Implementing a Government-wide Semantic Solution to Thesauri

Kenneth B. Sall, Science Applications International Corporation (SAIC) and

Ronald P. Reck, RRecktek LLC

April 20, 2006

XML Community of Practice (XML CoP)

Town Hall at the eGov Institute's KM Conference.

Agenda

- Problem

- Goals and Requirements

- Basic Thesaurus Terminology and IC Examples

- SKOS (Simple Knowledge Organisation System)

- Our SKOS Element Subset and Extensions

- SKOSaurus Pilot

- DTIC Thesaurus Examples

- Potential Next Steps

Problem Statement

- Government agencies need common vocabulary of (technical) terminology.

- Communication and data sharing is greatly enhanced when the semantics are clear.

- Various government groups approach this in different ways -- Microsoft Word, Excel, HTML, databases, and wiki pages: bulleted lists, tables, spreadsheets, acronym lists, etc.

- Need to focus on a common formats and standards that enable reuse and harmonization across Communities of Interest (COIs).

Goals and Requirements

- Allow constraining terms to one COI or sharing across COIs.

- Should benefit from ISO standards for thesauri.

- Enable term authors to use familiar tools (e.g., Excel).

- Leverage existing Microsoft formats to access expressive backend data stores.

- XML-based (RDF) solution with few required elements but many optional and/or repeatable elements.

- Multiple definitions of the same term must be permitted, with either same or different subject/context.

- Should support semantic relationships between terms: "search thesaurus".

4

% Thesauri Standards and Specifications

- **ISO 2788:1986** – Documentation - Guidelines for the establishment and development of monolingual thesauri

    - Developing a Thesaurus (mono-lingual)

    - ISO 5964:1986 – multi-lingual version

- ISO 1087:2000 - Vocabulary of Terminology

- ISO 704:2000 - Principles and Methods

- **ANSI/NISO Z39.19-2003** - Construction, Format, and Management

- ISO 15836:2003 - The Dublin Core metadata element set

- [Many more listed in paper.]

5

% Basic Thesaurus Terminology (1) [ISO 2788:1986]

- *Thesaurus* – list of concepts in a particular domain of knowledge together with explicit relationships

- *Concept* - unit of thought that exists in the mind as an abstract entity, independent of the term(s) that identify it (i.e., human language independent)

- *Concept Scheme* - set of concepts, optionally including statements about semantic relationships between those concepts.

  – Thesauri, classification schemes, subject heading lists, taxonomies, terminologies, glossaries and other types of controlled vocabularies

Basic Thesaurus Terminology (2) [ISO 2788:1986]

- ISO 2788 defines abbreviations for each thesaurus construct. These "generally recognized conventions" are useful for compactness and in automated processing.

- USE (or SEE) – *preferred* label for this concept follows

- UF = USE FOR – *alternate* label follows, may be a synonym but less preferred

    – e.g., ELECTRONIC INTELLIGENCE **UF** ELINT

          ↑   ↑
     Preferred label   Alternate label

Basic Thesaurus Terminology (3) [ISO 2788:1986]

- SN = Scope Note - to clarify or constrain the meaning

  – sometimes contains the concept's definition

- BT = Broader Than – another concept more *general* than this concept

- NT = Narrower Than – more *specialized* than this concept

- RT = Related To – concept that is *similar* in some way

Thesaurus Concept Example

MT Adult education
BT Education
BT Employee training
NT Continuing education
RT Community colleges
RT Compensatory education
RT Higher education
SN Educational programs/services, usually at the postsecondary level, for adults seeking personal, academic or occupational enhancement.

Source: GAO Thesaurus, Feb. 2005

Example: DTIC Thesaurus (1)

INTELLIGENCE
 NT ACOUSTIC INTELLIGENCE
 NT COUNTERINTELLIGENCE
 NT *ELECTRONIC INTELLIGENCE
 NT INTELLIGENCE(HUMANS)
 NT *MILITARY INTELLIGENCE
 NT PHOTOGRAPHIC INTELLIGENCE

INTELLIGENCE(MILITARY)
 use MILITARY INTELLIGENCE

DTIC Thesaurus

Italics added

Source:

Defense Technical Information Center

10

SAIC®
An Employee-Owned Company

DTIC Thesaurus (2)

MILITARY INTELLIGENCE
  EVALUATED INFORMATION CONCERNING AN ACTU
  AL OR POSSIBLE ENEMY THEATER OF OPERATIO
  NS.
 UF INTELLIGENCE(MILITARY)
 BT INTELLIGENCE
 NT AIR INTELLIGENCE
 NT ARMY INTELLIGENCE
 NT COMMUNICATIONS INTELLIGENCE
 NT ESPIONAGE
 NT NAVAL INTELLIGENCE
 NT STRATEGIC INTELLIGENCE
 NT *TACTICAL INTELLIGENCE

DTIC Thesaurus (3)

ARMY INTELLIGENCE
 INCLUDES EVERY PHASE AND HANDLING OF INF
 ORMATION FROM ITS EVALUATION COLLATION,
 SYNTHESIS, INTERPRETATION AND PRESENTATI
 ON, TO ITS DISSEMINATION BY THE ARMY.
 BT *MILITARY INTELLIGENCE

COMINT
 use COMMUNICATIONS INTELLIGENCE

COMMUNICATIONS INTELLIGENCE
 TECHNICAL AND INTELLIGENCE INFORMATION D
 ERIVED FROM FOREIGN COMMUNICATIONS BY OT
 HER THAN THE INTENDED RECIPIENTS.
 UF COMINT
 BT *MILITARY INTELLIGENCE

12

DTIC Thesaurus (4)

COUNTERINTELLIGENCE
 BT INTELLIGENCE

ELECTRONIC INTELLIGENCE
  THE TECHNICAL AND INTELLIGENCE INFORMATI
  ON DERIVED FROM FOREIGN NONCOMMUNICATION
  S ELECTROMAGNETIC RADIATIONS EMANATING F
  ROM OTHER THAN NUCLEAR DETONATIONS OR RA
  DIOACTIVE SOURCES.
 UF ELINT
 BT INTELLIGENCE
 NT RADAR INTELLIGENCE

ELINT
 use ELECTRONIC INTELLIGENCE

DTIC Thesaurus (5)

Note: We will see the data from these DTIC slides later in SKOSaurus.

