Mapping Taxonomies, Thesauri, and Ontologies

presented by

Heather Hedden

Hedden Information Management

SEMANTiCS
Karlsruhe 2019

About Heather Hedden

- Taxonomy consultant
  – Independent, through Hedden Information Management (since 2004)
  – Employed, through Project Performance Corporation, and contract
- Former staff taxonomist
  – At various companies: Gale/Cengage Learning, Viziant, First Wind
- Instructor of online and onsite taxonomy courses
  – Independently through Hedden Information Management
  – Previously at Simmons University - Library & Information Science School
- Author of *The Accidental Taxonomist* (2010, 2016, Information Today, Inc.)
- Former indexer of books and database content (articles, images, etc.)

Outline

1. Introduction to mapping knowledge organization systems (KOS)
2. Situations for KOS mapping
3. Method of mapping
4. Mapping examples
5. Standards for mapping
6. Tools for mapping
7. Mapping case study

Introduction

Knowledge organization systems (KOS):

- Taxonomies
- Thesauri
- Ontologies
- Other controlled vocabularies

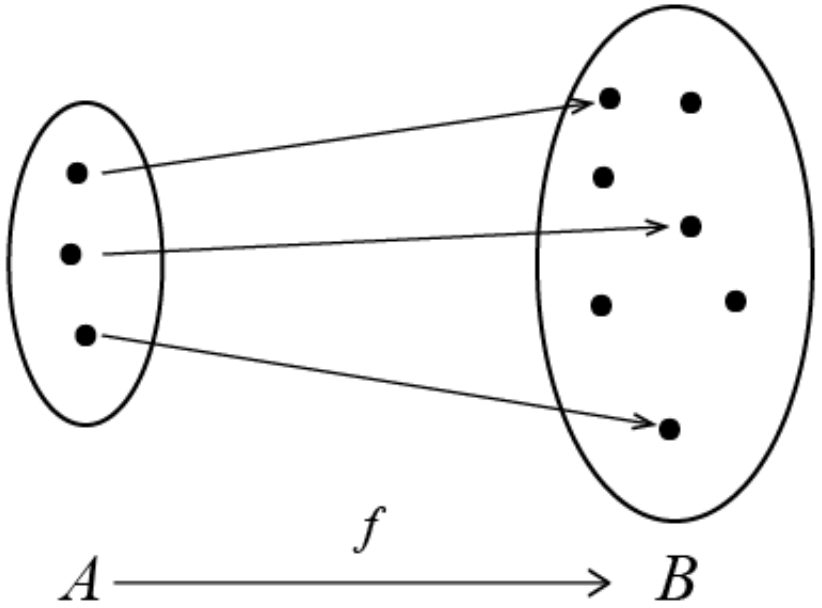
- Usually created for a specific use (specific content and audience)
- Occasionally created for wider, shared use
- Often are enhanced, or extended or adapted for additionally uses

Introduction

Mapping knowledge organization systems (KOS)

- A form of linking knowledge organization systems together
- Linking individual concepts in one KOS to concepts in another.
- Retaining them each as a distinct KOS.
- A KOS continues to be used for its original purpose plus added use through the mapped KOS.

