

THE ACCIDENTAL TAXONOMIST

Luděk Janda

Product documents librarian

PnT Portal, Red Hat

Sept-17 2012

Agenda

- Who I am
- Introduction to taxonomy
- Development of a taxonomy
- Testing & Maintenance
- Q&A

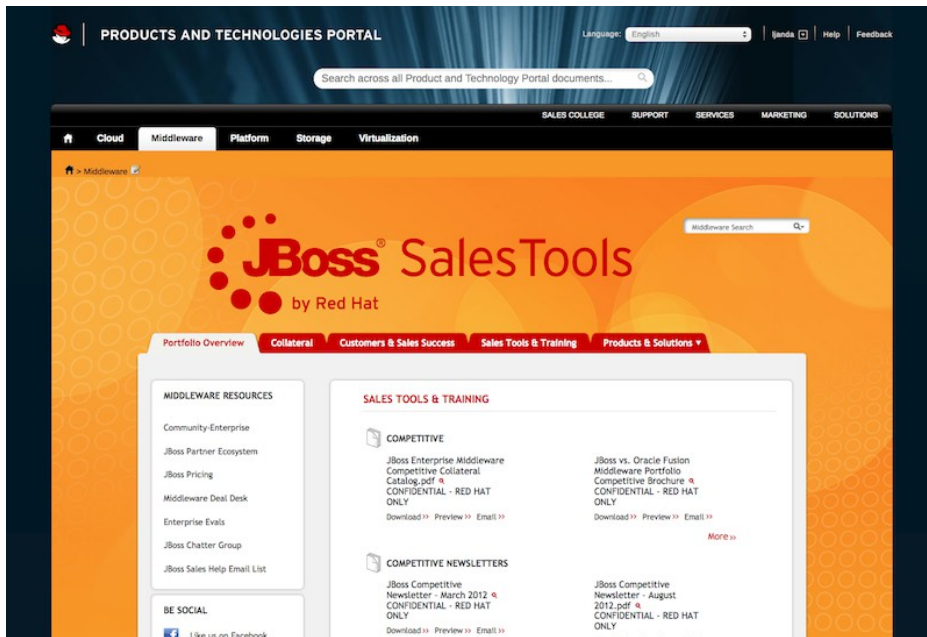
Who I am

Who am I?

- Luděk Janda
- Product Documents Librarian
 - Administrator of DMS (Alfresco)
 - Responsible of content model
 - Keeping the overall structure organized and every document in its place
- PnT Portal, <https://engineering.redhat.com/pnt/>
- Ultimate place where the sales and field teams could get all the documentation related to products

PnT Portal

- ~ 2660 documents
- 13 main sections
- 150 content owners
- Standardized folders structure
- Classification based on document types
- Regular review process (6 months)