NAVAL INTELLIGENCE
 BT *MILITARY INTELLIGENCE

RADAR INTELLIGENCE
 INTELLIGENCE CONCERNING RADAR OR INTELLI
 GENCE DERIVED FROM THE USE OF RADAR EQUI
 PMENT.
 BT *ELECTRONIC INTELLIGENCE

STRATEGIC INTELLIGENCE
 BT *MILITARY INTELLIGENCE

TACTICAL INTELLIGENCE
 BT *MILITARY INTELLIGENCE
 NT TERRAIN INTELLIGENCE

14

DTIC Thesaurus (6) – Navigation Interface

[DTIC
Thesaurus](#)

Example: CALL Thesaurus (1)

[CALL Thesaurus](#)

energy intelligence
evasion and escape intelligence
finished intelligence
foreign intelligence
geographic intelligence
intelligence cycles
intelligence disciplines
intelligence estimates
interdepartmental intelligence
joint intelligence
medical intelligence
national intelligence
national intelligence estimates
naval intelligence
political intelligence
raw intelligence
scientific and technical intelligence
technical intelligence
terrain intelligence

Related Terms:

intelligence and electronic warfare
intelligence publications
military capabilities studies

Send your comments to: CALL Thesaurus

Last updated: Aug 25, 2005 (cer)

Concept: intelligence

CALL Thesaurus (2)

CALL Thesaurus

Concept: finished intelligence

FEA Business Reference Model (BRM)

BUSINESS REFERENCE MODEL (BRM)
Services for Citizens: Intelligence Operations

Intelligence Operations involves collecting and analyzing information to meet the national security challenges of the US by processing reliable, accurate foreign intelligence, and disseminating intelligence products to policymakers, military commanders, and other consumers.

- **Dissemination**
  Consists of delivering the intelligence products to consumers.

- **Intelligence Analysis & Production**
  Converts large amounts of data to a form suitable for the production of finished intelligence to include translation, decryption, and interpretation of information stored on film and magnetic media through the use of highly refined photographic and electronic processes.

- **Intelligence Collection**
  Involves the gathering of raw data from multiple sources from which finished intelligence is produced.

- **Intelligence Planning and Direction/Needs**
  Establishes the intelligence requirements of the policymakers, the President, the NSC, military commanders, and other officials in major departments and governmental agencies.

Intelligence Operations in the BRM

18

SKOS
(Simple Knowledge Organisation System)

- Leverages ISO 2788 (and ISO 5964) by defining an RDF vocabulary based on the ISO standards [implied].

- Defines an XML element (SKOS property) for each thesaurus construct (USE, UF, BT, NT, SN, etc.) and many more.

- Semantic Web Best Practices and Deployment Working Group – W3C

- SKOS Working Drafts (W3C) and Related Efforts

  - SKOS Core Guide

  - SKOS Core Vocabulary Specification

  - Quick Guide to Publishing a Thesaurus on the Semantic Web

  - Also: SKOS Mapping, Extensions, API, Development Wiki

19

SKOS Core

- **SKOS Core** - model for expressing structure and content of *concept schemes*:

  – Thesauri

  – Classification schemes

  – Subject heading lists

  – Taxonomies

  – "Folksonomies"

  – Other types of controlled vocabulary

  – Concept schemes are also embedded in glossaries and terminologies.

  Source: SKOS Core Guide, November 2005

SKOS Core Vocabulary

- ## SKOS Core Vocabulary

  - Application of Resource Description Framework (RDF)

  - Can be used to express a concept scheme as an RDF graph

  - Can be linked to and/or merged with other RDF data by semantic web applications

  - Uses RDFS Classes and RDF Properties to describe Concepts and Concept Schemes

  - 26 Properties and 5 Classes

Source: SKOS Core Guide, November 2005

% SKOS Vocabulary: 5 Classes

CollectableProperty
Collection
Concept   *
ConceptScheme   (*)
OrderedCollection

* = the classes implemented in SKOSaurus pilot
Source: SKOS Core Vocabulary, November 2005

% SKOS Vocabulary: 26 Properties

altLabel   *
altSymbol
broader   *
changeNote
definition   *
editorialNote
example   *
hasTopConcept
hiddenLabel
historyNote
inScheme
isPrimarySubjectOf

isSubjectOf

member
memberList
narrower   *
note
prefLabel   *
prefSymbol
primarySubject
related   *
scopeNote   *
semanticRelation
subject   *
subjectIndicator
symbol

* = 9 properties implemented in SKOSaurus pilot

Source: SKOS Core Vocabulary, November 2005

23

Our SKOS Element Subset and Extensions (1)

- **`skos:Concept`** – contains all statements about properties for a given concept

- **`skos:prefLabel`** – USE; preferred handle for this concept; designator. [In SKOS, no two concepts in the same *concept scheme* may have same prefLabel.]

