

INFORMATION HIGHWAYS 2002
Canada's event for the e-content community

Sound foundations - taxonomies that work

Linda Farmer, Second Knowledge Solutions

Beth Golden, Factiva

David M. Scott, NewsEdge



Information Highways 2002

Taxonomies:

Naming & Structuring Your Content

Linda Farmer
Second Knowledge Solutions
lfarmer@k2s.ca
<http://k2s.ca>



Here's What We'll Cover

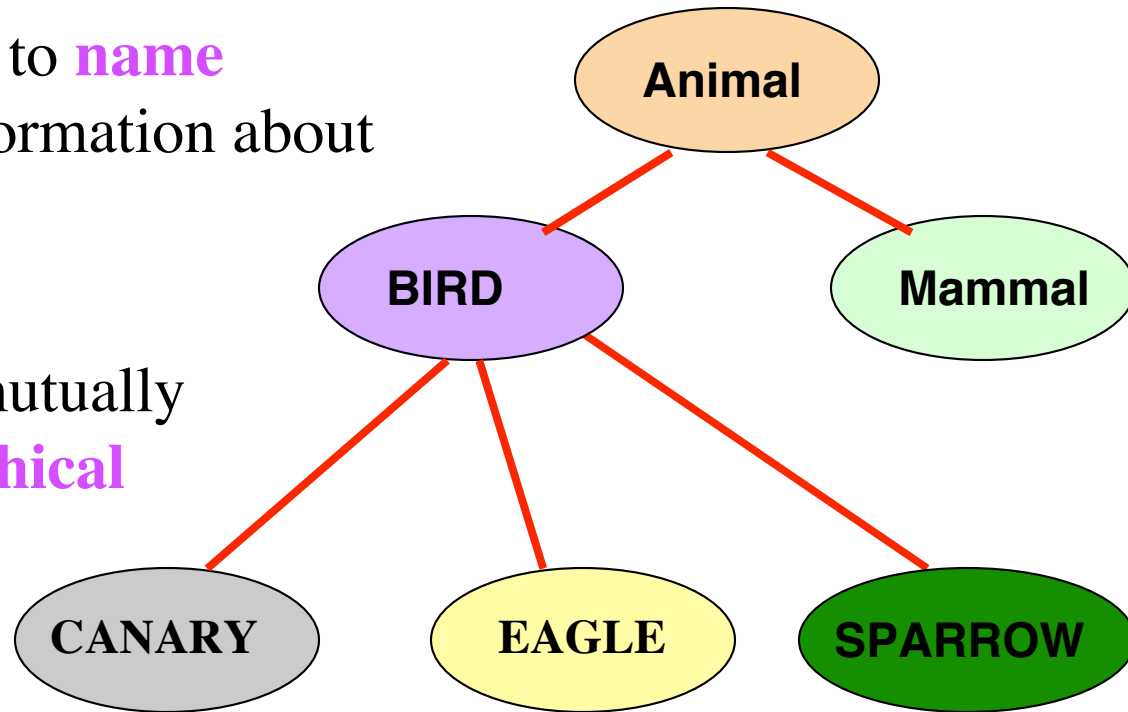
1. Building a taxonomy
2. Populating the taxonomy with content
3. Automatic categorization
4. Benefits and ROI
5. Future Directions



What is a taxonomy?

A systematic way to **name** and **structure** information about a given subject

A set of named, mutually exclusive **hierarchical categories**.



What is a thesaurus?

Vocabulary Control

A list of preferred and non-preferred terms accompanied by a standardized set of relationship indicators.

- hierarchical (BT), associative (RT) and equivalent/synonymous (UF)

Indexing/Word Mapping

Promotes consistency in indexing documents and retrieving more relevant search results

BIRD

NT CANARY

EAGLE

SPARROW

BT ANIMAL

RT BEAK

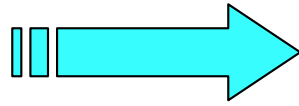
ORNITHOLOGY

WINGS

UF Aves



Thesaurus



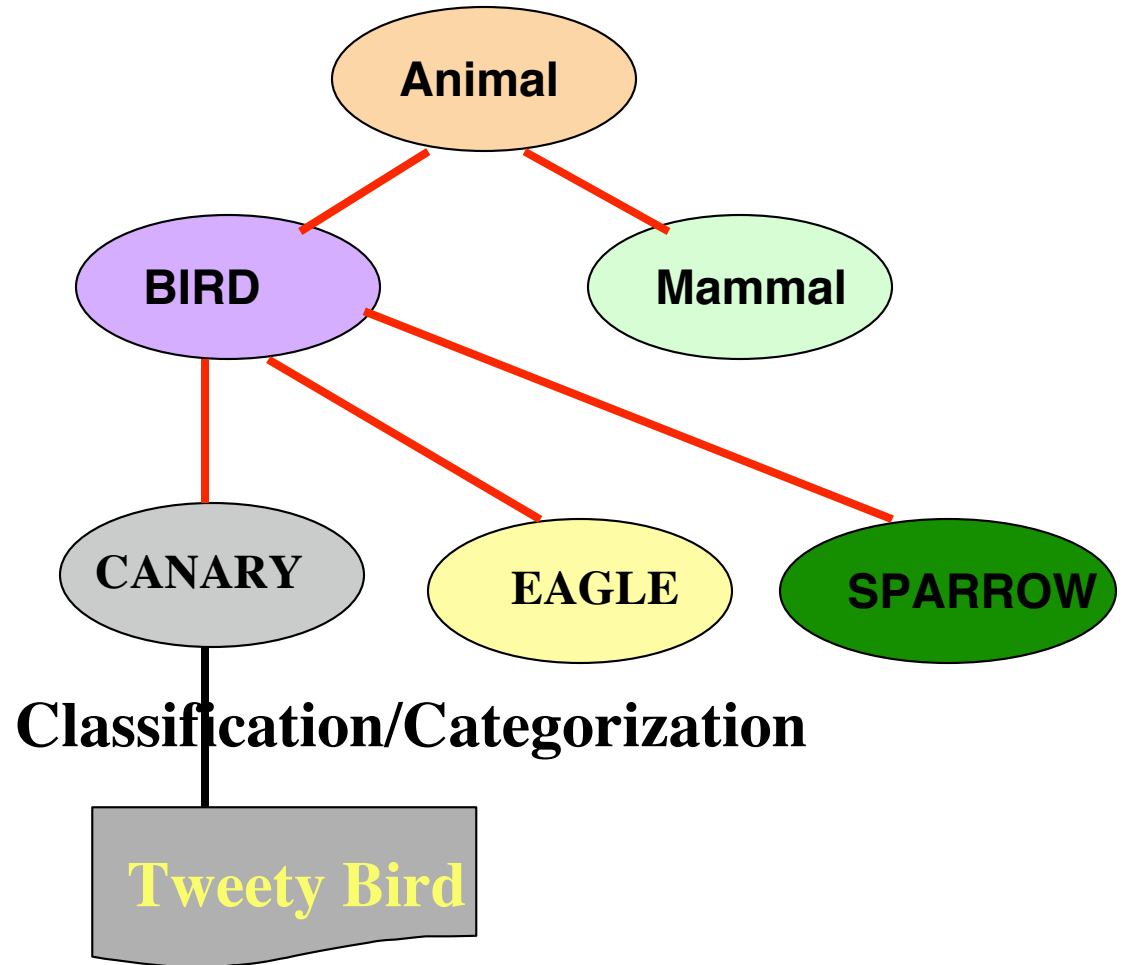
Taxonomy

BIRD
NT **CANARY**
EAGLE
SPARROW

BT **ANIMAL**

RT BEAK
ORNITHOLOGY
WINGS

UF Aves



What is an ontology?