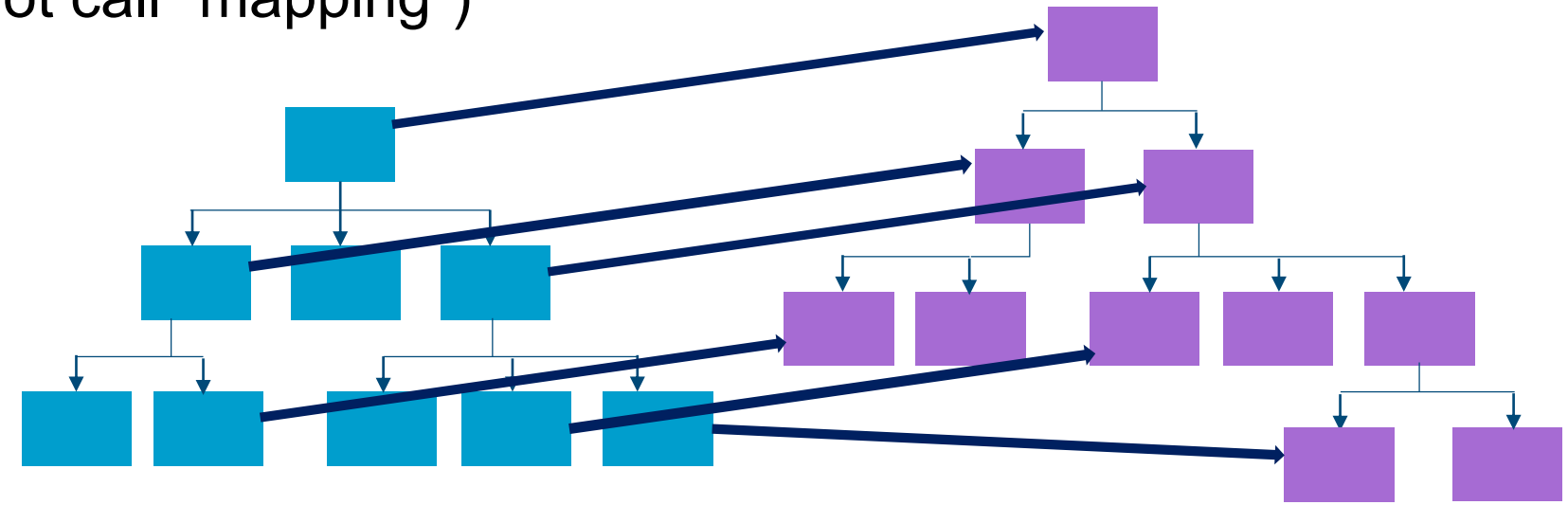
The name “mapping” might come from mathematical set theory, whereby elements in one set are mapped to elements in another set.



Introduction

Mapping types

- Directional from one KOS to another with sufficiently equivalent links, so that one KOS may be used for another.
- Directional from a term set to a KOS with equivalent and hierarchical links, so that a KOS can be enriched with added concepts.
- Bidirectional, with equivalent links, so that content can be shared.
- Bidirectional, with associative links, so that users can navigate to new content. (Might not call “mapping”)



Introduction

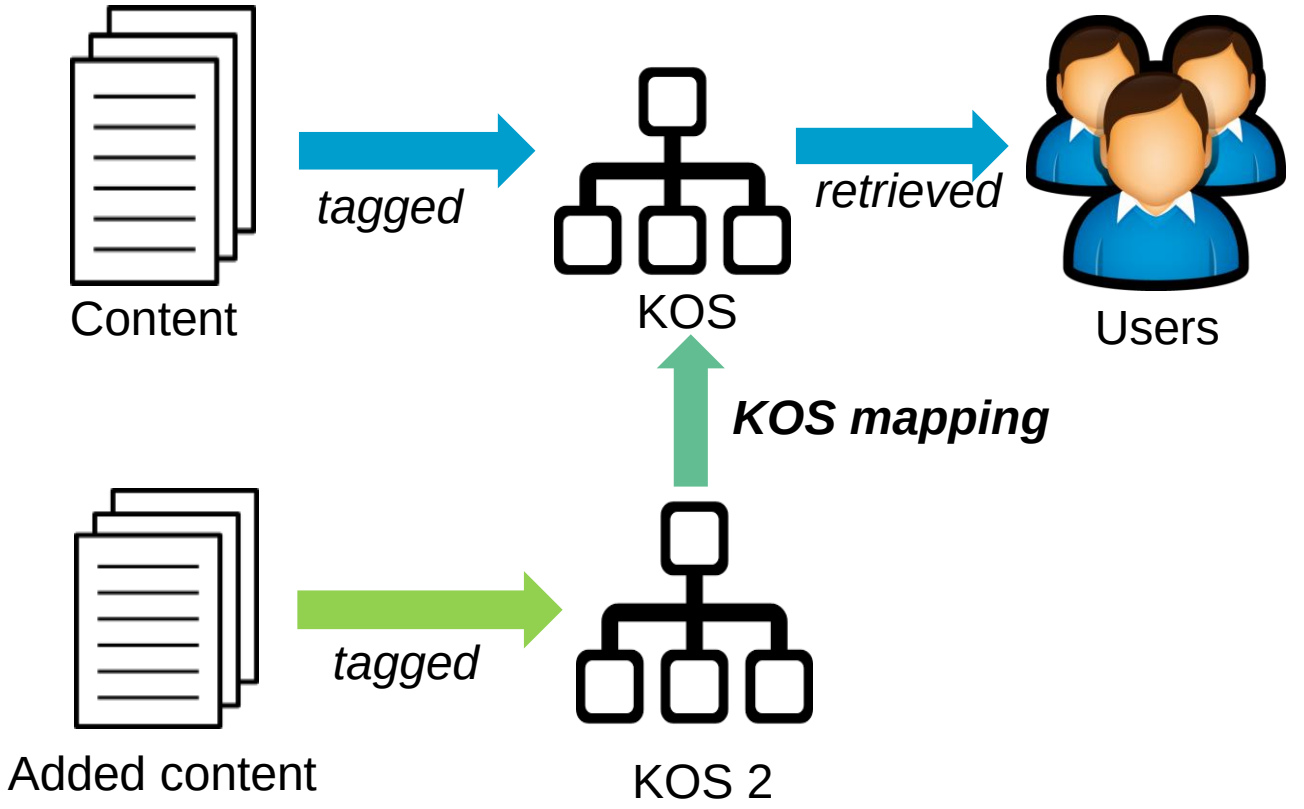
Crosswalk – a table of mappings between concepts in two or more structured vocabularies.