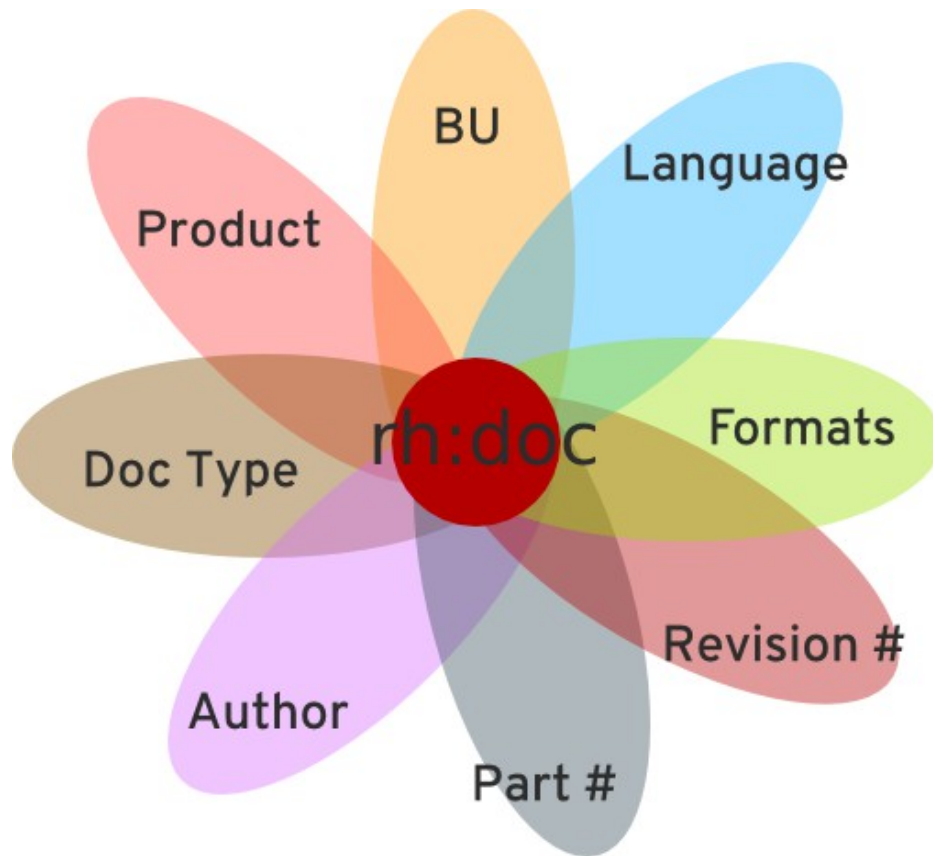
Paul Cormier's requirements

- Create unified look/feel for entry into Products + Technology to augment BU-specific wikis and sites
- Organize materials into an easy to use resource center for the materials most often accessed by sales and field teams
- Consistently format and refine sales tools and materials
- Instill discipline to keep content fresh and methodically updated

PnT Portal Taxonomy Efforts

- **Our existing taxonomy:**
 - File naming guidelines
 - <https://home.corp.redhat.com/wiki/file-naming-conventions>
- **New taxonomy:**
 - Should describe all the RH documents in their complexity (document types, context, stakeholders)
 - Make them trackable across the company
 - Make them identifiable (water marking)
- **Taxonomy Server**
 - Keep the taxonomy consistent across all CMSs.

RH Document

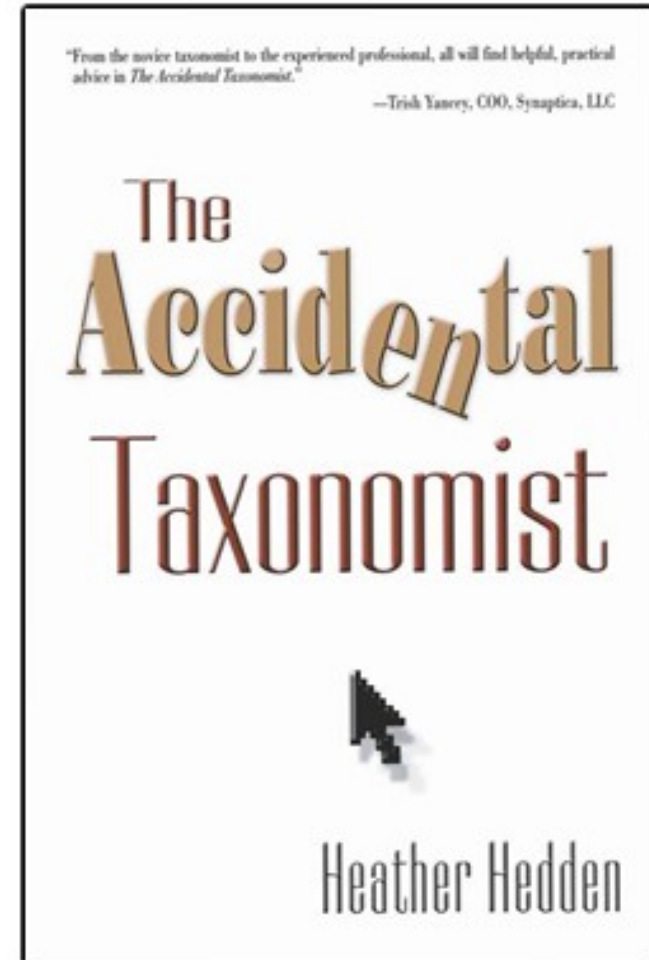


- BU (Cloud, Middleware)
- Product (RHEL 7)
 - pulled down from PP
- Doc Type (Datasheet, Brochure)
 - Leigh Blaylock's team is working on this
- Author (linked to LDAP)
- Part #
- Revision #
- Formats (Print, A4)
- Language (ja-JP)

Introduction to taxonomy

“The Accidental Taxonomist”

- Hedden, Heather: The Accidental Taxonomist. (Information Toady, 2010)
- “A person who have to come to the taxonomist work because a need arose within an organization where they were working in another capacity.”