Ontology = extended taxonomy

- Concepts, relations, facts & principles
- Catalogue of a world/domain
- How it works, how it's put together (B2B)

Thesaurus



Taxonomy



A taxonomy on steroids

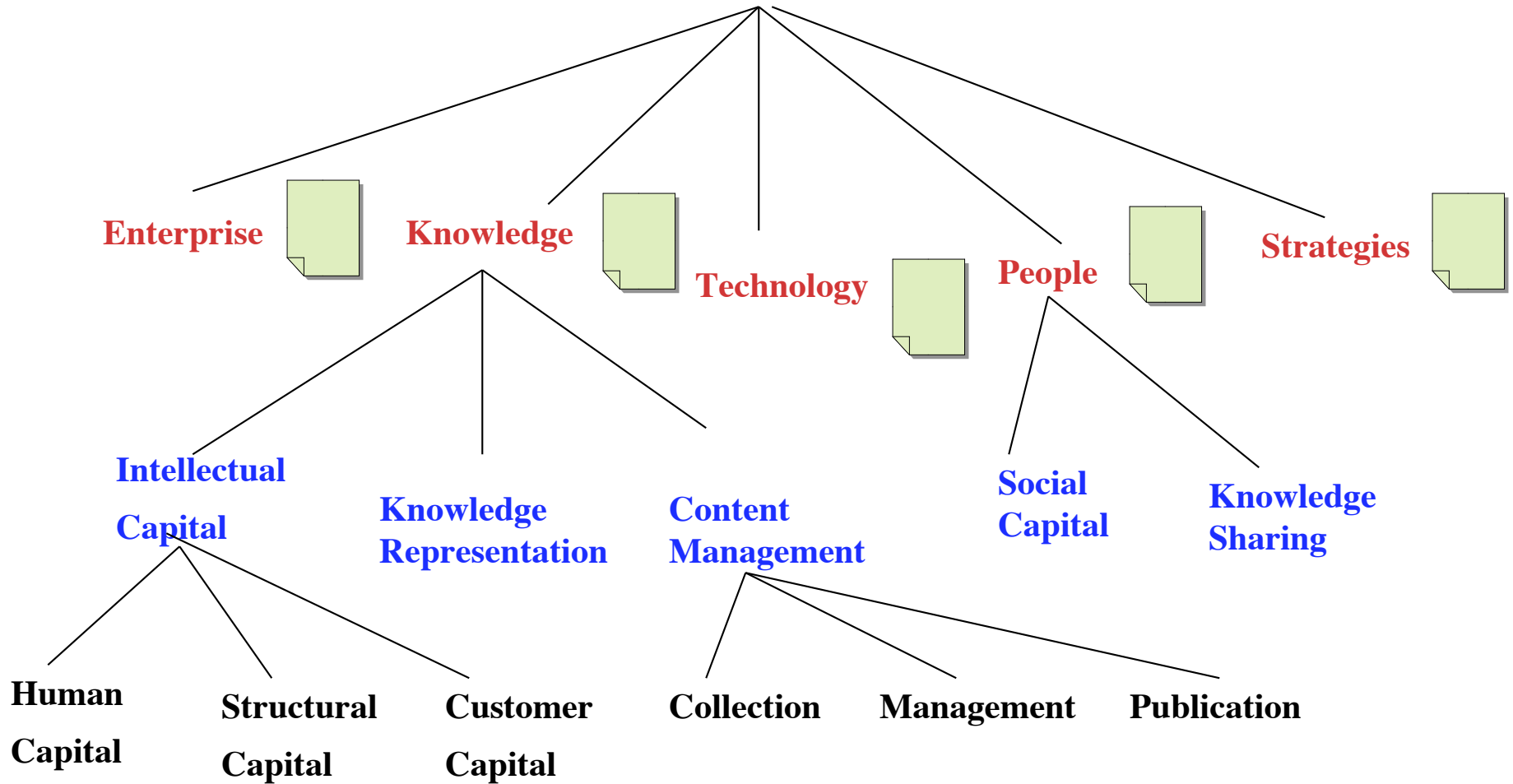


How Do You Build A Taxonomy?

1. **Business Goals:** What business process/ objective will this taxonomy serve?
2. Create a domain statement.
3. Do other taxonomies exist in this domain?
4. Bring together a sample set of documents & identify key concepts. Select name.
5. Determine relationships between concepts.
6. Create hierarchical categories for domain.
7. Choose implementation tools.



Domain = Knowledge Management





Challenges in Building a Taxonomy

1. Staff inexperienced, untrained in thinking taxonomically
2. Interesting but laborious process
3. Expensive and time-consuming part of a portal investment
4. Subject Matter Experts (SMEs) too busy
5. Translate taxonomy into a web/portal interface
6. Keeping it up-to-date



Populating Taxonomy with Content

1. Revisit existing content and assign categories from taxonomy.
2. Incorporate metadata tags for taxonomy category.



Challenges: Populating a Taxonomy

1. Existing corpus of documents is huge
Unstructured, untagged, unindexed.
*Who's going to do it? How long will it take?
Will we need to hire additional staff? How
much old stuff is worth keeping?*
2. Keeping up with large volume of new
content.



