

What Are Business Taxonomies? Why Use Them?

ARMA International
New Orleans, Louisiana, USA
Sept. 29 – Oct. 2, 2002

Denise Bruno, MLS Heather Richmond, CRM





Agenda

- **▲What**
- **▲When**
- **▲Where**
- **▲Why**
- **▲Who**







Objectives

- ▲To define a business taxonomy and its concepts, and to explain why it is important to have one.
- ▲To outline a list of steps to take to develop a business taxonomy and the pitfalls to avoid along the journey.
- **▲**To discuss how taxonomies and records management impact each other.





"The value of knowledge is largely tied to the way in which that knowledge is organized. If you can't find it, it's not likely to be of much use to you."

Marc Rapport
"Unfolding Knowledge"
Knowledge Management E-zine





Exercise

- **▲ Divide into groups**
- **▲** Pass out envelopes
- **▲**Go over directions
- **▲**Re-group after 7-8 minutes
- **▲** Group presentation





What company do you think that was?





idea noun thinking Definitions



Taxonomy (1)

- **▲** "Taxis" arrange, put in order
- **▲**"Onoma" name
- ▲Is the end result of the science, laws, or principles of classification





Taxonomy (2)

"Structures that provide a way of classifying things – living organisms, products, books – into a series of hierarchical groups to make them easier to identify, study, or locate. Taxonomies consist of two parts – structures and applications. Structures consist of the categories (or terms) themselves and the relationships that link them together. Applications are the navigation tools available to help users find information."

Jean Graef Montague Institute





Controlled Vocabulary

- ▲ An indexing language, i.e., a standardized set of terms and phrases authorized for use in an indexing system to describe a subject area or information domain.
- ▲ A collection of preferred terms that are used to assist in more precise retrieval of content.





Thesaurus

- ▲ A type of controlled vocabulary that shows the hierarchical (e.g. parent-child), associative (e.g. related) and equivalent (e.g. synonymous) relationships among terms
- **▲**Controlled vocabulary, thesaurus and classification structure (taxonomies) are often used interchangeably.





User Warrant

▲ Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system





"...the primary motives for developing an internal taxonomy were to improve information access and to save time by streamlining the search process."

"Taxonomies for Business: Access and Connectivity in a Wired World", TFPL Ltd.





Information Environment

- **▲** Paper
- **▲** Facsimiles
- **▲ Electronic docs**
- **▲** Email
- **▲ Chat boards**
- **▲White boards**
- ▲ Legacy databases

- **▲Instant messaging**
- **▲ Electronic forms**
- **▲** Workflow
- ▲ Video & audio
- ▲Intra- & internet materials
- **▲**Microforms





Information Environment

- ▲ No standards for info design or else too vague or incapable of being enforced
- ▲ Separate offices/divisions, many with own IT shops, build separate info systems
- **▲**Cultures of competitiveness or mistrust
- **▲**Legacy systems difficult to change
- **▲** Managers still looking for silver bullet





Value of Taxonomies

- **▲**Control the glut of information
- ▲ Filter, categorize and label information
- ▲ Aid discovery the serendipitous find, new associations
- ▲Improve retrieval, especially through controlled vocabularies
- **▲**Enhance full text search





What Can They Do?

▲Identification

◆Is this a service we can provide? Is there a product that will do this? Have we solved this problem before? Does this type of information exist?

▲ Discovery

What's related to the issue or problem at hand? Who is working on it?

▲ Delivery

◆If something is available, where or how do I get it?





"Creating a structure for internal information helps both to cut down on information overload and to speed recovery of critical knowledge."

Rebecca Lloyd "Building Better Knowledge Maps", Knowledge Management e-zine





Hierarchical Relationships

▲ Generic (Genus/Species)

- ◆ Inclusiveness
- ◆ Genus/Species
- ◆ Inheritance
- **♦** Transitivity
- Systematic & predictable rules for association & distinction
- Mutual exclusivity

- ◆Eye Diseases (from MeSH)
 - → Conjunctival Diseases
 - → Conjunctival Neoplasm
 - → Conjunctivitis
 - Keratoconjunctivitis
 - → Corneal Diseases







Hierarchical Relationships

▲ Whole-part (Tree)

- Systematic & predictable rules for distinction
- Does not assume genus/species
- One-way flow of information

- **◆** Automobile
 - **⇒**Body
 - **→**Engine Block
 - → Pistons
 - → Valves
 - **→**Interior
 - → Upholstery





Difference from Classification?

- **▲** More fashionable, chic!
- ▲ Just a different word for the same concept?
- ▲ Structure, content & application based
- ▲ Customized to reflect language, culture & goals of a particular enterprise
- **▲** Use different kinds of information resources
 - ◆ E-mails, memos, people & documents, books, reports & web pages.
- ▲ Created by multi-disciplinary teams
- ▲ Part of a process





Types of Taxonomies

- **▲**Function
- **▲** Department
- **▲**Subject
- **▲**Product/Service
- **▲** Location









Is One Taxonomy Enough?