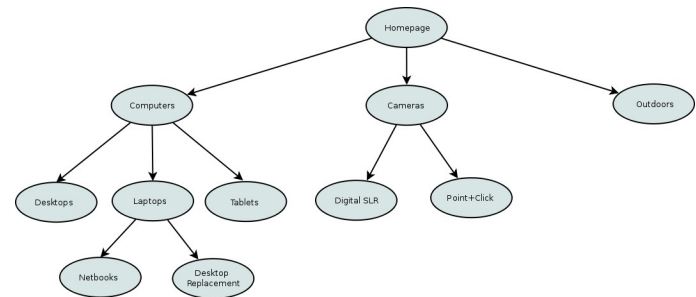
Taxonomy Roles Are Often Combined With Other Job Roles

Source: P. Lambe, Global survey of 184 taxonomy & metadata professionals, 2009



Taxonomy - definition

- Hierarchical classification or categorization system
 - *Narrower sense*
- Any means of organizing concepts of knowledge
 - *Broader sense*



Purposes of taxonomies

- Indexing or cataloging support
 - taxonomy provides a list of agreed-on terms for indexing or cataloging purposes
- Retrieval support
 - users take advantage of structured content
- Organization and navigation support
 - a taxonomy can provide a categorization or classification system for things

Books, Music, DVD

- [Books](#)
- [DVD](#)
- [Magazine Subscriptions](#)
- [Music](#)
- [Video](#)
- [Amazon Shorts](#)

Electronics & Office

- [Electronics](#)
- [Audio & Video](#)
- [Camera & Photo](#)
- [Office Products](#)
- [Software](#)
- [Computer & Video Games](#)
- [Computers](#)
- [Cell Phones & Service](#)

Kids & Baby

- [Toys & Games](#)
- [Baby](#)
- [Imaginarium](#)

Home & Garden

- [Bed & Bath](#)
- [Furniture & Décor](#)
- [Home & Garden](#)
- [Kitchen & Housewares](#)
- [Outdoor Living](#)
- [Pet Supplies](#)
- [Tools & Hardware](#)

Sports & Fitness

- [Exercise & Fitness](#)
- [Sports & Outdoors](#)

Gifts & Registries

- [Baby Registry](#)
- [Free e-Cards](#)
- [Gift Certificates](#)
- [Gift Store](#)
- [Wedding Registry](#)
- [Wish List](#)

E.g. retrieval support, source: amazon.com

Organization Support

The Information Architecture Institute site map, a navigational menu

Member Services	IA Network	Learning IA	About Us
<ul style="list-style-type: none">Join IAIMember BenefitsJob Board<ul style="list-style-type: none">Post a JobView current jobsMentoring Program<ul style="list-style-type: none">What is mentoring?What to expect from a mentor/protégéHow to choose a mentorSample mentoring agreementGrant Program<ul style="list-style-type: none">Progress Grant DetailsGrant RecipientsEvent SponsorshipEmail Lists<ul style="list-style-type: none">IAI NewsletterIAI Members ListJob Board ListIAI Meta ListSecond Life Initiative ListMember Center	<ul style="list-style-type: none">Networking GuideIAI Local Groups<ul style="list-style-type: none">Local Group FAQStart a Local GroupLocal Group Coordinator SurveyGeoLocate IAsFind a NetworkFind a Project<ul style="list-style-type: none">EducationIA GlossaryLocal GroupsIA LibraryIA ToolsTranslation of IASecond LifeJob BoardDiscuss IAContribute Documents	<ul style="list-style-type: none">Research<ul style="list-style-type: none">Salary SurveyIndustry SurveysIA White PapersIAI Grant ResearchEducation<ul style="list-style-type: none">CareersSchools Teaching IAIA CurriculumSelected IA BooksIA GlossaryElevator PitchesTestimonialsLibrary<ul style="list-style-type: none">Introduction to IARecently AddedSearch the LibraryAbout the LibraryBrowse by:<ul style="list-style-type: none">SubjectResource TypeAuthorLanguageAll ResourcesTools	<ul style="list-style-type: none">Our MissionPeople<ul style="list-style-type: none">Board of DirectorsStaffAdvisory BoardVolunteer Initiative LeadersFoundersPartnersGlobal InitiativesAnnual Report <hr/> <ul style="list-style-type: none">Utilities