Taxonomy Maintenance

1. Large volume of new content
2. Need to adjust taxonomy categories
3. Reclassify content to incorporate changes
4. Need good technical support



Volume of Information Imposes Constraints



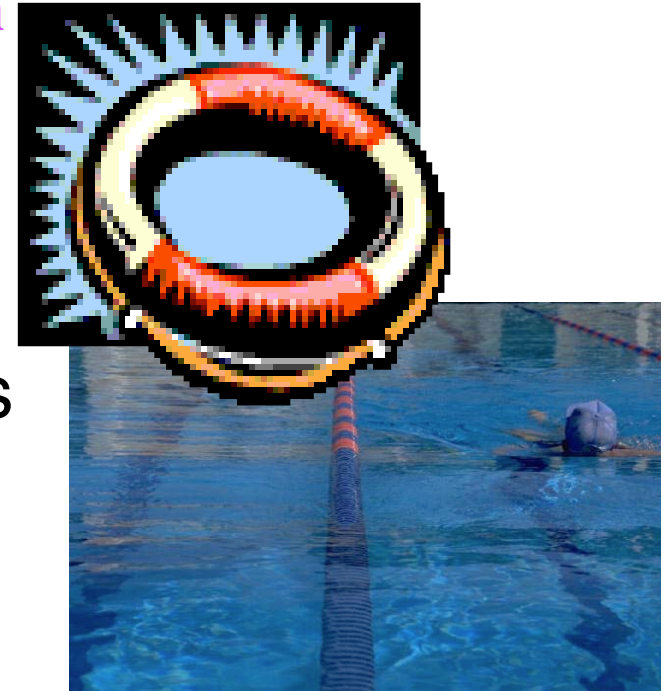
Categorization Approaches

1. Manual
2. Automatic
3. Hybrid/cyborg

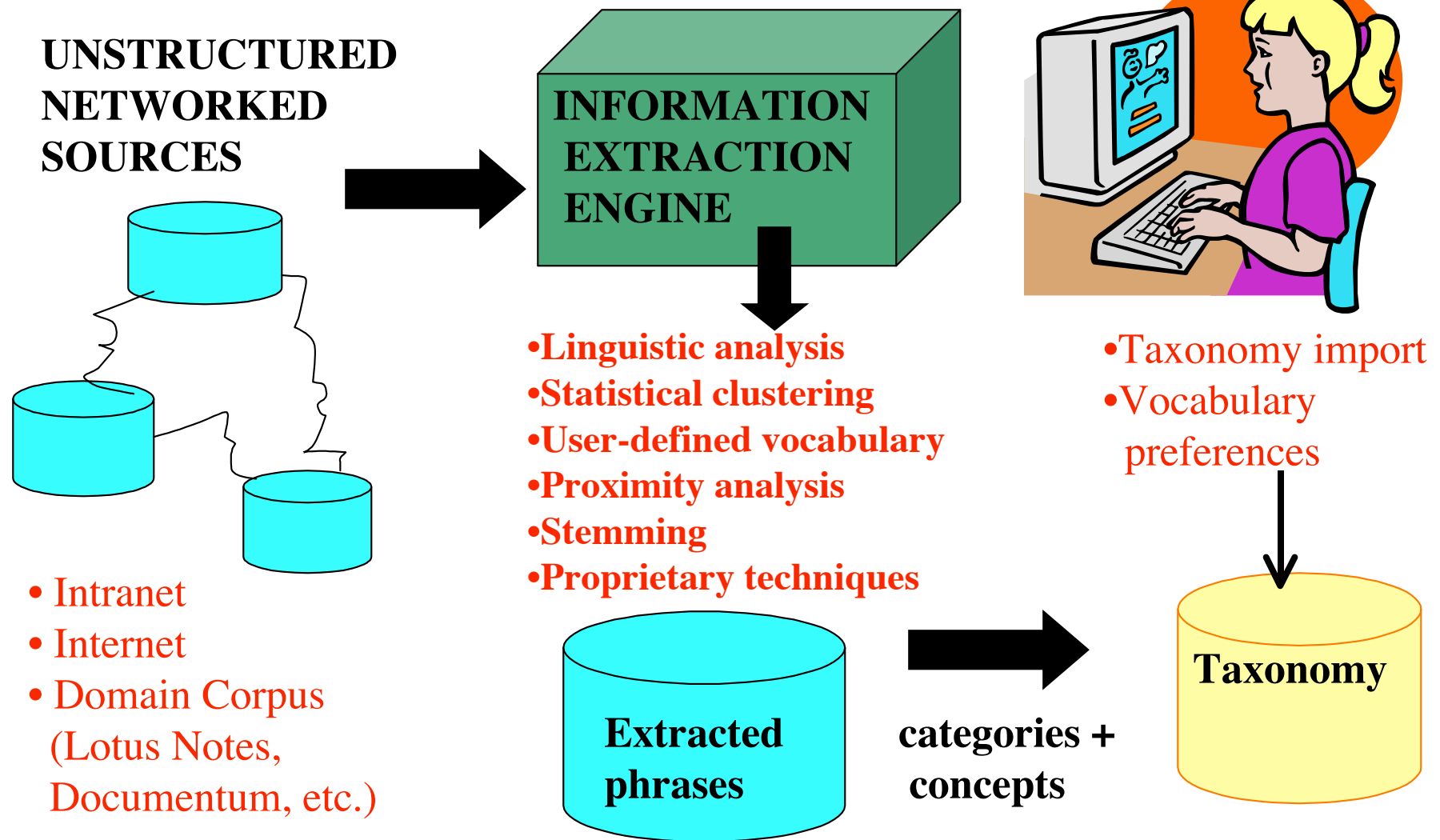
Categorization Software Tools

The new lifeline for the enterprise
swimming in unstructured information

1. Create **taxonomy** categories
2. **Classify** existing collections of unstructured content
3. Apply **metadata** to content