- **▲** Discrete bodies of knowledge
- **▲Global locations**
- **▲** Personalization
- **▲**Functional vs. Department view







"Though figuring out where to start can be frustrating, a good taxonomy is recognized as a central part of a knowledge management system."

Thomas Trimmer President, GrapeVine Technologies





Taxonomy Development

- ▲ Research revealed that the enterprises were aware of:
 - ◆ The need to develop a better information structure
 - Were prepared to commit substantial financial resources to the project
 - Understood the importance of using a standardized terminology
 - Recognized that user participation and feedback was important

Alan Gilchrist

"The Corporate Taxonomy – The Latest Tool in the Battle Against Information Overload"





Taxonomy Development

1. Planning and Data Gathering

- Define the stakeholders
- Define the goals, the scope, the rules
- Create a communication plan
- Conduct data gathering









Taxonomy Development

2. Build Draft Taxonomy

- Confirm type of taxonomy to use
- Research existing taxonomies
- If using application (e.g. EDMS, categorizing software), know capabilities/limitations of it





Taxonomy Development

2. Build Draft Taxonomy (cont.)

- Top-down vs. Bottom-up approach
- Begin to group docs and nest groups
- Begin to build CV, thesaurus
- Present draft to interviewees







Taxonomy Development

2. Build Draft Taxonomy (cont.)

- Refine hierarchy
- Seek advice from subject experts if needed
- Present draft to validation committee





3. Pilot

- Select one area as pilot or prototype and populate with sample docs
- Test with users





4. Refine and Finalize

- Refine hierarchy structure and bucket names
- Launch it in its entirety
- Market the launch





5. User Training

- Reiterate the value of the taxonomy
- Customize sessions





6. Ensure Continued Development

- Staffing
- Mechanism for updating

PROJECT → PROCESS





Role of RM in Taxonomy Development

- **▲**Part of the team
- **▲**Classification expertise
- **▲** Knowledge of organization
- **▲**Understanding of business processes
- **▲**Understanding of business terminology
- ▲ Appreciation for the need of information retrieval





Taxonomy Example

- **▲** Collection
 - **♦**Part
 - ⇒ Section
 - → Primary
 - Secondary
- **▲** Collection 2: ABC Company Management
 - ◆Part 4: Financial Management
 - ⇒ Section 05: Financial Reporting and Auditing
 - → Primary 03: Audit Working Papers (2-4-05-03)
 - Secondary 01: Audit Confirmations (2-4-05-03-01)