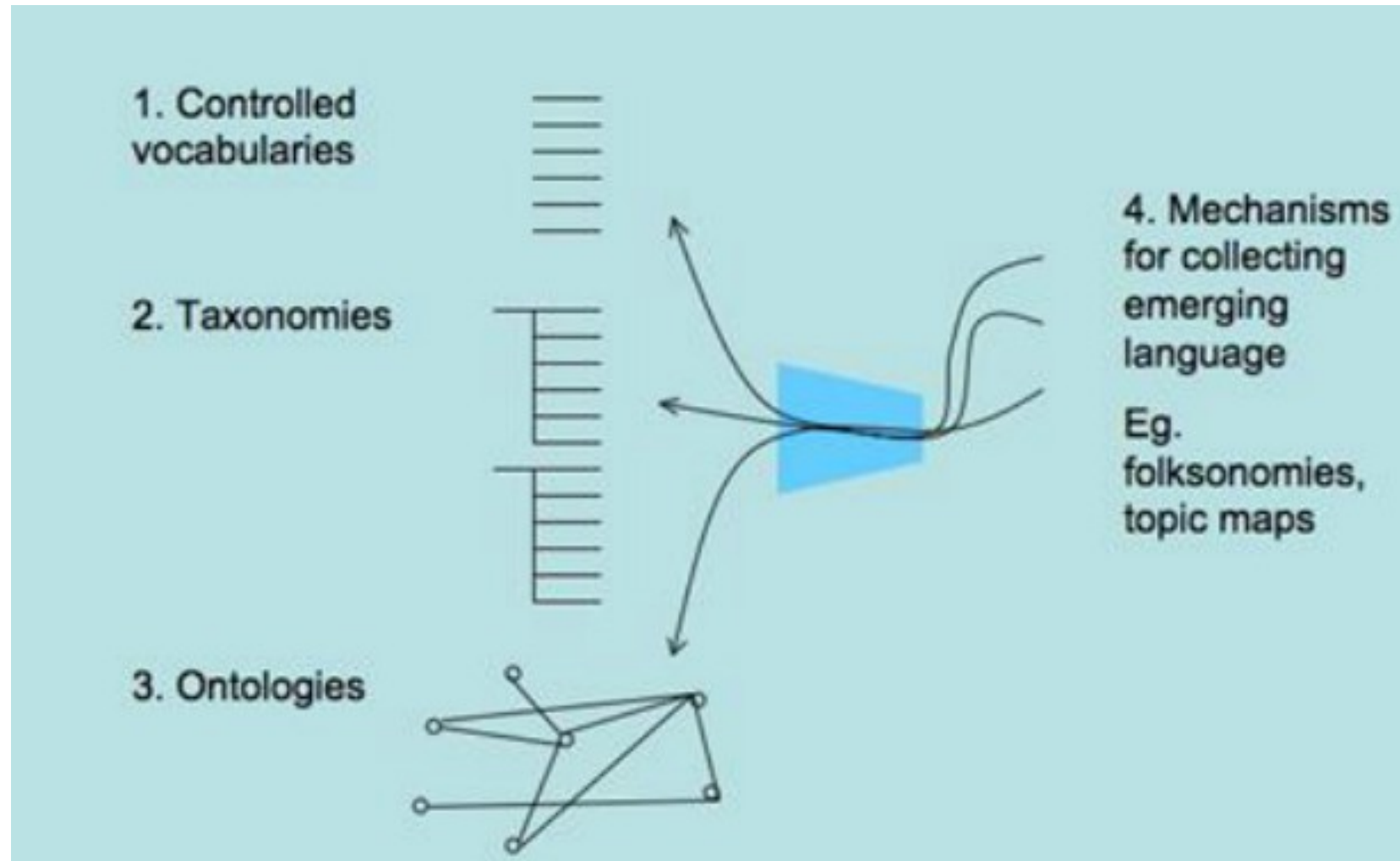
Kinds of taxonomies

- Controlled Vocabulary / Authority File
 - Restricted list of words
- Hierarchical Taxonomy
 - Collection of terms organized into a single large hierarchical structure.
- Faceted Taxonomy
 - Terms are grouped into multiple facets or hierarchies.
- Thesaurus
 - Controlled dictionary with detailed information about relationships among terms.