- Depending on systems used, a designated crosswalk table may or may not be created.
- A KOS managed in software with a mapping feature does not require a crosswalk, but a crosswalk file can be generated/exported.

Situations for KOS Mapping

An expanded set of content, tagged with a different KOS, will be retrieved by users with their existing KOS.

- The organization continues to provide only its KOS to its users to retrieve both its own content and added content.

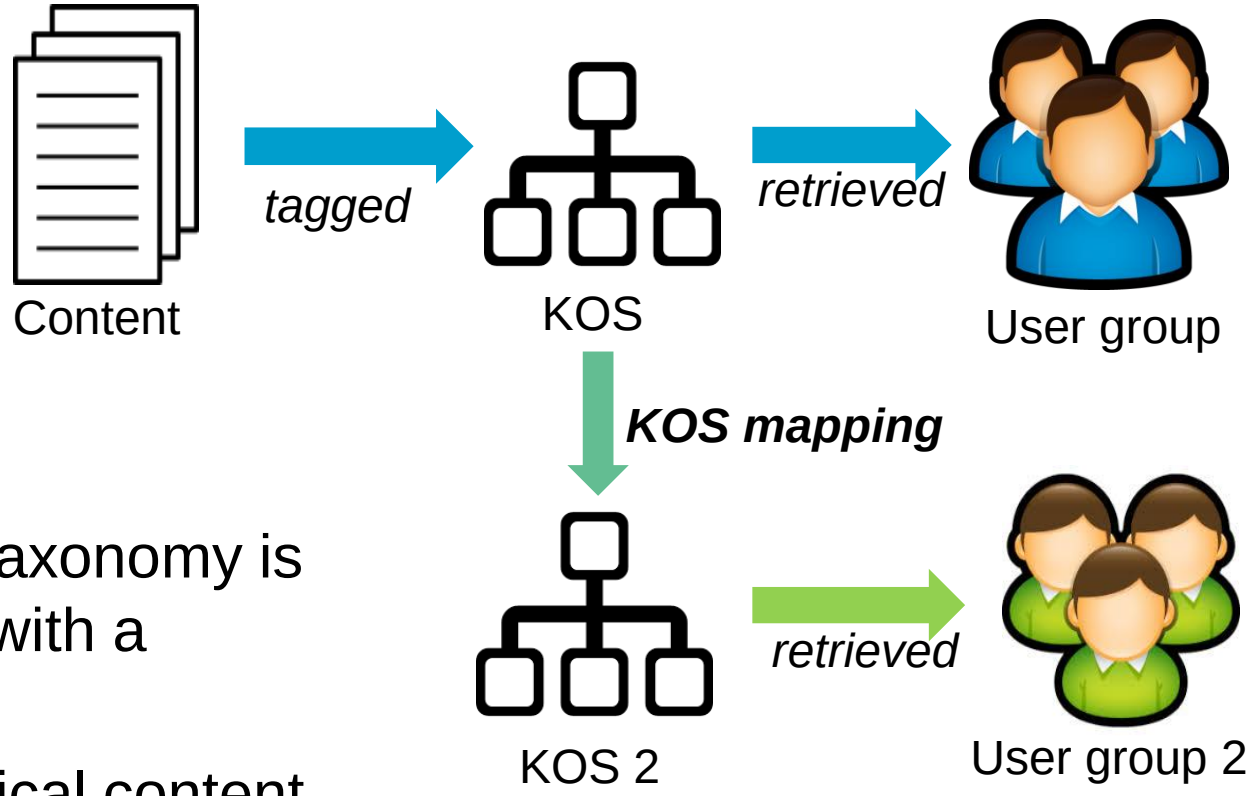


- A content publisher with a KOS partners with a specialized information vendor, with its own KOS, to expand its content offering.
- An organization with a KOS tagged to its internal content licenses content from an external source that is tagged with a different KOS.

Situations for KOS Mapping

A set of content will be retrieved by different audiences, each accessing their own KOS.

- Identical content will be retrieved by end-users with a new KOS. Rather than re-index, the new KOS (or more than one) will be mapped to the existing KOS.