Lotus Domino.doc

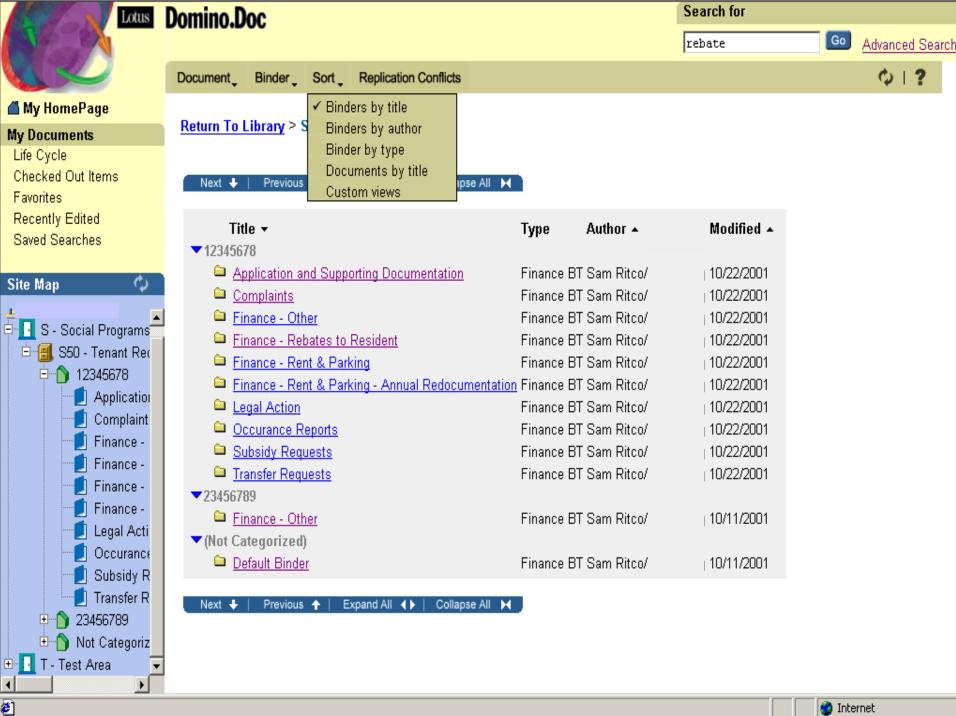
▲Library

- ◆File Room
 - **→** File Cabinet
 - → Binder Categories
 - Binders

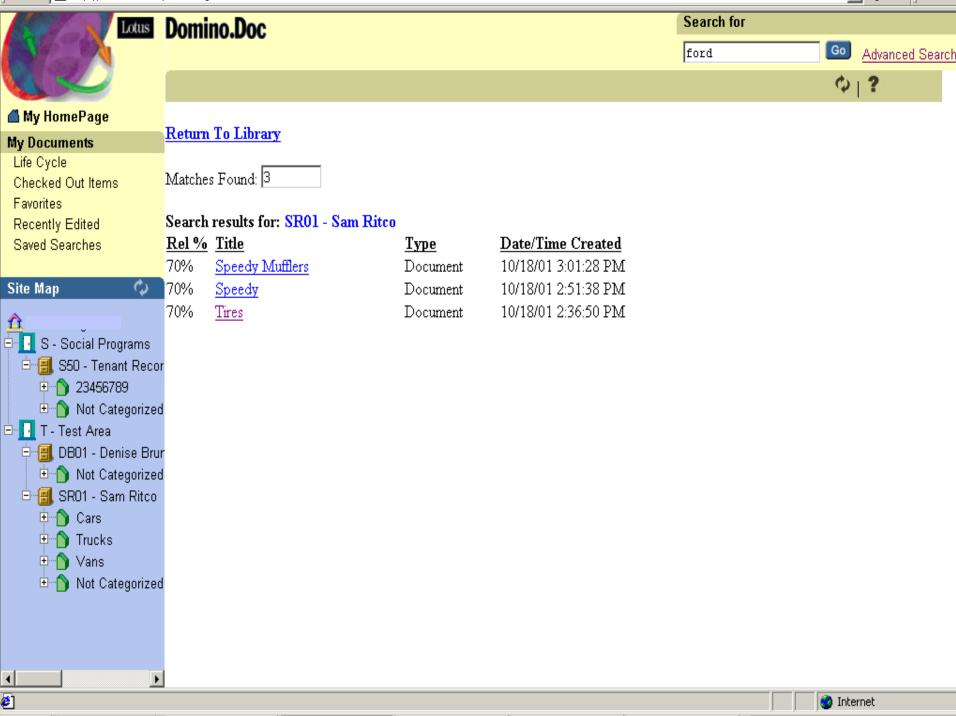
▲ Pilot Library

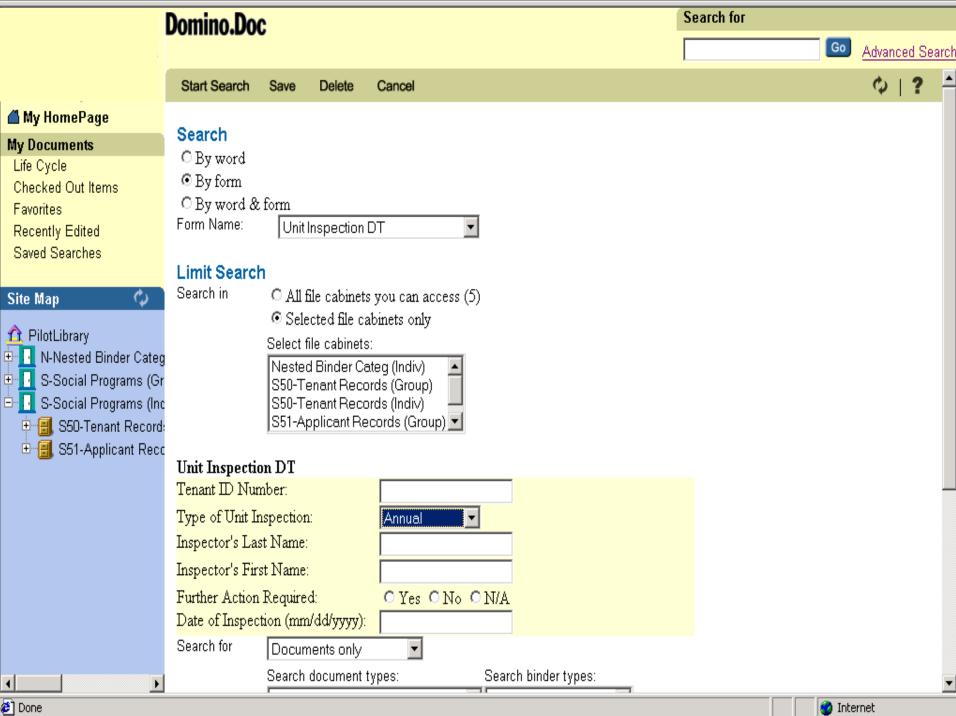
- ◆S Social Programs
 - ⇒S50 Tenant Records
 - **→** 12345678
 - Complaints
 - Finance Rent & Parking
 - Subsidy Requests