- **`skos:altLabel`** – UF; alternate handle; spelling variants; *can* be used for abbreviations or acronyms (but we don't)

- **`skos:related`**, **`skos:narrower`**, **`skos:broader`** – associated with, more specific, or more general than this concept

Our SKOS Element Subset and Extensions (2)

- **`skos:scopeNote`** – constrains meaning; ISO 2788 allows definitions to appear here (but we don't)

- **`skos:definition`** – statement or formal explanation of the meaning of a concept

- **`skos:example`** – used in a sentence

- **`skos:subject`** – topic; can be a `skos:broader`

Our SKOS Element Subset and Extensions (3)

- Pilot Extensions (non-SKOS)

  - **ABBREVIATON_OR_ACRONYM** – very common government need (could define as `rdfs:subPropertyOf skos:altLabel`)

  - **SOURCE** - official document names and URLs are preferred, but specific names of people or agencies are acceptable; (probably could define as `rdfs:subPropertyOf skos:note`)

  - **COI** – essentially a `skos:Collection` (with a potential `skos:ConceptScheme`)

SKOS Fragment: Military Intelligence

```
<skos:Concept rdf:about="http://ex.com/MILITARY_INTELLIGENCE#concept">
 <skos:prefLabel xml:lang="en">MILITARY INTELLIGENCE</skos:prefLabel>
 <skos:altLabel xml:lang="en">INTELLIGENCE(MILITARY)</skos:altLabel>
 <skos:definition xml:lang="en">EVALUATED INFORMATION CONCERNING AN
ACTUAL OR POSSIBLE ENEMY THEATER OF OPERATIONS</skos:definition>
 <skos:scopenote xml:lang="en">IC</skos:scopenote>
 <skos:source xml:lang="en">DTIC Thesaurus</skos:source>
 <skos:subject rdf:resource="http://ex.com/intelligence#concept"/>
 <skos:narrower rdf:resource="http://ex.com/AIR_INTELLIGENCE#concept"/>
 <skos:narrower rdf:resource="http://ex.com/ARMY_INTELLIGENCE#concept"/>
 <skos:narrower rdf:resource="http://ex.com/ESPIONAGE#concept"/>
 <skos:narrower rdf:resource="http://ex.com/NAVAL_INTELLIGENCE#concept"/>
 <!-- etc (several NTs omitted for space reasons) -->
</skos:Concept>
```

SKOSaurus Pilot

- Proof of concept

- Many simplifying assumptions

- Fabricated data (except for DTIC)

- About 100 man hours

- Ron Reck and Ken Sall

- Presented at XML 2005 Conference (Nov. 2005)

% SKOSaurus Pilot: Environment

- CGI script issues SOAP requests and uses RMI.

- The host operating system is Microsoft Windows XP with Service Pack 2.

- Dell Latitude D800 (1.69GHz) with 1G of RAM.

- The Windows XP host runs VMware 5.0 build 13124 to emulate a machine onto which the Solaris X86 operating system version 10 is installed.

- This is referred to as the guest operating system which runs the SKOSaurus system, consisting of:

  – Perl version 5.8.7 and various Perl modules

  – Java version 1.4.2.08

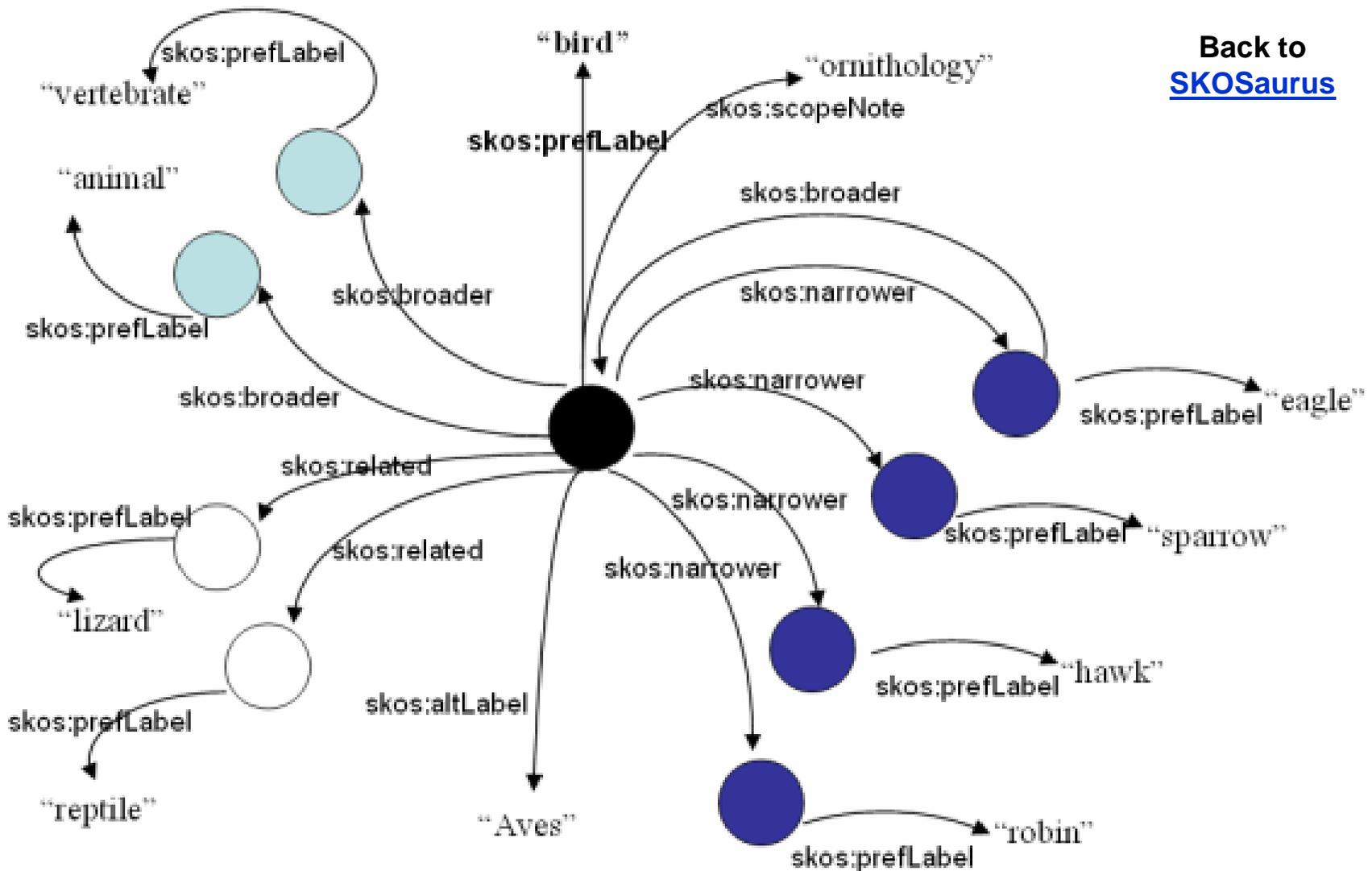 – Kowari server 1.1.0 Pre2

  – XSLT stylesheets

29

Main Use Cases (for Pilot)

- Concept Entry via Web Form

- File Upload of Excel Spreadsheet (as CSV)

- File Upload of SKOS (or RDF)

- Query of Concept Data Store

RDF Graph: Bird Example

31

% Illustrative Statements from RDF Graph

- An alternate label (`skos:altLabel`) for "bird" is "Aves".