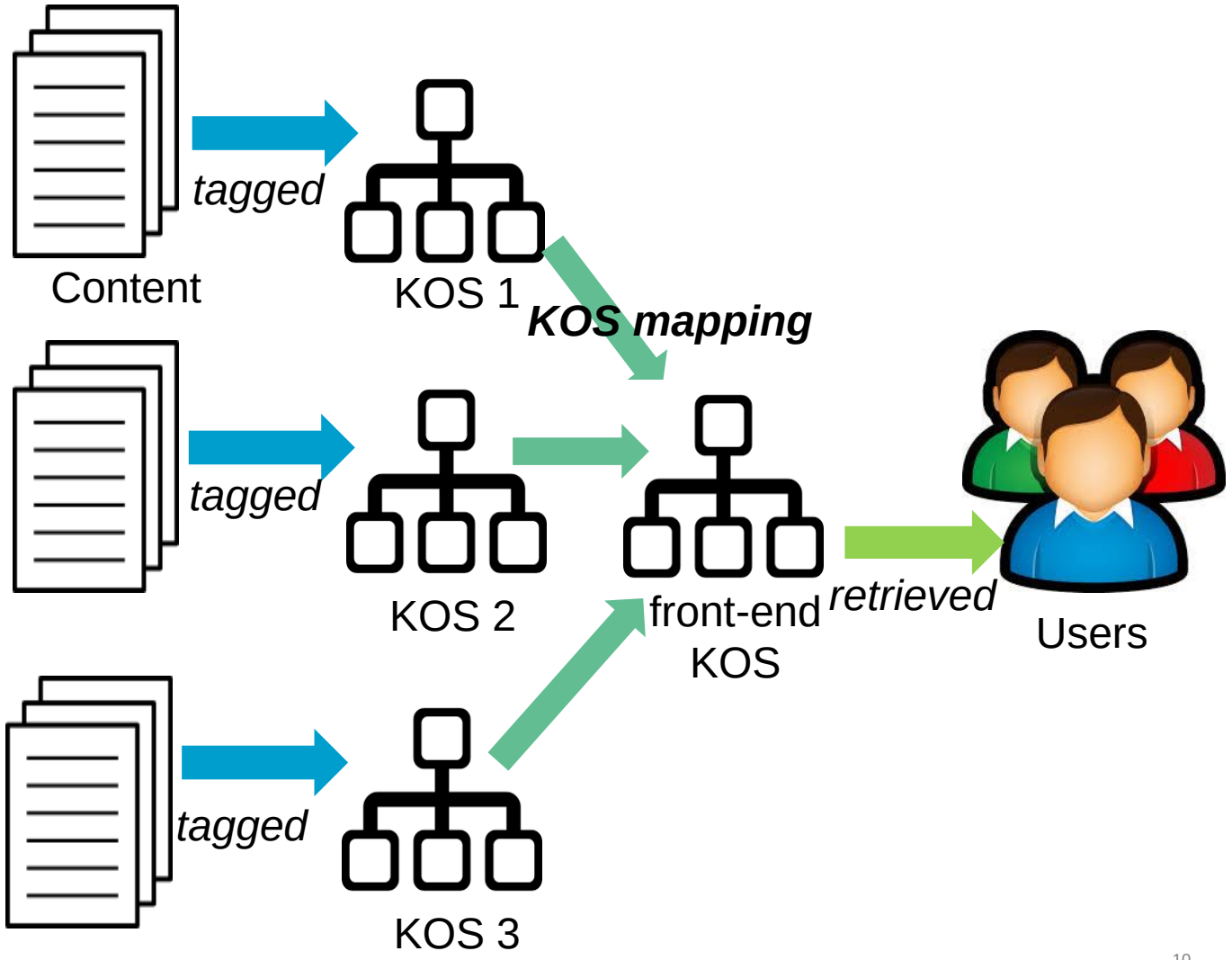


- Selected content with an enterprise taxonomy is made available on a public web site with a different public-facing taxonomy.
- A provider of scientific/technical/medical content with a technical thesaurus creates a simpler taxonomy aimed at laypeople.
- Content will be made available in a different language region (locale), and a comparable KOS already exists in that other language.

Situations for KOS Mapping