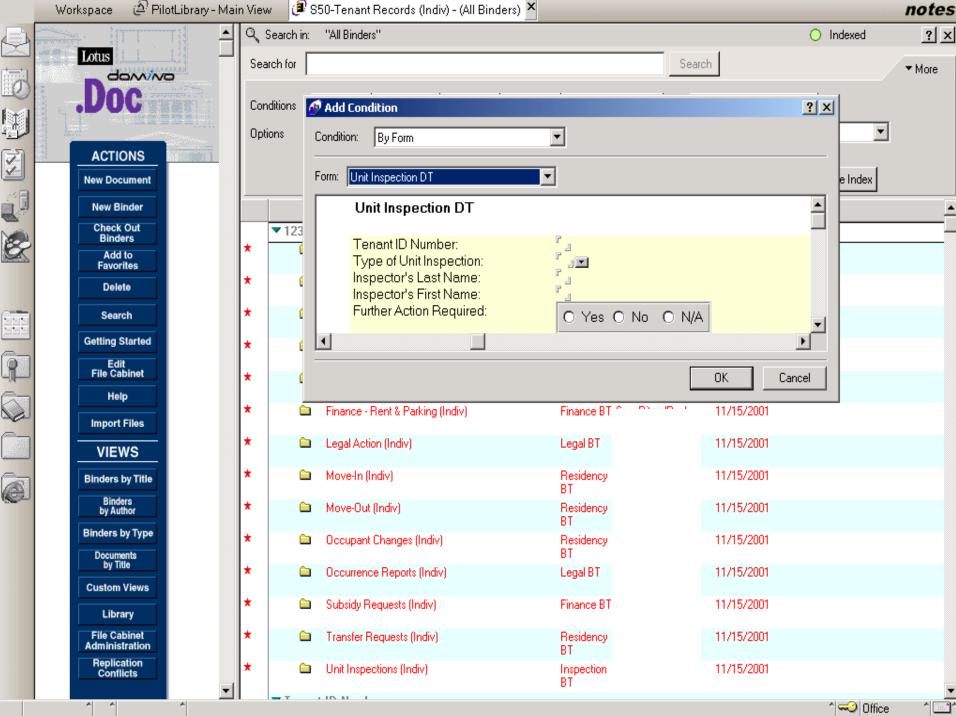


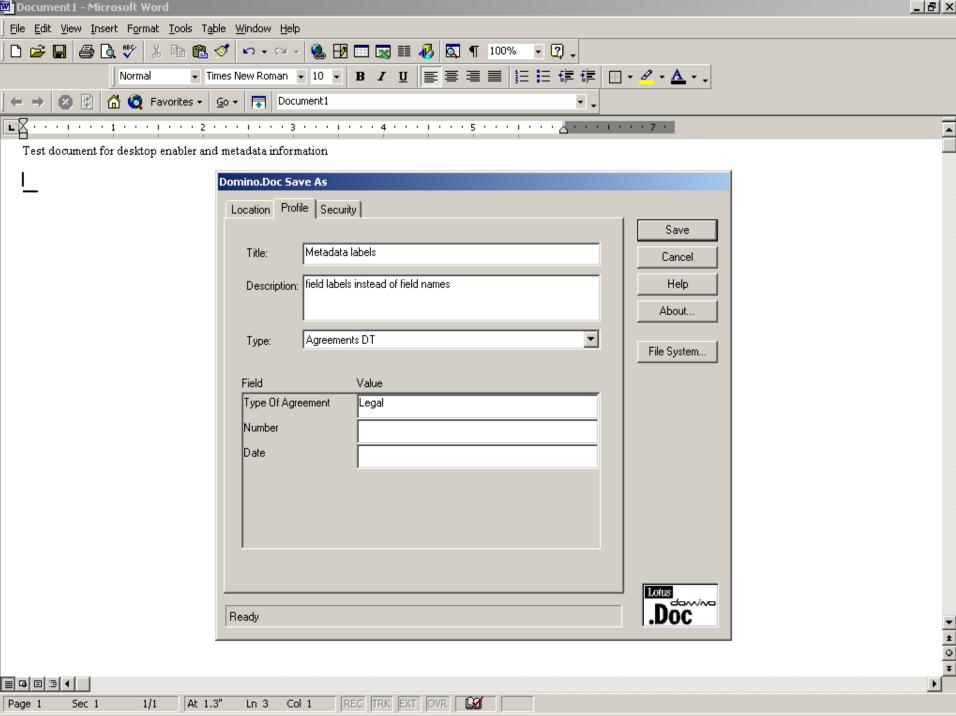


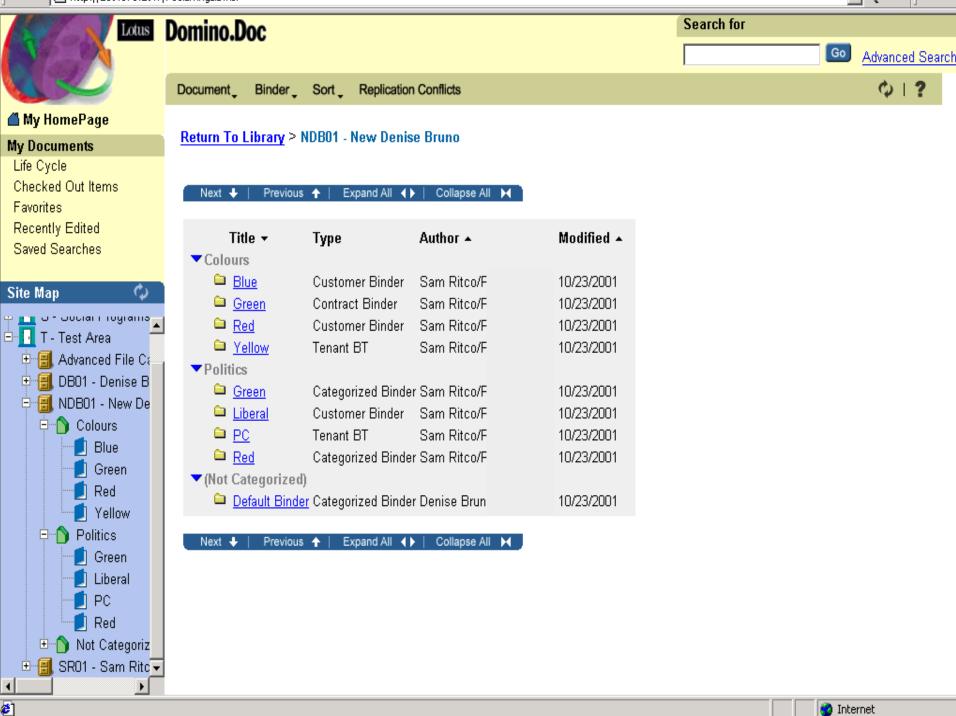


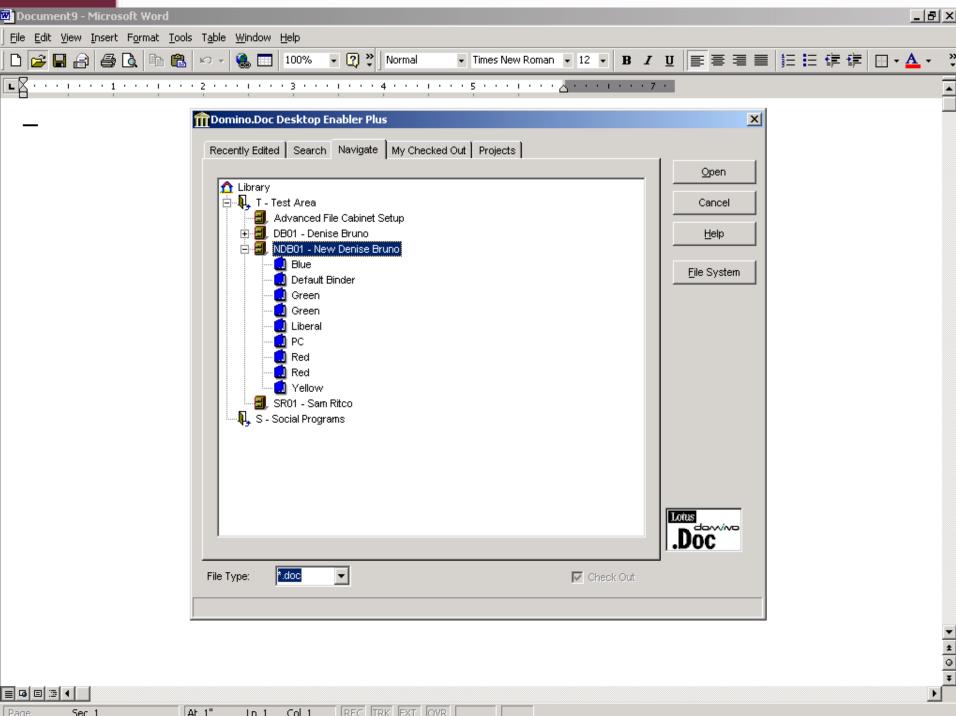














Automatic Categorization Software

- ▲ Automatic categorization software provides the potential means to automatically file documents into either a predefined taxonomy or self-defining categories.
- ▲It is the process by which technology is used to create clusters





Automatic Categorization Software

- **▲ Rules based (Semio, Inxight)**
 - ◆If-then statements define the clusters
 - **♦**Uses linguistics
 - ◆Not dependent on information in the collection
- **▲ Statistical clustering (Autonomy, Mohomine, Hummingbird)**
 - Employs mathematical algorithms
 - ◆Dependent upon information in the collection





CONDAR CONSULTING INC.