Kinds of taxonomies

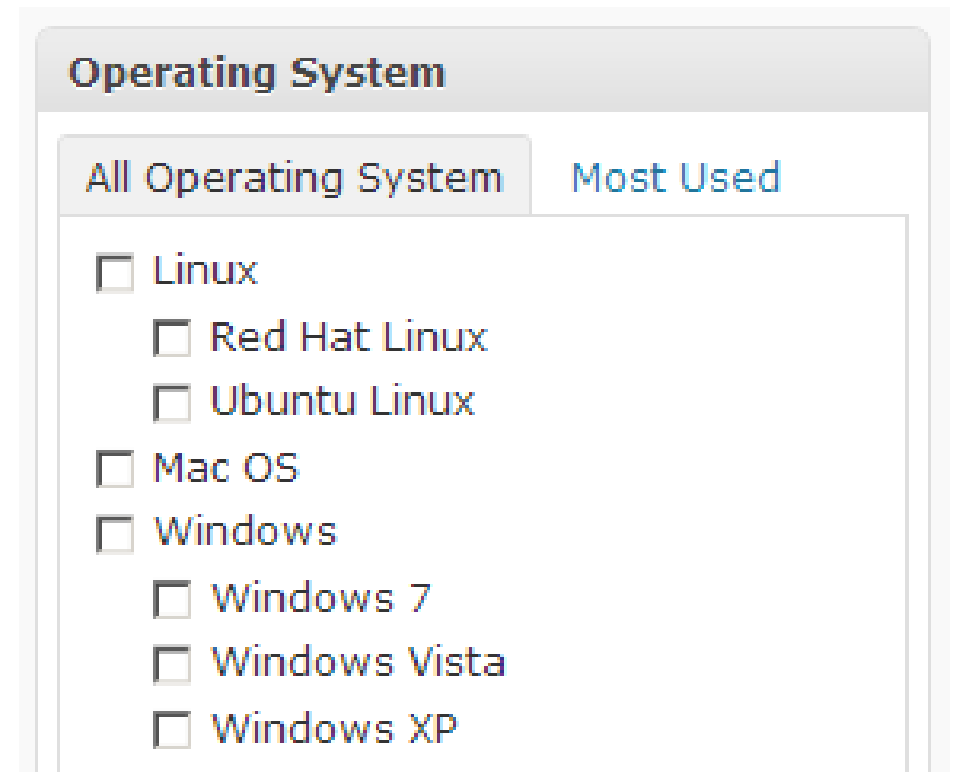
- Ontology
 - Dictionary with complex semantic relationships among terms
- Folksonomy
 - Collection of user-assigned uncontrolled tags.

Orders of Complexity



Hierarchical Taxonomy

- A collection of controlled vocabulary terms organized into a single large hierarchical structure.
- **BT broader term** – relation **child** → **parent**
- **NT narrower term** – **parent** → **child**



Facets

- Groups of terms (facets)
- Describe content from multiple angles, or attributes.
- Each facet represents its own hierarchy and is mutually exclusive.
- The objective is to allow for searching on multiple terms in combination.

The image shows a screenshot of Amazon search facets for 'Point-and-Shoot Digital Cameras'. The facets are organized into several sections:

- Category:** Electronics > Camera & Photo > Digital Cameras. **Point-and-Shoot Digital Cameras**
- Brand:** Any Brand
 - Canon (29)
 - Sony (14)
 - Panasonic (26)
 - Kodak (26)
 - Nikon (8)
 - General Imaging (11)
 - Pentax (8)
 - Olympus (36)
 - Samsung (25)
 - Fuji (15)
 - Casio (4)
 - Bushnell (11)
 - > See more...
- Seller:** Any Seller. **Amazon.com**
- Price:** Any Price
 - \$0-\$24 (32)
 - \$25-\$49 (12)
 - \$50-\$99 (48)
 - \$100-\$199 (126)
 - \$200-\$499 (83)
 - \$500-\$999 (1)
 - \$5000-\$9999 (1)
- Megapixels:** Any Number of Megapixels
 - 1.9 MP & Under (12)
 - 2 to 2.9 MP (2)
 - 3 to 3.9 MP (4)
 - 5 to 5.9 MP (10)
 - 6 MP & Up (106)
- Optical Zoom:** Any Optical Zoom
 - 1.9x & Under (22)
 - 2.0x to 3.9x (124)
 - 4.0x to 5.9x (36)
 - 6x to 9.9x (12)
 - 10x to 19x (15)
- Display Size:** Any Display Size
 - 1.9 in. & Under (20)
 - 2 to 2.9 in. (111)
 - 3 to 4.9 in. (11)
- Image Stabilization:** Any Image Stabilization
 - None (74)
 - Optical (37)
 - Electronic (5)
- Viewfinder Type:** Any Viewfinder Type
 - None (92)
 - Optical (41)
 - LCD (13)

Facets, source: amazon.com

Thesaurus

- Includes information about each term
- its relationships to other terms
- Contains synonyms or alternate expressions for each term entry.
- **BT/NT**
- **RT** – Related term
- **USE/UF** – Use/Used For
- **SN** – Scope Note
- German language **RT** German history
- Inundations **USE** Floods
- Floods **UF** inundations
- **Analyzers SN:** Excludes devices for performing mathematical analysis.

