saving Reports

A report can be saved on the system for later use. Reports are usually saved in portfolios. These are used to group together several reports in a logical way. For ex

SAS Marketing Automation 3.1 > SAS Campaign Management 7.4 > Features > Windows > Dialog Boxes > Data Management > Report > Campaign Report

Renaming and Deleting Reports

Portfolios and reports which you own can both be renamed and deleted. Follow the steps below: Click the Open icon. The Report and Portfolio Management dialog

SAS Marketing Automation 3.1 > SAS Campaign Management 7.4 > Features > Windows > Dialog Boxes > Data Management > Report > Campaign Report

Retrieving and Editing Reports

Saved reports can be opened for viewing, copying to the clipboard, and printing. You might also want to retrieve saved reports to run them against an updated data

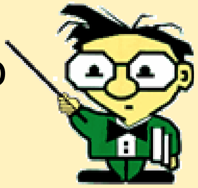
SAS Marketing Automation 3.1 > SAS Campaign Management 7.4 > Features > Windows > Dialog Boxes > Data Management > Report > Campaign Report

Reports > Edit Reports

system returns all search results "grouped by" category

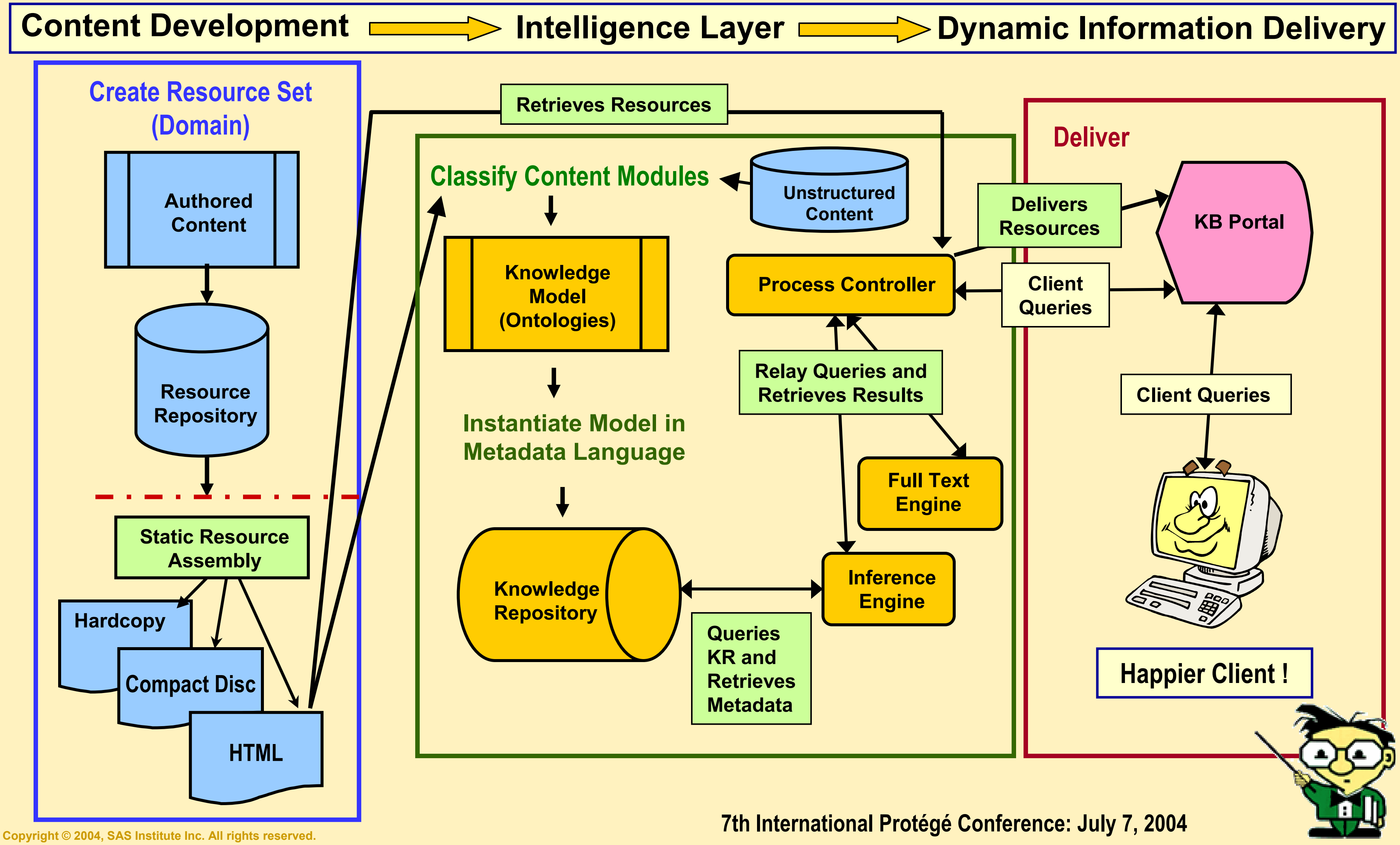
Our Vision for an Integrated Solution: The Value of XML-Based Modular Content