Automatic Categorization Software

▲ Catalogue-by-example (Autonomy, Inxight)

- Learns from previously catalogued examples
- "Training set" of exemplary documents
- ◆Iterative process





Automatic Categorization Software

▲ Advantages:

- Perceived cost savings, fewer resources
- Minimizes time to complete classification process
- ◆Identifies unknown associations
- Ability to handle large unstructured databases
- Frees humans from routine tasks





Automatic Categorization Software

▲ Disadvantages:

- Limited accuracy
- Short docs, very large docs, many subject docs
- Often requires large collections
- Potentially misleading
- Security issues with existing systems





CONDAR CONSULTING INC.



UK Government Category List

Top Two Levels

Agriculture, environment and natural resources

Energy and fuel

Environmental protection

Farming

Fisheries

Forestry

Horticulture

Land

Plants and animals

Waste management

Water resources

Arts, recreation and travel

Culture and creativity

Gambling and lotteries

Parks and gardens

Passports and visas

Sports and recreation

Tourism

Transport

Business and industry

Business practice and regulation

Business sectors

Companies

Consumer affairs

Energy and fuel

International trade

Transport

Education, careers and employment

Career development

Education and skills

Employment relations

Equal opportunities

Occupational health and safety

Unemployment and jobseeking

Working hours, terms and conditions

Government, politics and public administration

Central government

Civil Service

Constitution

Devolved administrations

Electoral system

Honours system

Local government

Political parties

Public administration

Regional policy

Health, nutrition and care

Animal health

Benefits

Care

Carers and health professionals

Disabled people

Family planning

Food and drink

YY 1.





UK Government Category List

Complete List

Agriculture, environment and natural resources

- Energy and fuel
- Energy conservation
- Fossil fuels
- Nuclear energy
- Renewable energy
- Environmental protection
- Climate and weather
- Coastal erosion and protection
- Environmental monitoring
- Nature conservation
- . . Pollution
- Farming
- Arable farming
- Common Agricultural Policy
- . . Crops
- . . . Genetically modified food and crops
- Livestock farming
- Fisheries
- . . Common Fisheries Policy
- Forestry
- Horticulture
- Common Agricultural Policy
- Land
- Agricultural land
- Countryside
- . . Access to countryside
- . . Flooding

- Visual arts
- Gambling and lotteries
- National Lottery
- Parks and gardens
- Passports and visas
- Sports and recreation
- . Athletics
- . Football
- . Hunting
- Water sports
- . . Winter sports
- Tourism
- Transport
- Air transport
- Freight transport
- . . Inland waterways
- Ports and maritime transport
- Public transport
- Rail transport
- Roads and road transport
- . . Road safety

Business and industry

- Business practice and regulation
- . Advertising
- . . Business development
- Business management
 - Charities



UK GCL Index (ISO Standard)

Annual leave Working hours, terms and conditions See: Anti-social behaviour and disorder IIF : Football hooliganism Football violence Hoolidanism Vandalism BT: Crime Apprenticeships Adult education and training See: Arable farming UF: Cereal production Wheat growing BT: Farming RT: Crops Plant health Archaeology See: Heritage Architecture UF: Churches (architecture) BT: Culture and creativity RT: Building and construction

Heritage





Semio Software

!Farming +agriculture +farming +hill farming +intensive farming +organic farming -arable -cattle -dairy -fur -livestock -pig -poultry -sheep !Arable Farming +arable farming +cereal production +wheat growing !Common Agricultural Policies +cap +common agricultural policy !Crops +crop -genetically modified food and !Genetically Modified Food And Crops +genetically modified food and crop !Livestock Farming +animal breeding





Agriculture, Environment And Natural Resources

Energy And Fuels, Farming, Fisheries, ...

Business And Industries

Business Practice And Regulation, ...

Economics And Finance

Capital And Financial Markets, Taxs, ...

Government, Politics And Public Administration

Central Governments, Civil Services, ...

Information And Communication

Communications Industries, Intelligences, ...

People, Communities And Living

Benefits, Charities, Death And Funerals, ...

Arts, Recreation And Travels

Culture And Creativities, Tourisms, ...

Crime, Law, Justice And Rights

Animal Rights And Welfares, Crimes, Laws,

Education, Careers And Employments

Education And Skills, ...