A front-end KOS will be used to retrieve various content sets, each tagged with its own KOS.

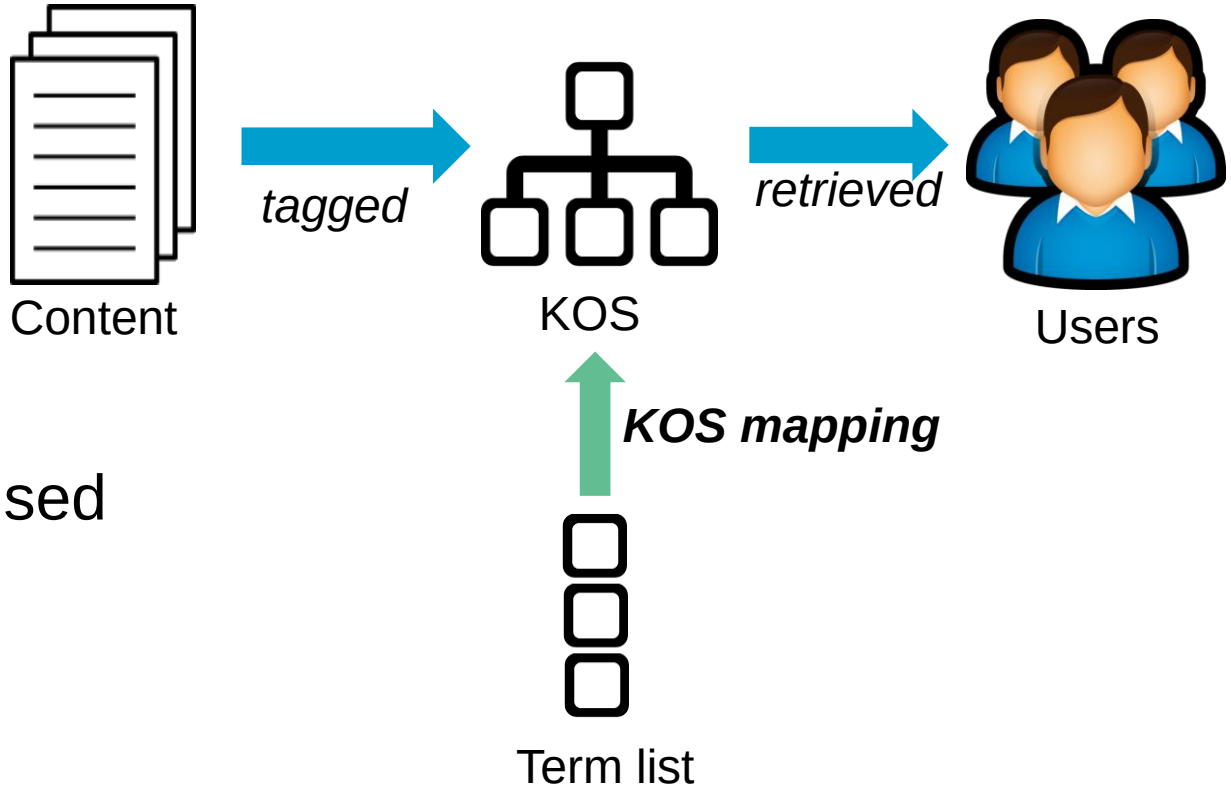
- A vastly expanded set of content can be accurately retrieved.
  - A knowledge graph is built to aggregate data from multiple repositories or data silos, each with its own KOS.
  - An enterprise search is based on “federated search.”
  - A search engine product taxonomy is mapped to, in order to increase SEO.