- The concepts with the preferred label "vertebrate" and "animal" are broader than the concept with the preferred label "bird".

- There are four specializations of birds listed ("robin", "hawk", "sparrow" and "eagle"), each indicated as `skos:narrower` than "bird".

- The concepts "lizard" and "reptile" are `skos:related` to the "bird" concept in some way.

- Among various concepts which might have the `skos:prefLabel` of "bird", the one illustrated is constrained to ornithology, according to `skos:scopeNote`. This distinguishes the concept from "bird", such as in the informal term for a (young) woman.

% OWL Statements About SKOS

- **`skos:broader owl:inverseOf skos:narrower`**

- **`skos:narrower owl:inverseOf skos:broader`**

and

- **`skos:broader` is an `owl:TransitiveProperty`**

- **`skos:narrower` is an `owl:TransitiveProperty`**

- [RDF/OWL version of SKOS Core](#)

Example Spreadsheet: birds

prefLabel	altLabel	ABBRE	definition	subject	scopeNote
bird	Aves		warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings	ornithology	ornithology
bird	chick		informal term for a (young) woman	slang	slang; more likely to be used in the UK
shuttlecock	birdie; cock; shuttle; bird		A shuttlecock is a high-drag projectile used in the sport of badminton. It has an open conical shape| with a rounded head at the apex of the cone traditionally made of cork and a skirt traditionally of sixteen overlapping goose feathers.	sports	This term is applies only to the sport of badminton.
birdie	birdy		A score on an individual hole that is one stroke below par.	golf	This term applies only to golf.
eagle			any of various large keen-sighted diurnal birds of prey noted for their broad wings and strong soaring flight	ornithology	This terms refers to ornithology.
eagle			a score of two strokes under par on a hole	golf	This term applies only to golf.
Eagles			The Eagles are an American rock music group that originally came together in Los Angeles| California in the early 1970s.	music	This applies only to the rock music group.
Eagles	Philadelphia Eagles		Philadelphia football team	sports	This is a football team.
robin	American Robin		The American Robin (Turdus migratorius) is a	ornithology	This terms refers to

prefLabel	SOURCE	example	narrower	broader	related
bird	WordNet [http://wordnet.princeton.edu/] and http://en.wikipedia.org/wiki/Bird	A bird is an animal with feathers. All birds have wings but not all birds can fly. Ostriches| emus and penguins are birds that can't fly. Parrots| pelicans and wedge-tailed eagles are birds that can fly. All birds hatch from eggs. All birds are warm blooded.	eagle; sparrow; hawk; robin; kiwi	vertebrate; chordate; animal	birding; birdwatching; reptile; lizard
bird	http://wordnet.princeton.edu/	The young man whistled as the blonde bird walked by.		female	dame; doll; wench; skirt
shuttlecock	http://en.wikipedia.org/wiki/Birdie			badminton	
birdie	http://golf.about.com/cs/golfterms/g/bldef_birdie.htm				par; bogey; even; eagle
eagle	WordNet [http://wordnet.princeton.edu/] and http://en.wikipedia.org/wiki/Eagle		bald eagle	bird	raptor
eagle	WordNet [http://wordnet.princeton.edu/]				
Eagles	http://en.wikipedia.org/wiki/Eagles	The Eagles compilation| Their Greatest Hits (1971-1975)| with US sales of 28|000|000 is the best selling album of all time.	Glenn Frey; Don Henley; Joe Walsh; Timothy Schmit	California Rock	Linda Ronstadt; Poco; James Gang
Eagles	http://www.philadelphiaeagles.com/default.jsp	The Philadelphia Eagles are part of the NFL.	Donovan McNabb	NFL	Washington Redskins
robin	http://en.wikipedia.org/wiki/American_Robin	The American Robin has gray upperparts and head| and orange underparts| usually		thrush; bird	blackbird; European Robin

34

esentat

Single Row of Spreadsheet

prefLabel	altLabel	ABBRE	definition	subject	scopeNote
bird	Aves		warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as wings	ornithology	ornithology

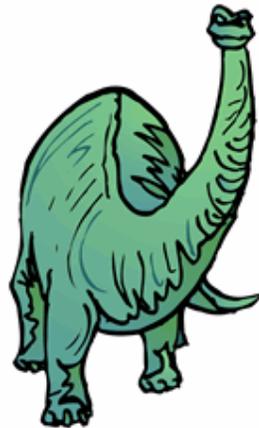| SOURCE	example	narrower	broader	related
WordNet [http://wordnet.princeton.edu/] and http://en.wikipedia.org/wiki/Bird	A bird is an animal with feathers. All birds have wings but not all birds can fly. Ostriches| emus and penguins are birds that can't fly. Parrots| pelicans and wedge-tailed eagles are birds that can fly. All birds hatch from eggs. All birds are warm blooded.	eagle; sparrow; hawk; robin; kiwi	vertebrate; chordate; animal	birding; birdwatching; reptile; lizard

% Spreadsheet Conventions (Pilot)

- One row per concept, sparse or densely populated.

- New row for different definition or homonym (e.g., bird). [SKOS conflict: no duplicate prefLabels.]