Health, Nutrition And Cares

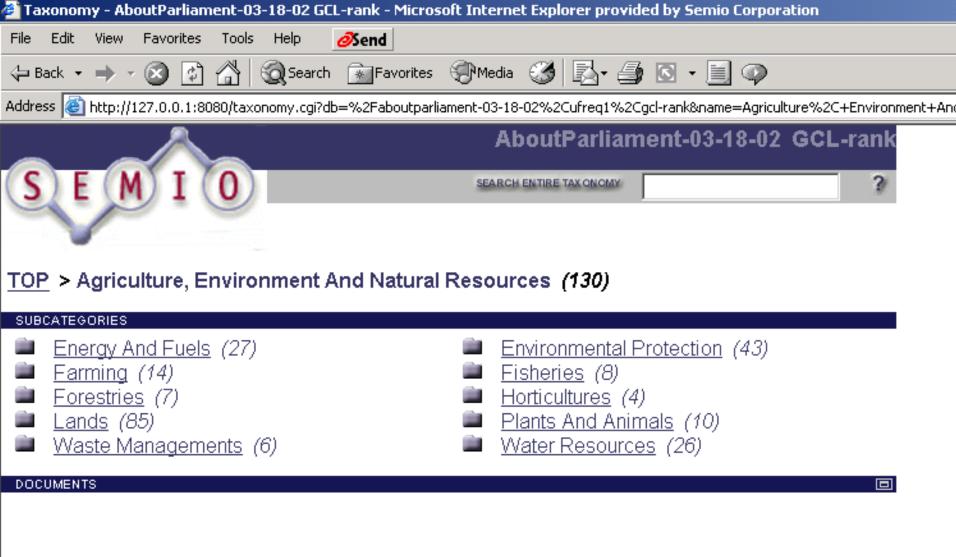
Animal Healths, Benefits, Cares, Healths, ...