Situations for KOS Mapping

A term list is mapped to a KOS to enrich the KOS.

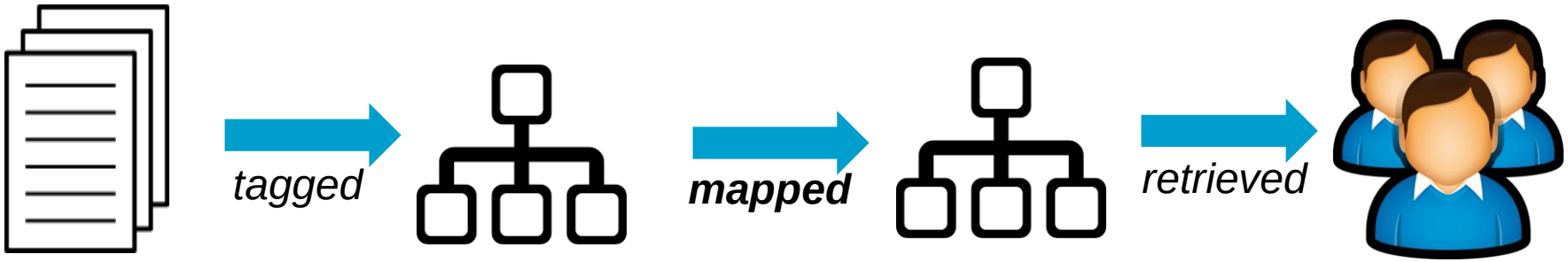
- A vastly expanded set of content can be accurately retrieved.
  - Terms from search engine logs are mapped to a KOS to add alternative labels.
  - Terms from an open source or licensed vocabulary are mapped to a KOS.



Method for Mapping

Mapping methodology/theory

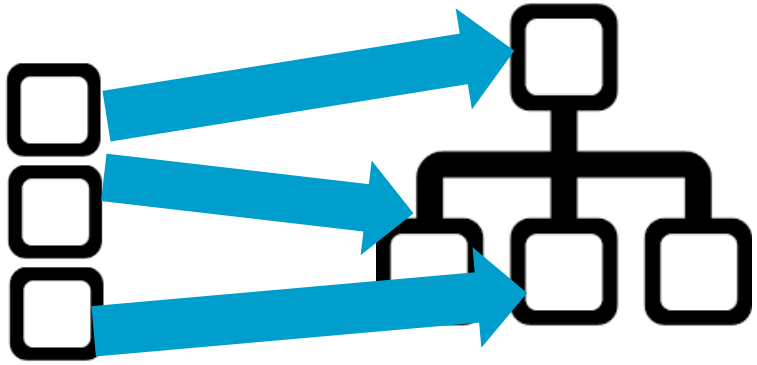
- Mapping direction: ***from*** a tagged taxonomy (source) ***to*** the retrieval/user-interface taxonomy (target)
- Consider the tagged-taxonomy/source terms as variants (alternative labels) for the retrieval taxonomy/target terms.
  - Equivalent meaning is for the context.
  - Narrower-to-broader matches are OK: a narrower concept in a tagging taxonomy may be mapped to a broader concept in the retrieval taxonomy, if no equivalent exists in the retrieval taxonomy.
  - Many-to-one mappings are OK.