A term in a thesaurus

The Thesaurus of ERIC Descriptors:

Acoustics

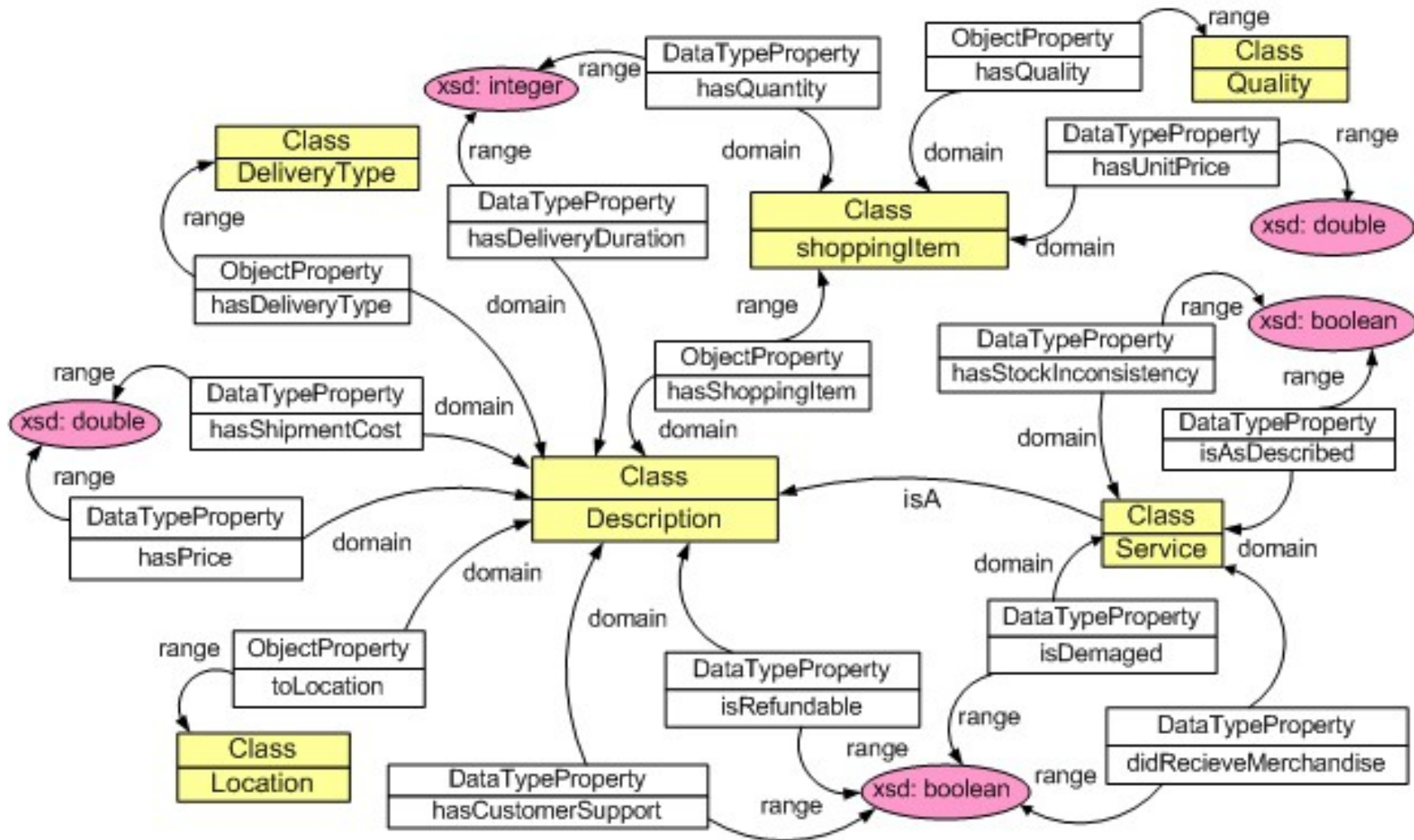
- SN Science of sound—includes the study of the transmission of sound through various media or in various enclosures
- U Sound
 - Sound Transmission
 - Sound Waves
- NT Psychoacoustics
- BT Sciences
- RT Acoustic Insulation
 - Acoustic Phonetics
 - Acoustical Environment
 - Architecture
 - Audio Equipment
 - Auditory Stimuli
 - Noise (Sound)
 - Physics
 - Sound Effects

[SN = Scope Note, U = Used for, NT = Narrower Term, BT = Broader Term, and RT = Related Term]

Ontology

- Complex relationships between terms
- Semantic relationships
 - The relationship is given a name and is unilateral. (E.g. Relationship between two persons - “**Paul is a son of Mike**”).
- The terms are accompanied by specific attributes in a more structured format, such as properties, features, characteristics, or parameters.