- The heading row should not be removed or modified.

- Column order is invariant.

- Since several elements are repeatable, use semi-colon to indicate iteration. [Configurable.]

- A limitation in our pilot parser requires the author to use the pipe symbol ("|") instead of a comma within a cell. [Config.]

- Any number of rows can be included, but there must be no blank rows or separator rows.

- File > Save As Comma Separated Values (*.csv).

SKOSaurus: Home Page

XML CoP SKOSaurus Presentation: April 20, 2006

SKOSaurus: Manage COIs

SKOSaurus: Upload CSV or SKOS

SKOSaurus: Upload Feedback

Birds.csv

/usr/tmp/CGItemp61741

MIME Type: *application/vnd.ms-excel*

Upload received. See the upload or the resulting SKOS or additional bootstrapped URIs

Successfully loaded 128 statements from file:/var/apache2/htdocs/incoming/csv-1130627035.txt into rmi://skosaurus.rrecktek.com/server1#drm

Successfully loaded 138 statements from file:/var/apache2/htdocs/incoming/csv-1130627035.txt.bootstrap into rmi://skosaurus.rrecktek.com/server1#drm

Generated SKOS files   Datastore for COI
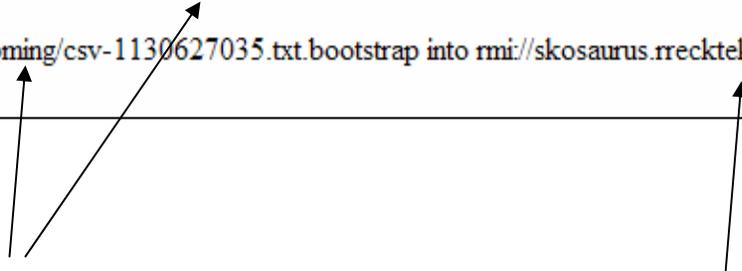
SKOSaurus: Generated SKOS Excerpt (1)

```xml
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  - <skos:Concept rdf:about="http://skos.rrecktek.com/drm/bird#concept">
      <skos:prefLabel xml:lang="en">bird</skos:prefLabel>
      <skos:altLabel xml:lang="en">Aves</skos:altLabel>
      <skos:definition xml:lang="en">warm-blooded egg-laying vertebrates characterized by feathers and forelimbs modified as
        wings</skos:definition>
      <skos:scopenote xml:lang="en">ornithology</skos:scopenote>
      <skos:source xml:lang="en">WordNet [http://wordnet.princeton.edu/] and http://en.wikipedia.org/wiki/Bird</skos:source>
      <skos:subject rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/ornithology#concept" />
      <skos:example xml:lang="en">A bird is an animal with feathers. All birds have wings but not all birds can fly. Ostriches, emus a
        birds that can't fly. Parrots, pelicans and wedge-tailed eagles are birds that can fly. All birds hatch from eggs. All birds are
        blooded.</skos:example>
      <skos:broader rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/vertebrate#concept" />
      <skos:broader rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/chordate#concept" />
      <skos:broader rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/animal#concept" />
      <skos:narrower rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/eagle#concept" />
      <skos:narrower rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/sparrow#concept" />
      <skos:narrower rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/hawk#concept" />
      <skos:narrower rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/robin#concept" />
      <skos:narrower rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/kiwi#concept" />
      <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/birding#concept" />
      <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/birdwatching#concept" />
      <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/reptile#concept" />
      <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/lizard#concept" />
    </skos:Concept>
```

41

SKOSaurus: Generated SKOS Excerpt (2)

```xml
- <skos:Concept rdf:about="http://skos.rrecktek.com/drm/bird#concept">
    <skos:prefLabel xml:lang="en">bird</skos:prefLabel>
    <skos:altLabel xml:lang="en">chick</skos:altLabel>
    <skos:definition xml:lang="en">informal term for a (young) woman</skos:definition>
    <skos:scopenote xml:lang="en">slang; more likely to be used in the UK</skos:scopenote>
    <skos:scopenote xml:lang="en">slang; more likely to be used in the UK</skos:scopenote>
    <skos:source xml:lang="en">http://wordnet.princeton.edu/</skos:source>
    <skos:subject rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/slang#concept" />
    <skos:example xml:lang="en">The young man whistled as the blonde bird walked by.</skos:example>
    <skos:broader rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/female#concept" />
    <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/dame#concept" />
    <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/doll#concept" />
    <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/wench#concept" />
    <skos:related rdf:resource="http://skosaurus.rrecktek.com/bootstrap/drm/skirt#concept" />
  </skos:Concept>
- <skos:Concept rdf:about="http://skos.rrecktek.com/drm/shuttlecock#concept">
    <skos:prefLabel xml:lang="en">shuttlecock</skos:prefLabel>
    <skos:altLabel xml:lang="en">birdie</skos:altLabel>
    <skos:altLabel xml:lang="en">cock</skos:altLabel>
    <skos:altLabel xml:lang="en">shuttle</skos:altLabel>
```

42

SKOSaurus: Web Form

43

SKOSaurus: Kowari Model Dump: Query

XML CoP SKOSaurus Presentation: April 20, 2006

SKOSaurus: Kowari Model Dump: Result

bject $predicate $object from <rmi://skosaurus.rrecktek.com/server1#drm> where $subject $predicate $object;		
	predicate	object
drm/bird#concept	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	http://www.w3.org/2004/02/skos/core#Concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#prefLabel	"bird"
drm/bird#concept	http://www.w3.org/2004/02/skos/core#altLabel	"Aves"
drm/bird#concept	http://www.w3.org/2004/02/skos/core#altLabel	"chick"
drm/bird#concept	http://www.w3.org/2004/02/skos/core#definition	"warm-blooded egg-laying vertebrates characterized by feathers an
drm/bird#concept	http://www.w3.org/2004/02/skos/core#definition	"informal term for a (young) woman"
drm/bird#concept	http://www.w3.org/2004/02/skos/core#subject	http://skosaurus.rrecktek.com/bootstrap/drm/ornithology#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#subject	http://skosaurus.rrecktek.com/bootstrap/drm/slang#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#example	"A bird is an animal with feathers. All birds have wings but not all b
drm/bird#concept	http://www.w3.org/2004/02/skos/core#example	"The young man whistled as the blonde bird walked by."
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/birding#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/birdwatching#conce
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/reptile#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/lizard#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/dame#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/doll#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/wench#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#related	http://skosaurus.rrecktek.com/bootstrap/drm/skirt#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#broader	http://skosaurus.rrecktek.com/bootstrap/drm/vertebrate#concept
drm/bird#concept	http://www.w3.org/2004/02/skos/core#broader	http://skosaurus.rrecktek.com/bootstrap/drm/chordate#concept