- ***Why XML?***
  - Accepted standard-already used by a critical mass
  - Facilitates platform independent interoperability
  - Provides a content development framework that supports modular writing
- ***Why Modular Writing?***
  - Reusability
    - Controls work redundancy
    - Reduces semantic heterogeneity
      - using the same terminology to mean different things
      - using different terminology to mean the same thing
  - Facilitates content classification: “about” one thing
  - Enables advanced information retrieval and delivery techniques
    - dynamic assembly of complex resources that are relevant to a user’s current context



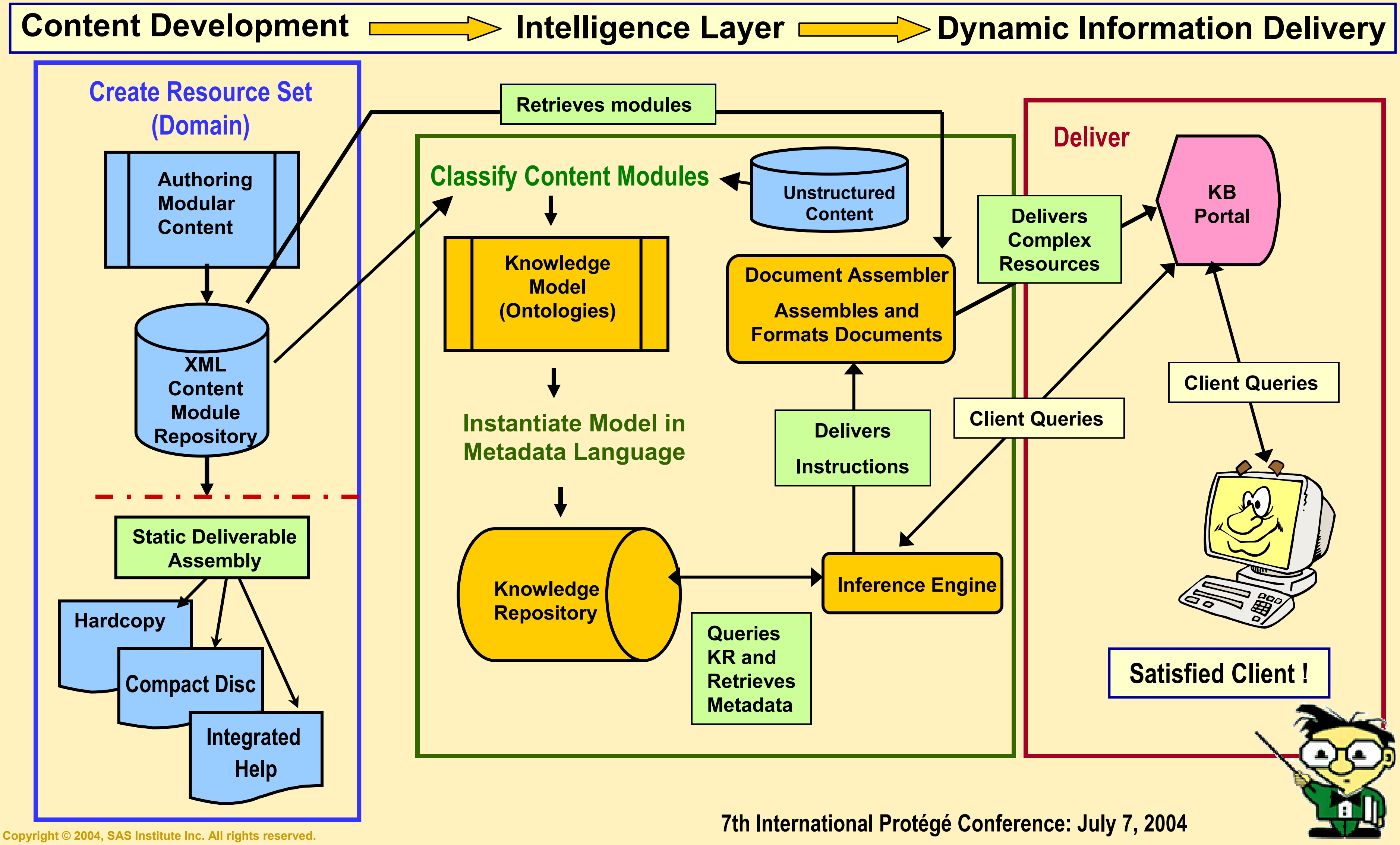
7th International Protégé Conference: July 7, 2004

Knowledge Base System Architecture



7th International Protégé Conference: July 7, 2004

One Vision for an Integrated Solution


Content Development
Intelligence Layer
Dynamic Information Delivery
Create Resource Set (Domain)
Authoring Modular Content
XML Content Module Repository
Static Deliverable Assembly
Hardcopy
Compact Disc
Integrated Help
Retrieves modules
Classify Content Modules
Unstructured Content
Knowledge Model (Ontologies)
Instantiate Model in Metadata Language
Knowledge Repository
Document Assembler
Assembles and Formats Documents
Delivers Instructions
Inference Engine
Queries KR and Retrieves Metadata
Delivers Complex Resources
Client Queries
Deliver
KB Portal
Client Queries
Satisfied Client !

7th International Protégé Conference: July 7, 2004

Questions/Comments?

sas

The Power to Know.

Thanks to Contributors: Dee Stribling and Chris Goolsby