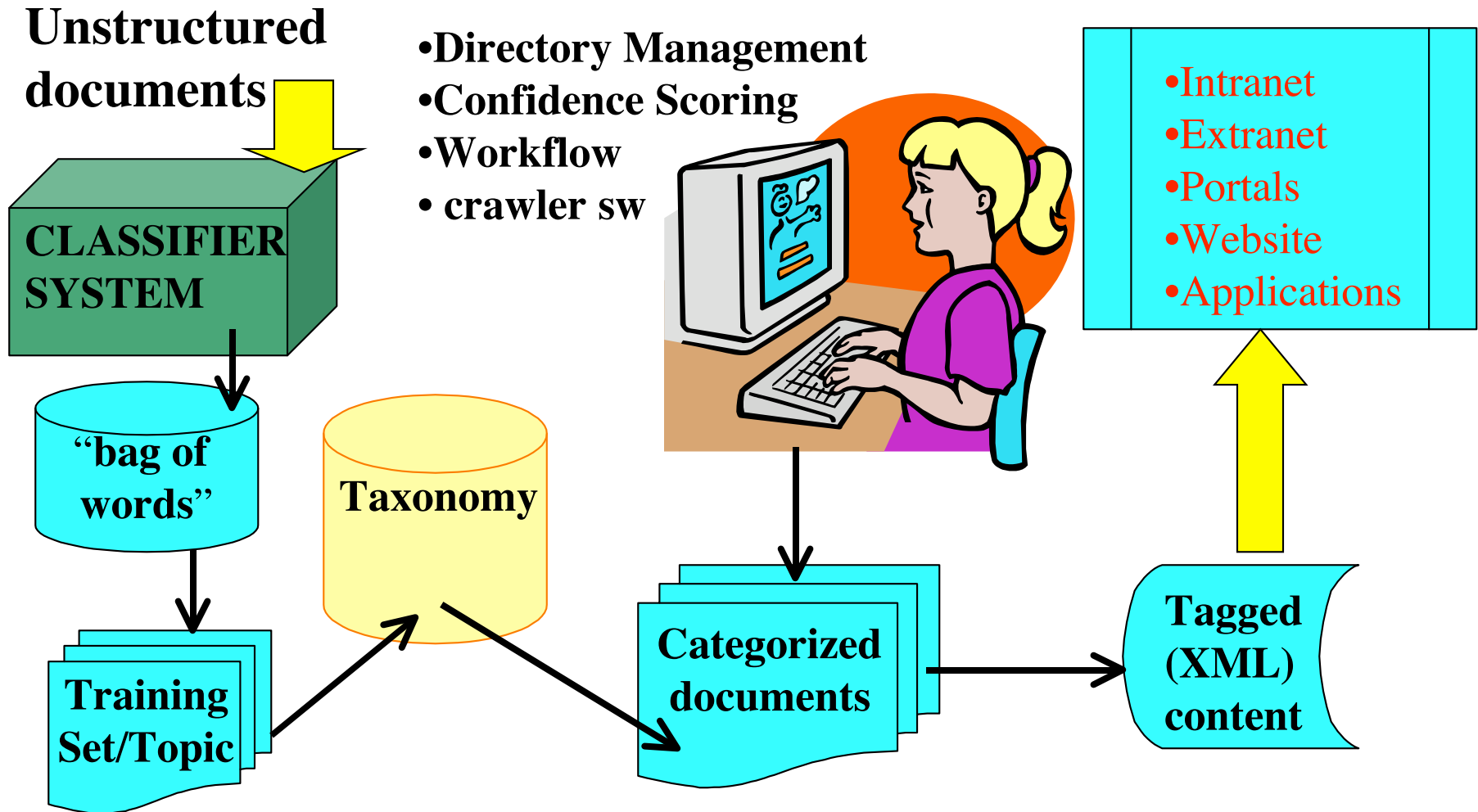


Phase 1: Taxonomy Creation

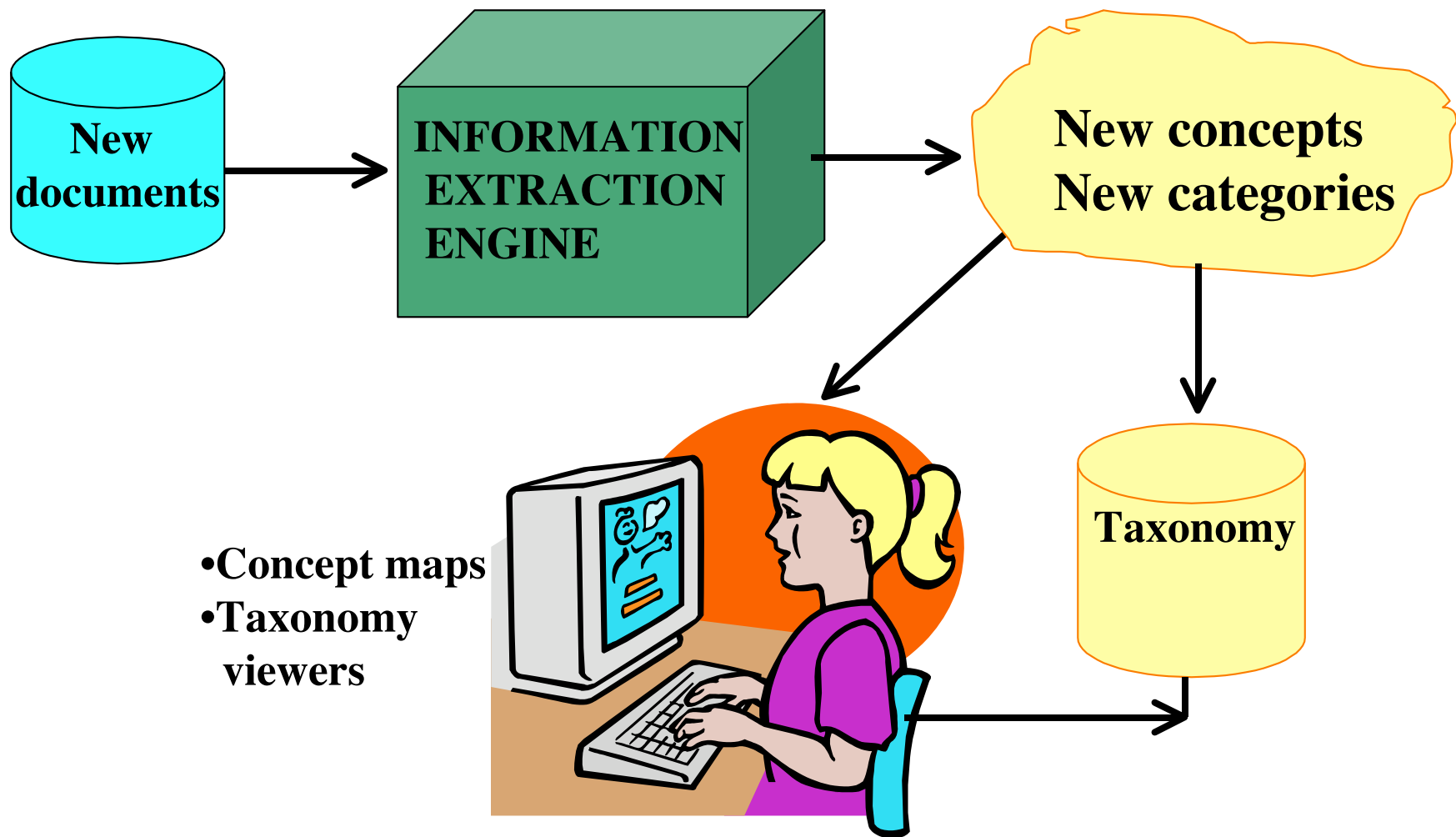




Phase 2: Categorization



Phase 3: Taxonomy Maintenance



Some Automatic Categorizers

Semio

Autonomy

**ArchiText
(Yellowbrix)**

GlobalWisdom

Cartia

Inxight



Mohomine

Quiver

Metacode



Benefits of a Taxonomy

Your information is now content:

1. Organized around a purpose
2. Named and given context & meaning
3. Structured for quick access
4. Tagged for computer manipulation



Which means that ...