SKOSaurus: Intuitive Search

XML CoP SKOSaurus Presentation: April 20, 2006

SKOSaurus: DTIC Data:
17K concepts (in 73K lines) → 70K SKOS statements

```
ABLATIVE MATERIALS
 BT *ABLATION
 BT  MATERIALS

ABNORMAL PSYCHOLOGY
   INCLUDES GENERAL INVESTIGATIONS OF IRREG
   ULAR MENTAL PHENOMENA, INCLUDING BEHAVIO
   R OR MENTAL DISORDERS, DREAMS, HALLUCINA
   TIONS, AND MENTAL RETARDATION. FOR PREVE
   NTION, DIAGNOSIS, AND THERAPY OF MENTAL
   DISTURBANCES, SEE PSYCHIATRY.
 UF   PSYCHOPATHOLOGY
 BT *PSYCHOLOGY
 NT *MENTAL DISORDERS

ABNORMALITIES
 NT   CONGENITAL ABNORMALITIES

ABO INCOMPATIBILITY
 BT *BLOOD GROUP DISORDERS

ABORT

ABORTION

ABRASION
 BT  WEAR

ABRASION RESISTANCE
 use WEAR RESISTANCE

ABRASIVE BLASTING
 UFC  SANDBLASTING
 BT *SURFACE FINISHING
```

```
ANATOMICAL MODELS
 BT  MODELS

ANATOMY
 BT *BIOLOGY
 NT  ABDOMEN
 NT *BODY FLUIDS
 NT *CARDIOVASCULAR SYSTEM
 NT  CEREBROVASCULAR SYSTEM
 NT *DIGESTIVE SYSTEM
 NT *EXTREMITIES
 NT  FISTULAS
 NT *HEAD(ANATOMY)
 NT *HIP
 NT  HUMAN BODY
 NT *INTEGUMENTARY SYSTEM
 NT *MUSCULOSKELETAL SYSTEM
 NT  NECK(ANATOMY)
 NT *NERVOUS SYSTEM
 NT *ORGANS(ANATOMY)
 NT *REPRODUCTIVE SYSTEM
 NT *RESPIRATORY SYSTEM
 NT  SHEATHS(ANATOMY)
 NT  SHOULDERS(ANATOMY)
 NT  THORAX
 NT *URINARY SYSTEM

ANCHOR CHAINS
 BT  CHAINS

ANCHORS
 NT  ANCHORS(MARINE)
 NT  ANCHORS(STRUCTURAL)
 NT  GRAPNELS

ANCHORS(MARINE)
 UF  SHIP ANCHORS
```

[DTIC Thesaurus](#)

Source: Defense Technical Information Center

Intuitive Search 1: Organizations (prefLabel)

SKOSAURUS

Results for Term :ORGANIZATIONS:	
type	Concept
prefLabel	ORGANIZATIONS
prefLabel	
altLabel	ASSOCIATIONS
narrower	MILITARY ORGANIZATIONS
narrower	CORPORATIONS
narrower	LABOR UNIONS
narrower	SCIENTIFIC ORGANIZATIONS
narrower	SOCIETIES
narrower	TASK FORCES

Organizations: Military Organizations (NT)

SKOSAURUS

Results for Term :'MILITARY_ORGANIZATIONS':	
type	Concept
prefLabel	MILITARY_ORGANIZATIONS
prefLabel	
altLabel	ORGANIZATIONS(MILITARY)
broader	ORGANIZATIONS
narrower	MILITARY_FORCES(UNITED_STATES)
narrower	ARTILLERY_UNITS
narrower	BATTLE_GROUP_LEVEL_ORGANIZATIONS
narrower	CORPS_LEVEL_ORGANIZATIONS
narrower	MILITARY_FORCES(FOREIGN)
narrower	MILITARY_RESERVES
narrower	NATO
narrower	SQUADRONS
narrower	WING_LEVEL_ORGANIZATIONS

Organizations: Mil Orgs: Military Reserves (NT)

SKOSAURUS

Results for Term :'MILITARY_RESERVES':	
type	Concept
prefLabel	MILITARY_RESERVES
prefLabel	
broader	MILITARY ORGANIZATIONS
narrower	NATIONAL GUARD
narrower	RESERVE OFFICER TRAINING CORPS

SKOS "definition" Property

SKOSAURUS

Results for Term :'ELECTRONIC_INTELLIGENCE':	
type	Concept
prefLabel	ELECTRONIC_INTELLIGENCE
altLabel	ELINT
definition	THE TECHNICAL AND INTELLIGENCE INFORMATION DERIVED FROM FOREIGN NONCOMMUNICATIONS ELECTROMAGNETIC RADIATIONS EMANATING FROM OTHER THAN NUCLEAR DETONATIONS OR RADIOACTIVE SOURCES.
broader	INTELLIGENCE
narrower	RADAR_INTELLIGENCE