Example of a domain ontology dealing with retail management



Development of a taxonomy

Concepts & Terms

- Concepts
 - *Things or ideas that make up a taxonomy*
 - *E.g. **affection***
- Terms
 - *Labels for concepts. Most common, generic designation*
 - *E.g. various expressions for affection: **love, passion...***

Preferred and Nonpreferred Terms

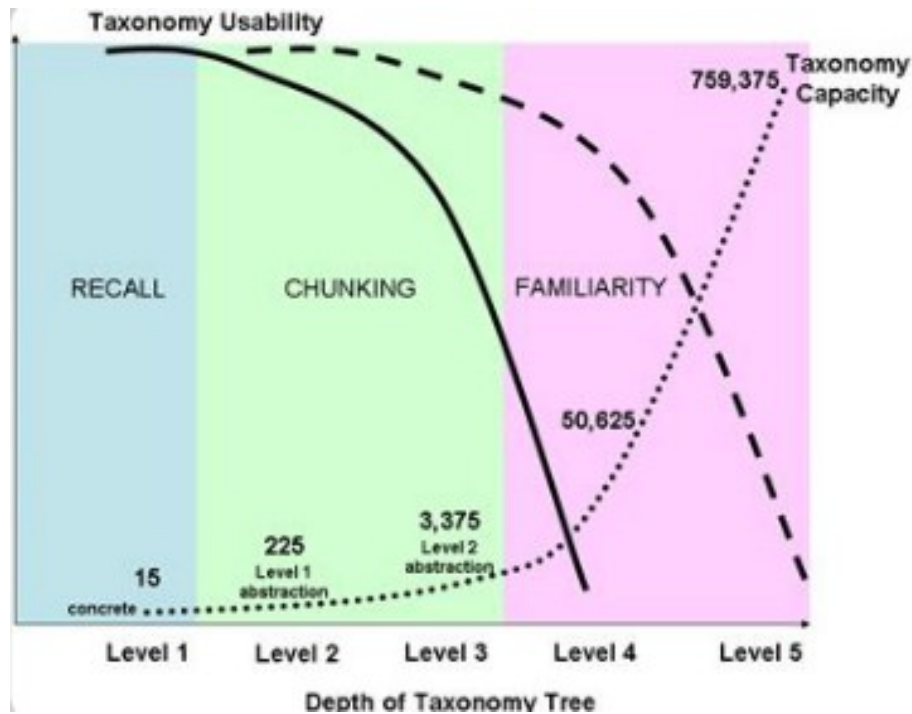
- **Candidate term**
 - Each concept/term we consider
- **Preferred term**
 - Is the official displayed word or phrase for the concept.
- **Nonpreferred term/s**
 - Various synonyms, or other sufficiently equivalent words or phrases.
- Ex. general public: *preferred t.: doctor*
- *Nonpreferred t.: physician*

Rules for Term Selection

- Appropriate
- Belonging
- Consistency
- Currency
- Distinctiveness
- Implication
- Novelty
- Standardization
- Structure
- Technical Accuracy
- Transparency
- Usability
- Warrant

Natural Capacity of a taxonomy tree structure

- Optimal lists have 7+/-2 items. Beyond 12-15 items lists become difficult to understand.



9 Principles of Taxonomy Design

- **1. Intuitive**

- The taxonomy structure reflects natural working or usage habits, or well-known structures (such as organizational structure...).

- **2. Unambiguous**

- Users do not have more than one obvious option for where to place content or find content they need.

- **3. Hospitable**

- The taxonomy can accommodate all new content without the need for significant revision.

9 Principles of Taxonomy Design

- **4. Consistent and Predictable**

- The taxonomy provides sufficient context for users to be able to navigate the structure quickly and accurately. (e.g. alphabetic order consistently applied)

- **5. Relevant**

- The taxonomy reflects user language and user perspective on how content is organized and connected.

- **6. Parsimonious**

- The taxonomy structure offers no more and no less than what is required for the content that is to be accommodated.

9 Principles of Taxonomy Design

- **7. Meaningful**

- The outcomes of searches are accurate in terms of the usefulness of the content retrieved.

- **8. Durable**

- The taxonomy does not need frequent change or expansion and rarely requires radical change or reorganization.

- **9. Balance**

- When the taxonomy is populated with content, there are relatively even quantities of content across the taxonomy.