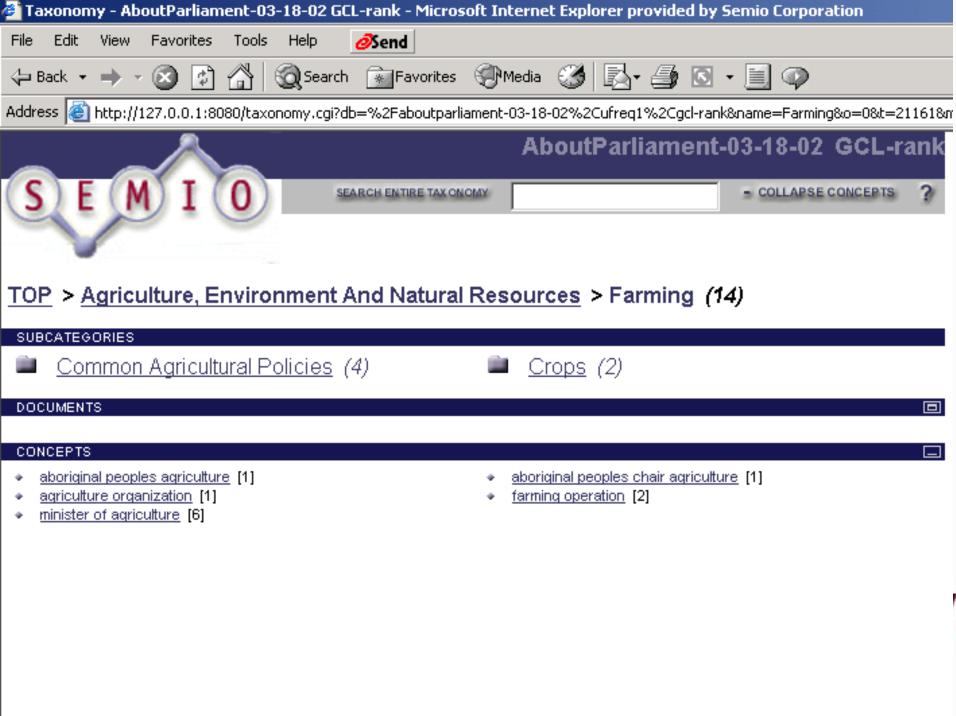
International Affairs And Defences

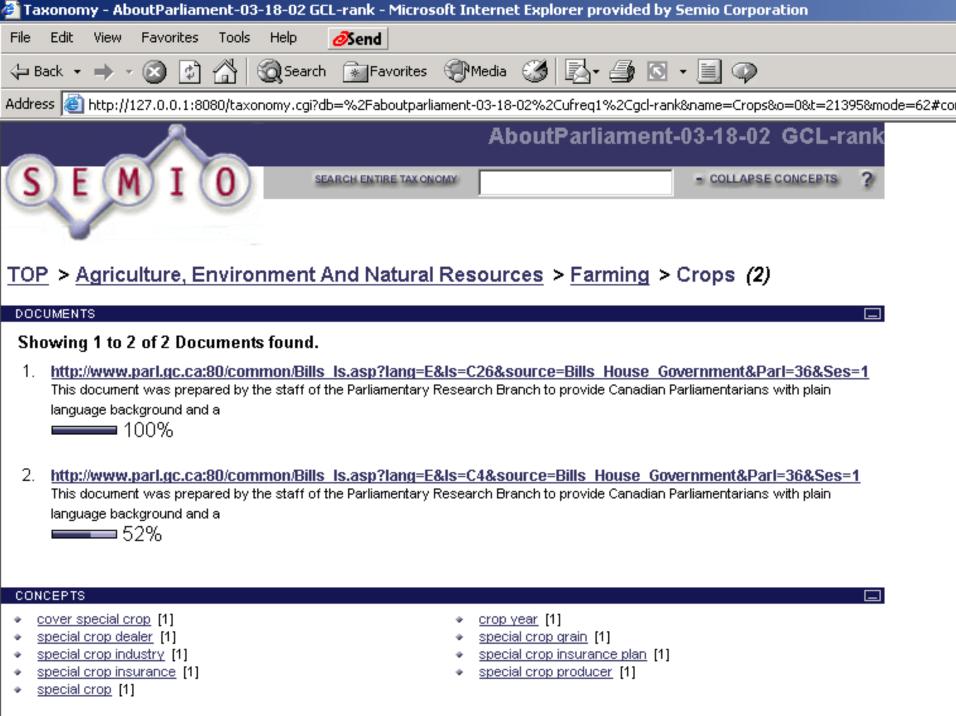
Defences, Embassies And Consulates,

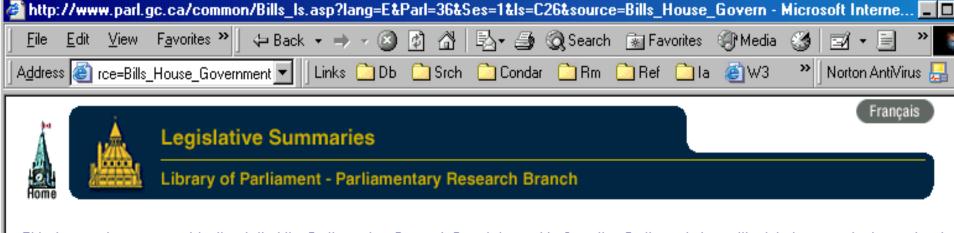
Science, Technology And Innovation

<u>Information Technologies,</u> ...









This document was prepared by the staff of the Parliamentary Research Branch to provide Canadian Parliamentarians with plain language background and analysis of proposed government legislation. Legislative summaries are not government documents. They have no official legal status and do not constitute legal advice or opinion. Please note, the Legislative Summary describes the bill as of the date shown at the beginning of the document.

BILL C-26: AN ACT TO AMEND THE CANADA GRAIN ACT AND THE AGRICULTURE AND AGRI-FOOD ADMINISTRATION MONETARY PENALTIES ACT AND TO REPEAL THE GRAIN FUTURES ACT

Prepared by Daniel Shaw Economics Division

> 30 January 1998 Revised 1 May 1998



BACKGROUND

DESCRIPTION AND ANALYSIS

- A. The Special Crops Producer Insurance Plan and Dealer Licensing Regime
 - Definitions
 - 2. Special Crops Dealers' Licences and Security Exemption
 - 3. Special Crops Insurance Plan and its Administration
 - 4. Special Crops Dealer Obligations
 - 5. Connecting the Act to the AAAMPA
 - 6. Other Matters
- B. Enforcement of the Canada Grain Act
- C. The Grain Futures Act

<u>COMMENTARY</u>



Automatic Categorization Software

▲Things to consider:

- Not truly taxonomy software
- Use humans and software
- Require a controlled vocabulary or thesaurus
- ◆Run software on custom data set





CONDAR CONSULTING INC.



"Too much faith is placed in technology – companies think that by purchasing technology they're purchasing a classification system. People have to be involved in the solution or it will never work"

Glenn Kelman Vice President Marketing, Plumtree Software





Managing Taxonomies Strategically

- **▲**Prepare for people issues
- ▲ Serve the real needs of users than produce an "ideal" textbook taxonomy
- ▲ Educate stakeholders about the real story of categorization software
- ▲ Plan for evolution and growth
- ▲It is part of a process not just a product





Case Study Exercise

▲Review case study

▲ Answer/discuss questions within

your group

▲ Discussion







Last Word

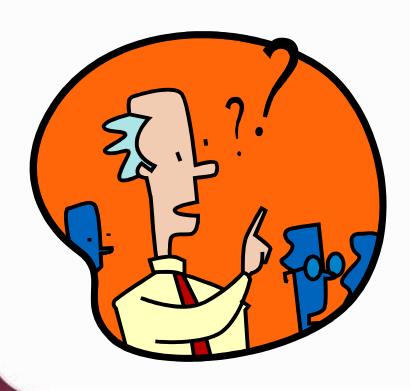
"Good Taxonomies, based on the use of classification and controlled vocabularies, result in more efficient information retrieval. This ensures better productivity and less user frustration. Where do taxonomies fit into the information architecture paradigm? They are one part of it, though they may not always be referred to as a taxonomy."

Liz Edols "Taxonomies Are What?"





Questions & Answers



Email

Heather

hrichmond@condar.ca

Denise

denisebruno@sympatico.ca

www.condar.ca





The End Thank you!

CONDAR CONSULTING INC.