Taxonomy Subcommittee Status

GK	T23@	A23@	PF
ZF	A24@	T23@	PD
O29@	A27@	A26@	P1F
O30@	T24@	A27@	#F
O31@	S24@	G13@	@F
A23@	T26@	E2@	&F
TF	A25@	PD	
A6@	A28@	PD	
G56F	U25@	PD	

COMPUTER HISTORY MUSEUM

18 February 2004

Paul McJones
Mike Blasgen
Dag Spicer
&
Sharon Brunzel
Mary Cicalese
Bernard Peuto

How taxonomy fits in with metadata

- Taxonomy is a kind of *descriptive* metadata
- CHM uses a qualified Dublin Core approach
  - <Subject> element give the *topic* of a thing (e.g., book about AI, photo of Seymour Cray)
  - <Subject.ObjectCategory> element says what a physical object *is*, e.g. a calculator
  - <Subject.SoftwareCategory> element says what a piece of software *is*, e.g., a payroll program

Review: approach

- Select an initial taxonomy
- Consider questions such as:
  - Uses: guide collecting vs. guide search
  - How does taxonomy relate to SCC metadata?
  - Does single taxonomy work across 50+ years?
  - OK to have several taxonomies?
- Outcome: initial taxonomy + recommendations for its evolution

Experiment

- Classified Booch Classic Software list (http://www.computerhistory.org/cgi-bin/wishlist.cgi) using:
  - UN Standard Products and Services Code
    - Commercial orientation (143 categories)
    - http://www.unspsc.org/
  - ACM Computing Classification System
    - Bibliographic orientation
    - http://www.acm.org/class/

Summary of findings

- Both taxonomies are plausible
  - UN SPSC is more appropriate for specific programs/systems (<Subject.SoftwareCategory>)
  - ACM CCS is more appropriate for describing topics (<Subject>)

Statistics: Booch vs. UN SPSC

Booch	#	UN SPSC
operating system	40	Operating system
language	29	Compiler and decompiler
application	15	9 categories: Analytical, Business function specific, CAD, Flight control, GUI dev., …
game	15	Computer game; Action game; Adventure game
text editing	14	Word processing; Content authoring & editing
web	11	8 categories: Email, Info. exchange, Networking, Internet browser, Conferencing, …
scientific	6	Analytical or scientific; Expert system
database	5	Data management and query
AI	4	Expert system
-	2	Analytical or scientific; Compiler
algorithms	2	<none>
utility	2	Security and protection; Transaction security/virus protection
assembler	1	Development
cmd. & control	1	Industry specific
framework	1	Development environment
laboratory	1	Compiler and decompiler
music	1	Music or sound editing
video editing	1	Video creation and editing

Miscellaneous notes (1)

- UN SPSC lacks categories *interpreter*, *assembler*, *linker*
- Both lack category *virus* (but not anti-viruses)
- Best category for "vanilla" text editors: ed, emacs, TECO, vi, uedit is *Content authoring and editing*
- Best category for Collected algorithms of the ACM, Linpack, Macsyma, Matlab, Netlib is *Analytical or scientific*
- Is Elizia (sic) *Expert system* or *Computer game or entertainment*?

Miscellaneous notes (2)

- Hypercard and MacApp are *Development environment software* or *Graphical user interface development software*
- *How about* ThingLab?
- Logic Theorist, SAINT, SHRDLU, Mycin are *Expert system software*
- MacDraw and MacPaint are *Content authoring and editing software* or *Graphics or photo imaging software*

Miscellaneous notes (3)

- Mysteries in Booch's list:
  - Simscript is a *classic editor?*
  - What is uedit?
  - MAD is a *classic scientific application* (not Michigan Algorithm Decoder?)
  - Bravo, PageMaker are *application* rather than *text editing*

Proposal

- Adopt UN SPSC for <Subject.SoftwareCategory>
- Maintain list of exceptions (additions, refinements) as we begin collecting
  - A Wiki would be nice
- Revisit when collection size approaches 1000