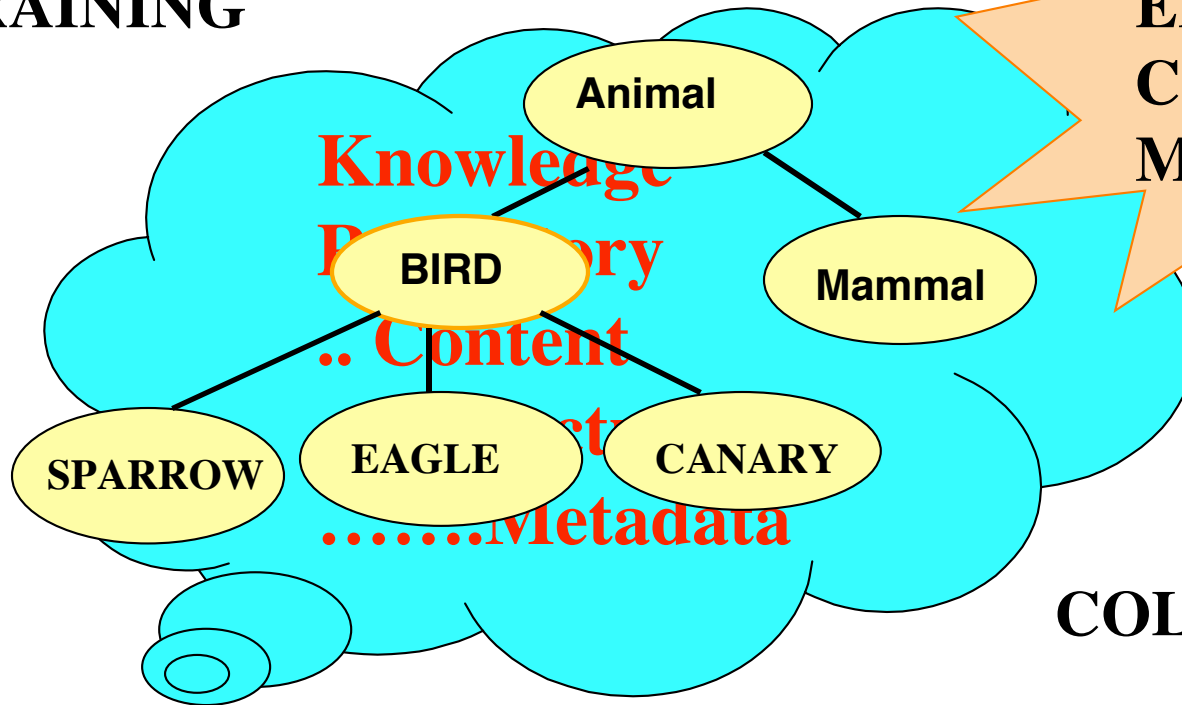
- Information is easier to find.
- It can be shared and reused.
- Duplication of project efforts is minimized
- Full text searching is enhanced, more effective
- Content quality improved with context and visibility across the enterprise
- Provides navigation paths through corporate knowledge
- Shows up what information is missing.

AND...