Unbalanced taxonomy - list of packages

source: `$ rpm -qa --queryformat "%{GROUP}\n" | sort`

1 Applications/Development	5 Text Editors/Integrated Development Environments (IDE)	26 Applications/Publishing
1 applications/system	6 Development/Debuggers	33 System Environment/Kernel
1 Application/System	6 Public Keys	40 Unspecified
1 Development/Documentation	7 Applications/Communications	47 System Environment/Daemons
1 Development/Java	7 Applications/Databases	57 Development/Tools
1 Libraries	7 Development/System	60 User Interface/Desktops
1 Lifelong Kindergarten Group at the MIT Media Lab	8 Documentation	64 User Interface/X
1 System/Libraries	9 Applications/File	69 Applications/Multimedia
1 Utilities	14 Applications/Editors	85 Applications/Internet
2 Amusements/Games	14 Applications/Emulators	90 Applications/System
2 Amusements/Graphics	15 Applications/Archiving	93 Development/Languages
2 System Environment/Shells	20 User Interface/X Hardware Support	158 Development/Debug
2 Utilities/Printing	21 Applications/Productivity	186 System Environment/Base
4 Development/Libraries/Java	25 Applications/Text	323 Development/Libraries
5 Applications/Engineering	26 Applications/Publishing	662 System Environment/Libraries

Development of Taxonomies

- **1. Research**
 - including stakeholders interviews and a content audit
- **2. Strategy**
 - developing the taxonomy plan and governance document
- **3. Design**
- **4. Implementation**
 - actually building the taxonomy
- **5. Administration**
 - maintenance of the taxonomy

Research of Sources

- **Content**

- The content or material that will be indexed or categorized with the taxonomy. It's also called **Content Audit** or **Content Warrant**

- **People**

- including the taxonomy owners, subject matter experts, and sample users. Also called **User Warrant**

- **External Sources**

- Also called **Standards Warrant**
- e.g. ISO 29845:2011 – document types of technical product documentation

Example of how a spreadsheet can be set up for a content audit.

	A	B	C	D	E	F	G	H	I	J	K
1	Content Inventory - ITS, Web section										
2	URL: http://www.its.monash.edu.au/web/										
3											
4	Page ID	Page title	URL	Template	Access	Keywords	Description	Maintainer	Approver	Status	Notes
5	1	Policy and guidelines	/web/policy	Section entry page	Public	policy, guidelines	Official policies and	WRD	Sue Steele	OK	
6	1.1	Webmaster policy	/web/policy	Policy document	Monash staff	webmaster, policy		WRD	Sue Steele	Redundant	This policy
7	1.2	Web server policy	/web/policy	Policy document	Monash staff	web server, policy		WRD	Sue Steele	Trivial	
8	1.3	Chowning policy	/web/policy	Policy document	Monash staff	chowning, policy		WRD	Sue Steele	Review	Shorten, k
9	1.4	Accessibility policy	/web/policy	Policy document	Public	accessibility, disability, policy		WRD	Sue Steele	OK	
10	1.4.1	Accessibility contacts	/web/policy	Standard document	Public	accessibility experts		WRD	Sue Steele	OK	
11	1.4.2	Accessibility resources	/web/policy	Standard document	Public	accessibility, resources, tools, links		WRD	Sue Steele	OK	
12	2	Web workshops	/web/workshops	Section entry page	Public	workshops, training	Presentation notes	WRD	Sue Steele	OK	
13	2.1	Optimising images for the web	/web/workshops	Presentation	Public	images, optimisation		WRD	Sue Steele	OK	
14	2.2	Introduction to usability	/web/workshops	Presentation	Public	usability, user-centred design, UCD		WRD	Sue Steele	OK	
15	2.3	Migrating to the CMS	/web/workshops	Presentation	Monash staff	CMS, Teamsite, migration		WRD	Sue Steele	OK	

Testing & Maintenance

Testing

- Purposes of testing processes:
 - Test a taxonomy design for usability
 - Gain further insight into User Audit
- 4 kinds of test:
 - Open card sorts
 - Closed card sorts
 - Scenario-based tests
 - Balance/load testing

Open-card sorts

- Topic terms written on cards.
- We give these card sets to our test subjects (individuals or groups).
- They sort them into clusters and then give category labels to the clusters.
- Information about natural categorization patterns of our target user communities.



Closed card sorts

- We have a number of topic cards.
- Users are asked to sort them into a pre-defined set of categories.
- Test the predictability of a two level taxonomy that we drafted.



Taxonomy maintenance

- “Taxonomy is never done”
 - Hedden, Heather: *The Accidental Taxonomist*, page 321.
- Terms must be updated and modified in keeping with evolving topics
- Taxonomist's duties: documentation, test-searching, indexing support, terms updates

Q & A