SKOSAURUS

Results for Term :'COMMUNICATIONS_INTELLIGENCE':	
type	Concept
prefLabel	COMMUNICATIONS_INTELLIGENCE
altLabel	COMINT
definition	TECHNICAL AND INTELLIGENCE INFORMATION DERIVED FROM FOREIGN COMMUNICATIONS BY OTHER THAN THE INTENDED RECIPIENTS.
broader	MILITARY_INTELLIGENCE

SKOSAURUS

Results for Term :'MILITARY_INTELLIGENCE':	
type	Concept
prefLabel	MILITARY_INTELLIGENCE
altLabel	INTELLIGENCE(MILITARY)
definition	EVALUATED INFORMATION CONCERNING AN ACTUAL OR POSSIBLE ENEMY THEATER OF OPERATIONS.
broader	INTELLIGENCE
narrower	AIR_INTELLIGENCE
narrower	ARMY_INTELLIGENCE
narrower	COMMUNICATIONS_INTELLIGENCE
narrower	ESPIONAGE
narrower	NAVAL_INTELLIGENCE
narrower	STRATEGIC_INTELLIGENCE
narrower	TACTICAL_INTELLIGENCE

Note: Definitions not shown in other screenshots.

ation: April 20, 2006

SAIC
An Employee-Owned Company

Intuitive Search 2: Intelligence (prefLabel)

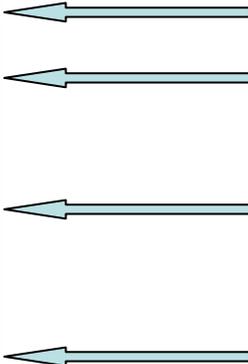| type	Concept
prefLabel	INTELLIGENCE
prefLabel	
narrower	ACOUSTIC INTELLIGENCE
narrower	MILITARY INTELLIGENCE
narrower	COUNTERINTELLIGENCE
narrower	ELECTRONIC INTELLIGENCE
narrower	INTELLIGENCE(HUMANS)
narrower	PHOTOGRAPHIC INTELLIGENCE

52

Intelligence: Military Intelligence

SKOSAURUS

Results for Term :'MILITARY_INTELLIGENCE':	
type	Concept
prefLabel	MILITARY_INTELLIGENCE
prefLabel	
altLabel	INTELLIGENCE(MILITARY)
broader	INTELLIGENCE
narrower	AIR_INTELLIGENCE
narrower	ARMY_INTELLIGENCE
narrower	COMMUNICATIONS_INTELLIGENCE
narrower	ESPIONAGE
narrower	NAVAL_INTELLIGENCE
narrower	STRATEGIC_INTELLIGENCE
narrower	TACTICAL_INTELLIGENCE

XML CoP SKOSaurus Presentation: April 20, 2006

Intelligence: Military Intelligence: narrower

SKOSAURUS

Results for Term :'NAVAL_INTELLIGENCE':	
type	Concept
prefLabel	NAVAL_INTELLIGENCE
prefLabel	
broader	MILITARY_INTELLIGENCE

SKOSAURUS

Results for Term :'ARMY_INTELLIGENCE':	
type	Concept
prefLabel	ARMY_INTELLIGENCE
prefLabel	
broader	MILITARY_INTELLIGENCE

SKOSAURUS

Results for Term :'COMMUNICATIONS_INTELLIGENCE':	
type	Concept
prefLabel	COMMUNICATIONS_INTELLIGENCE
prefLabel	
altLabel	COMINT
broader	MILITARY_INTELLIGENCE

SKOSAURUS

Results for Term :'TACTICAL_INTELLIGENCE':	
type	Concept
prefLabel	TACTICAL_INTELLIGENCE
prefLabel	
broader	MILITARY_INTELLIGENCE
narrower	TERRAIN_INTELLIGENCE

54

(Intelligence: MILINTEL) or (Unconventional Warfare) : Subversion: Espionage

SKOSAURUS

Results for Term :'ESPIONAGE':	
type	Concept
prefLabel	ESPIONAGE
prefLabel	
broader	MILITARY INTELLIGENCE
broader	SUBVERSION

SKOSAURUS

Results for Term :'SUBVERSION':	
type	Concept
prefLabel	SUBVERSION
prefLabel	
broader	UNCONVENTIONAL WARFARE
narrower	ESPIONAGE

SKOSAURUS

Results for Term :'UNCONVENTIONAL_WARFARE':	
type	Concept
prefLabel	UNCONVENTIONAL_WARFARE
prefLabel	
broader	WARFARE
narrower	GUERRILLA WARFARE
narrower	TERRORISM
narrower	COUNTERINSURGENCY
narrower	COUNTERTERRORISM
narrower	SUBVERSION
narrower	EVASION
narrower	SABOTAGE
narrower	SECTARIAN VIOLENCE
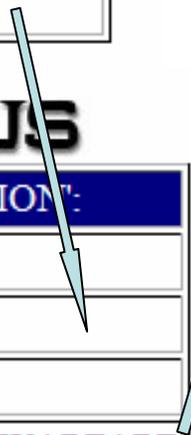
Unconventional Warfare: Subversion: Terrorism

SKOSAURUS

Results for Term :'TERRORISM':	
type	Concept
prefLabel	TERRORISM
prefLabel	
broader	UNCONVENTIONAL WARFARE
narrower	ASSASSINATION
narrower	DOMESTIC TERRORISM

So What?

Now, Connect the Dots!