Taxonomy Is Now a Knowledge Framework

TRAINING



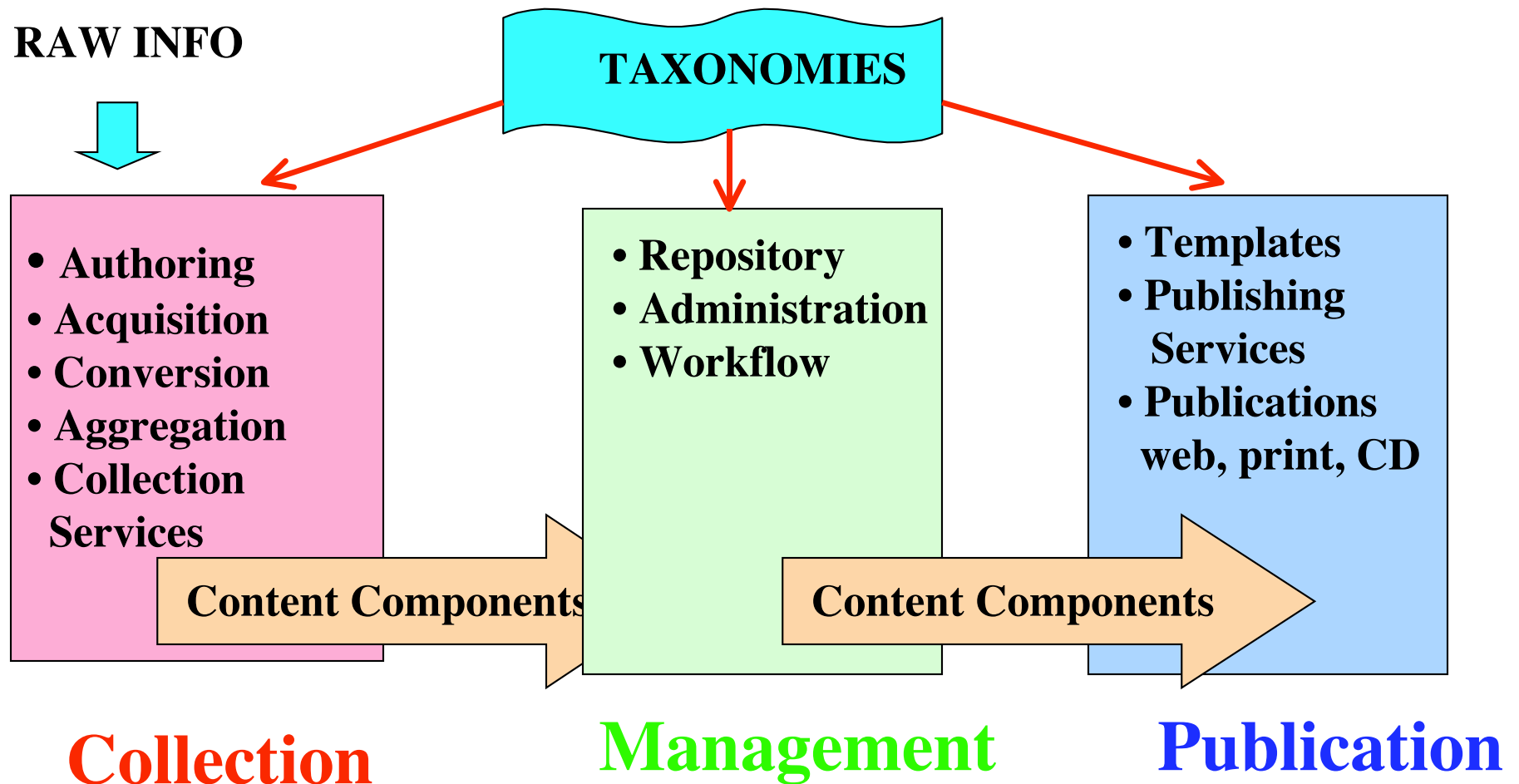
**ENTERPRISE
CONTENT
MANAGEMENT**

COLLABORATION

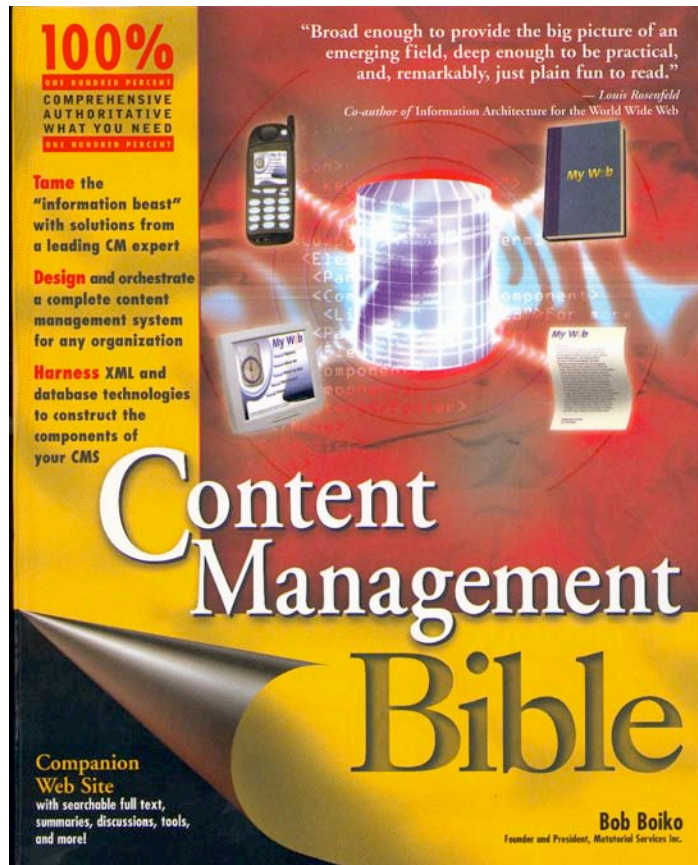
PERSONALIZATION



Content Management Process



Content Management Resource



Bob Boiko **Content Management Bible 2002**

www.metatorial.com

Upfront Expense and ROI

- What is the business case for taxonomies?
- Does the cost of creating & maintaining a taxonomy outweigh its benefits?





Some Costs

- Categorization software
\$125,000 +
- Content management software
- Staff to work on setting up taxonomy
- Continual maintenance
- Modifications to content management processes
- Technical staff support



Some Returns on Investment

- **Enhanced productivity**
e.g. 30% reduction in time searching for info.
10 - 40% increase in productivity.
- **Improved staff effectiveness**
Find info. easier. Reduced development cycles,
effective sales calls, enhanced customer support, etc.
- **Streamlining of processes**
Enhanced info. sharing with reduction in duplication
of effort



The Big Questions

Can our enterprise afford what it takes to effectively organize our knowledge?

Do we have any alternatives or is this the cost of doing business?

Are there other ways to remain competitive in an information-critical marketplace?



One Person's View

“If you think knowledge is expensive,
try ignorance.”

- Marc Auckland, Chief Knowledge Manager,
British Telecom
At a CKO Summit, September 2000

Future Trends

3rd wave of Internet-related software - the “**semantic wave**” involving meaning and understanding.

High performance **knowledge processing**



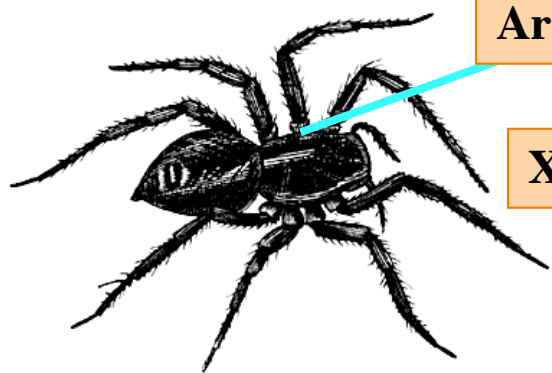
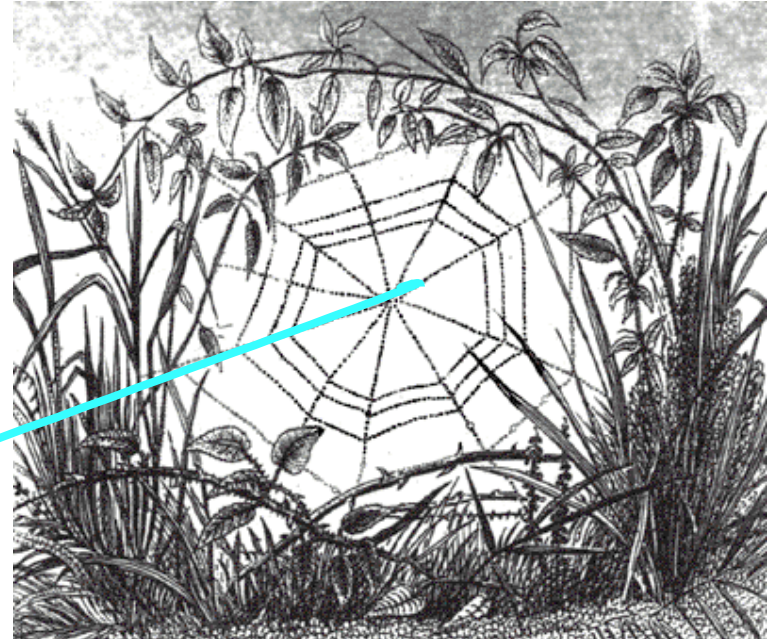
Knowledge Bases (KBs)

- Merging of collaborative multiple KBs created by distributed teams of domain experts

Spinning The Semantic Web

The “**smart network**” that understands the meaning of words and the logical relationships among them.

Replaces the “**web of links**” with the “**web of meaning**”.



Artificial intelligence

XML/RDF

Databases

Natural Language Processing

Intelligent Agents

Distributed Ontologies

Topic Maps

Berners-Lee & W3C (www.w3.org)

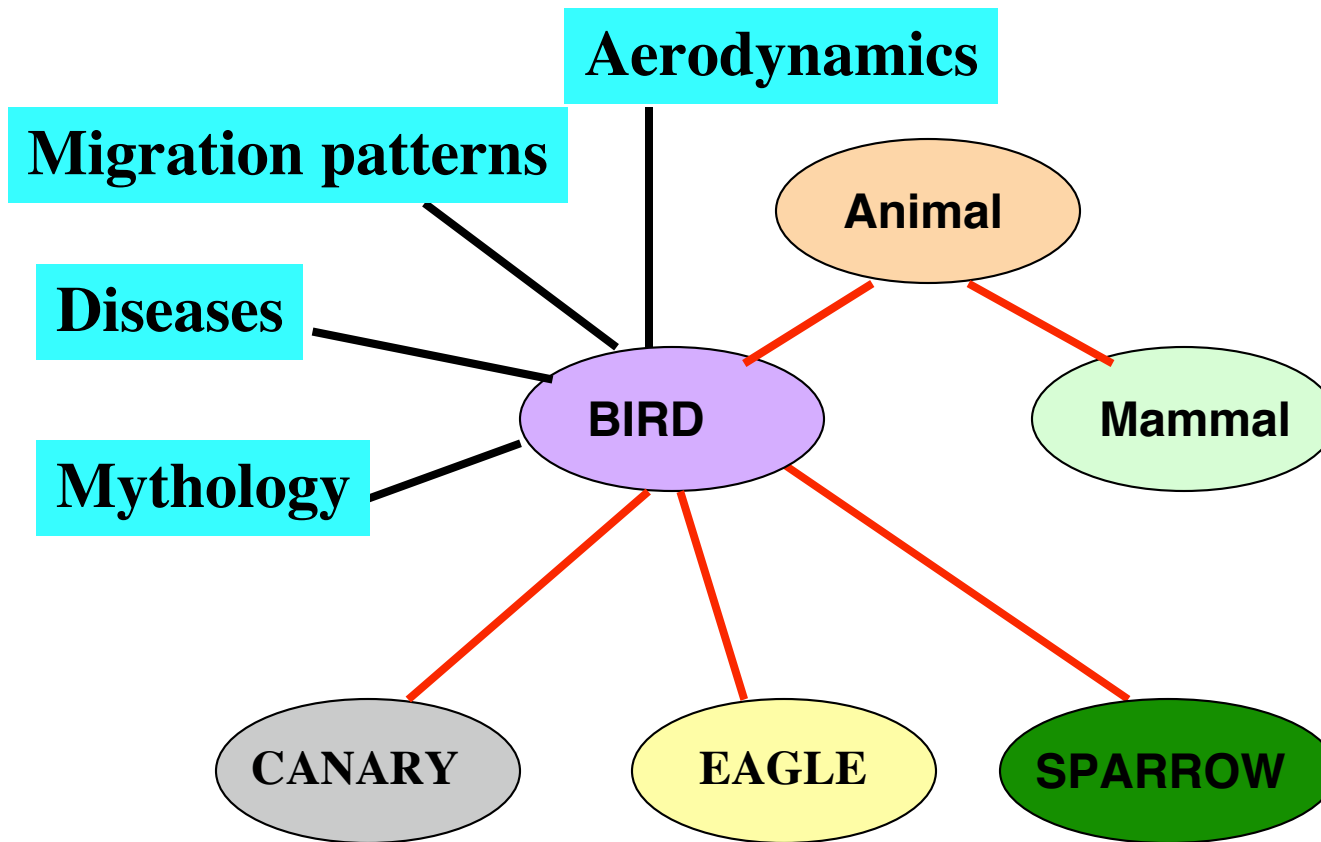
Distributed Ontologies

- Taxonomy + metadata about the **properties** for each category or class
- Encoded with logic-based language that enables automated reasoning by **intelligent agents** and web applications
- Goal: Provide highly reusable, extensible, long-lived semantic structure for content



Taxonomy on steroids

Ontology for Class Named "Bird"





Application Areas

- E-commerce
- Enterprise integration
- Digital libraries
- Medicine
- Biology
- Bioinformatics
- Geographic information systems
- Legal information systems



Topic Maps- Hot New Technology

- New paradigm for knowledge navigation and synthesis (*based on the index to a book*)
- Provide a navigation map or style sheet for an information set
- Emerging ISO standard (XTM 1.0)
www.topicmaps.org