Intelligence

Unconventional Warfare

Acoustic
Intel

Counterintel

Military Intelligence

Subversion

Terrorism

Sabotage

Electronic
Intel

Photographic
Intel

Air
Intel

Army
Intel

Espionage

Domestic
Terrorism

Communications
Intel

Naval Intel

Strategic
Intel

Tactical Intel

Intelligence(Human)

Terrain Intel

57

Learning
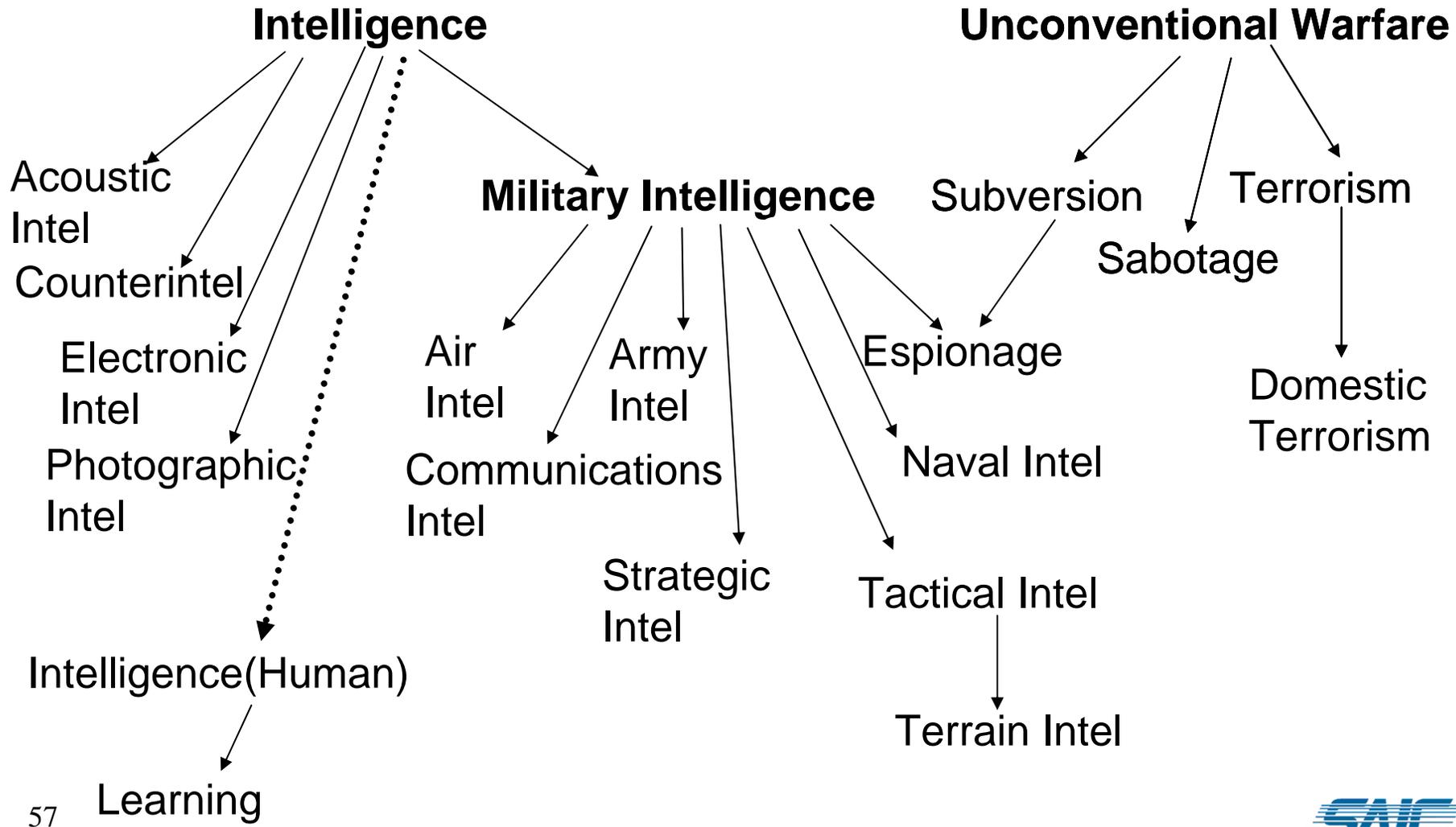
Graphical Interface: Arlington Library – AcornWeb.org

How Does This Help Solve IC Problems?

- Allows concept descriptions in human-friendly Microsoft Office formats.

- Converts relationships to XML-based format; can manipulate with common XML tools.

- XML is really RDF and SKOS, which are *machine-friendly* formats. RDF was designed to be manipulated by machines.

- Semantic Web – moving from humans to machines; *we want computers to do the work for us.*

- SKOSaurus concepts are ripe for integration with commercial semantic tools: METS, Content Analyst, Siderean, Factiva, Images, etc.

Potential Next Steps

- Class of problems with CONOPS (next slide)

- Graphical interface (topicmap-like?)

- Add on Reporting Tools (paper and graphics) – analytic searching, display portions, etc.

- Search across COIs

- Access control mechanisms

- Edit existing concepts

- Ingesting other common formats

- Integration with commercial semantic products

Potential Next Steps: Class of Problems

- Glossary management

- De-confliction (detection and resolution)

- Data Reference Model (DRM) artifact problem

  - Conceptual Taxonomies

  - Conceptual Model to Logical Model

- CONOPS to be developed for specific applications of these concepts

Summary and Conclusions

- Semantic concepts help reveal less than obvious associations.

- SKOS is a useful vocabulary for implementing a thesaurus.

- The U.S. Government would benefit from a unified approach to thesauri, especially when sharing terminology within and across Communities of Interest.

- Our approach assumes government term authors want to work in Excel, not XML/RDF/SKOS (although we permit SKOS upload).

- Other SKOS implementations are worth considering (e.g., Java-based NBII SKOS Thesaurus client).

- We hope W3C considers SKOS for the Recommendation Track.

Resources: Semantics and Thesauri

- SKOS home page   http://www.w3.org/2004/02/skos/

- XML 2005 Conference Proceedings and Slides

- W3C Semantic Web Activity home page

- Willpower Glossary of Terms Related to Thesauri

- W3C Semantic Web News and Events Archive

- SKOS: A language to describe simple knowledge structures for the web – A. Miles, XTech 2005 or paper

- SKOS Core Tutorial for DCMI 2005 – A. Miles, or PDF

- NBII SKOS Thesaurus

- SICoP: Semantic Interoperability (XML Web Services) Community of Practice – Brand Niemann et al

- Sall's Earlier Glossary Work