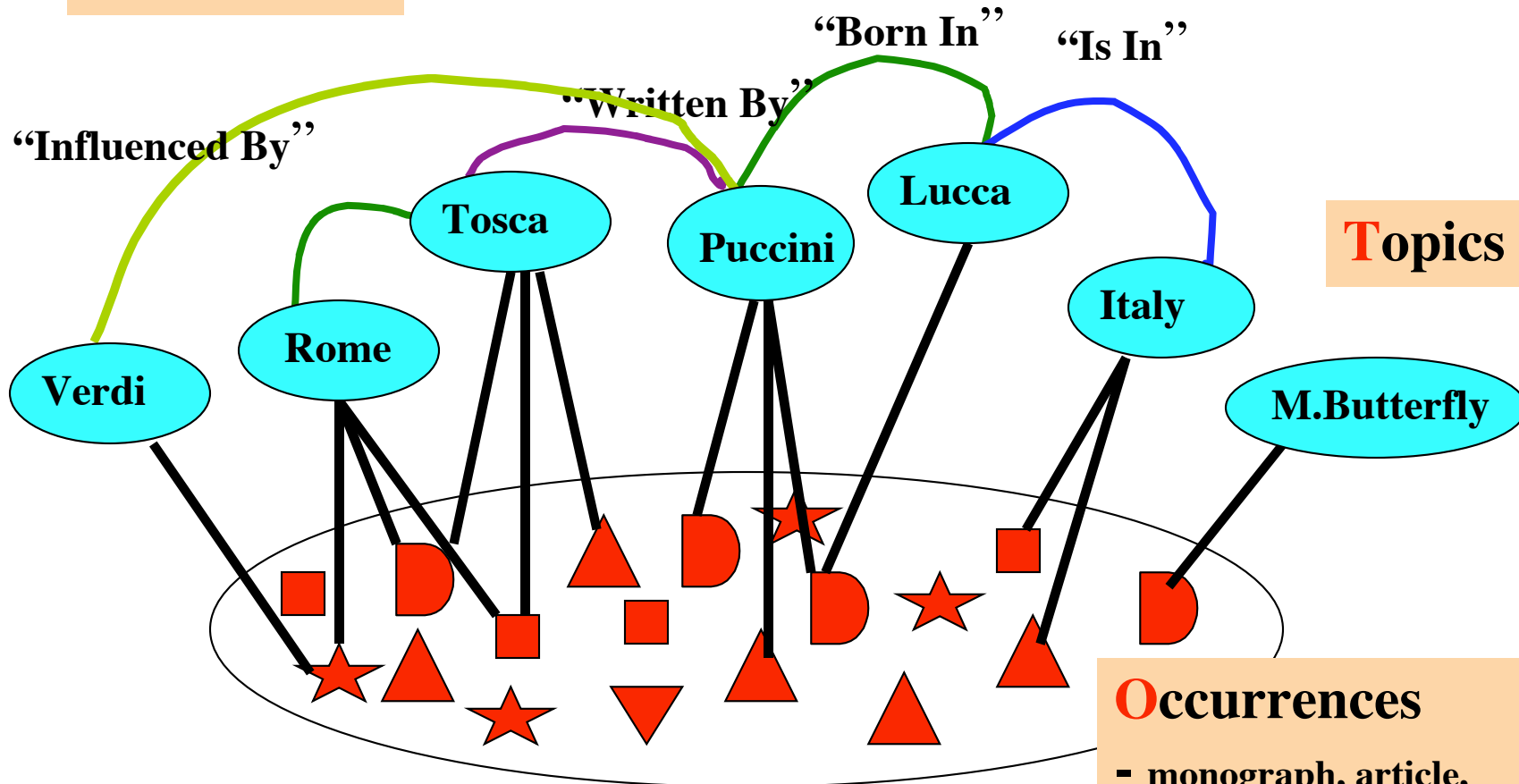


The TAO of Topic Maps

Info. Set = **Opera**

Associations

Topics



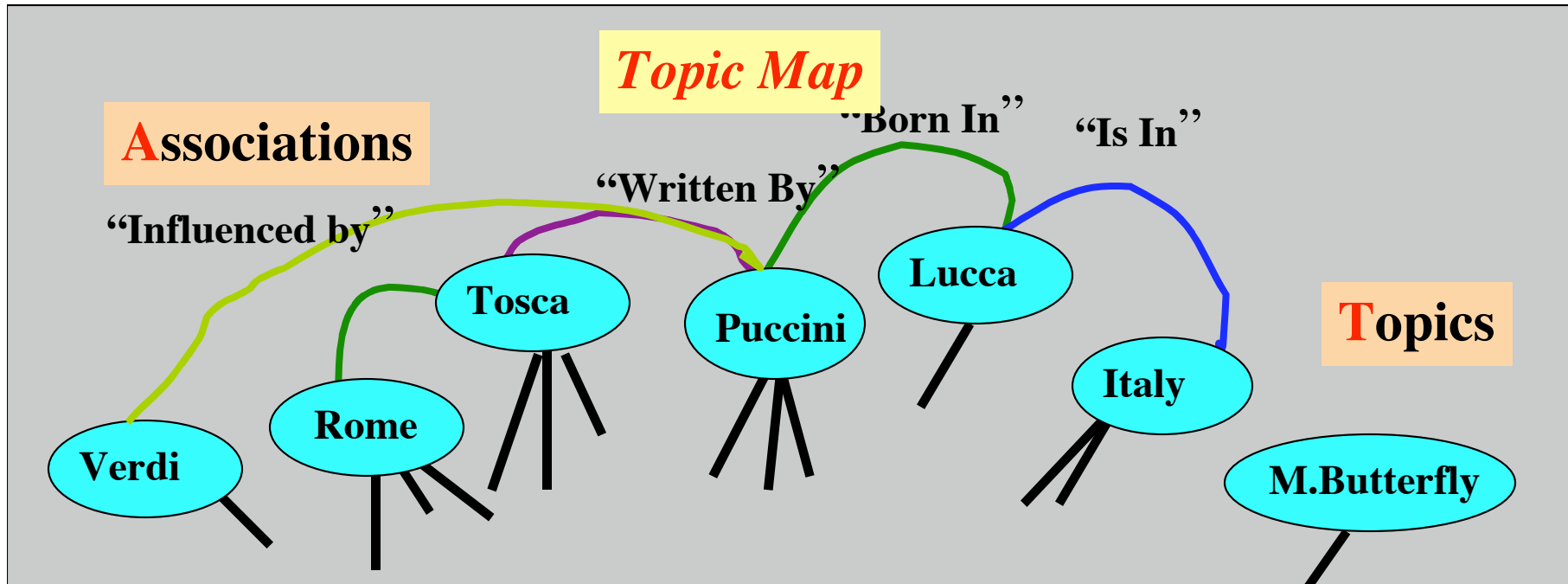
Occurrences
- monograph, article,
picture, commentary, etc.



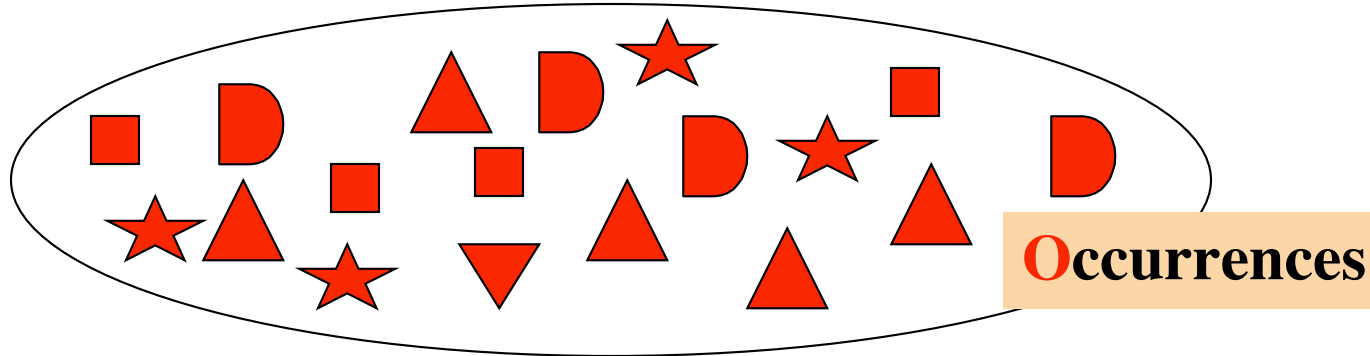


A Topic Map

RDF/XML/XTM



Info. Set =
Opera





Applications of Topic Maps

- Enhance & extend existing taxonomies
- Provide customizable, personalized routes to information
- Build a structured semantic network link over web, portals and intranets.
- Apply multiple maps (views) to the same information pool
- Maps collectable, interchangeable



Parting Words

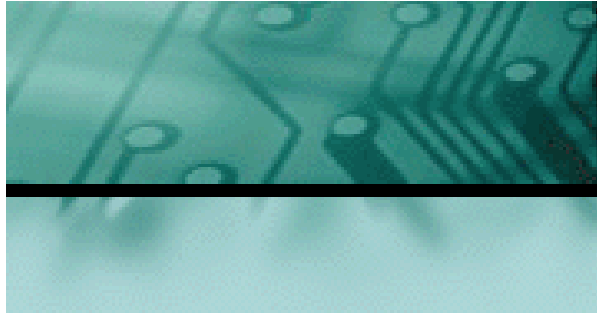
- Create a taxonomy that reflects the needs and organizational logic of your business - nothing more, nothing less.
- Plan to commit enough people with expertise in the subject matter to create and maintain the taxonomy
- Keep up with & evaluate the taxonomy technology.
- Used a phased approach to implementation
- Make enterprise content management part of your taxonomy initiative.



Linda Farmer

lfarmer@k2s.ca

<http://k2s.ca>



INFORMATION HIGHWAYS 2002
Canada's event for the e-content community

Sound foundations - taxonomies that work