Method for Mapping

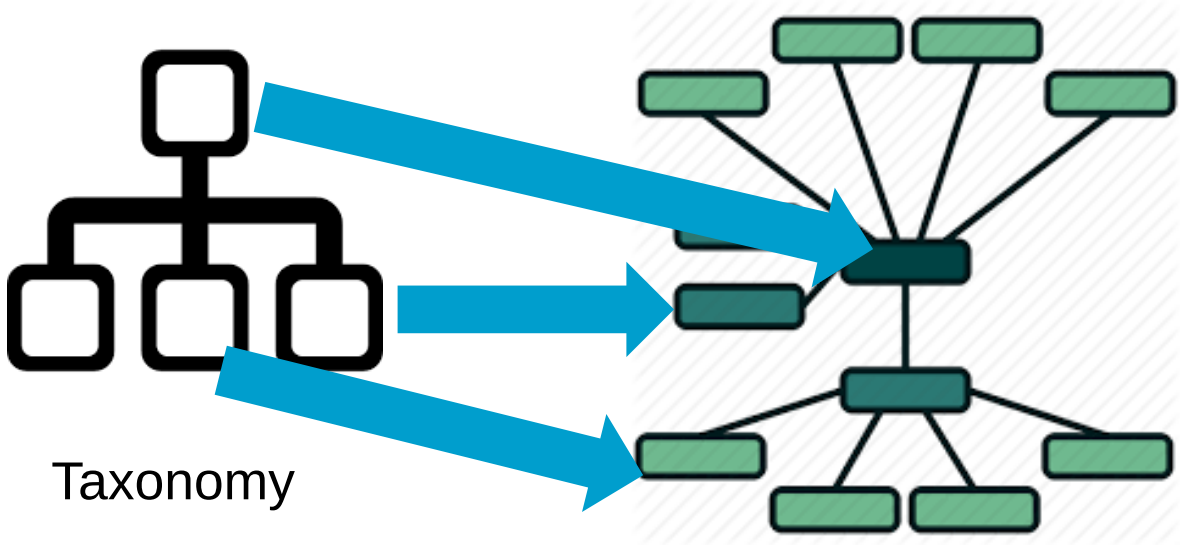
Mapping methodology/theory

- Focus on the meaning of concepts.
- Relationships between concepts *within* a KOS generally do not matter.
- Can map between term lists, taxonomies, thesauri, ontologies
- The type of KOS does not impact the direction of mapping, although the usual case is from simpler to more complex KOS.



Term list

Taxonomy



Taxonomy

Ontology

Method for Mapping

Directional mapping is easier when:

- The scope of both is identical.
- The retrieval KOS has fewer terms than the tagged KOS.
- The tagged KOS is more specific/granular than the retrieval KOS.

Directional mapping is more complex when:

- Mapping from a hierarchical taxonomy to a faceted taxonomy
- There is inconsistency, and one KOS is more detailed (with more specific/granular concepts) in some areas, and the other KOS is more detailed in other areas.

Directional mapping does not work when:

- From a *faceted* taxonomy to a hierarchical taxonomy, thesaurus, or ontology

Method for Mapping

Mapping technique/steps

- Identify which KOS is the tagged/mapped-**from** taxonomy, and which KOS is the retrieval/mapped-**to** taxonomy.
- Use a software tool or scripts to compare both, to obtain exact matches and close matches.
- Human review confirms and approves automatically proposed close matches.
- Human review attempts to identify mappings for unmatched concepts, but some will remain unmapped and cannot be utilized.
- If all tagged content is required for inclusion, then new concepts need to be added to the retrieval KOS.

Method for Mapping

Automatic mappings, without requiring review, comprises:

- Exact match concepts, ignoring only capitalization and diacritics
- Concept in tagged/source KOS is an exact match to a synonym (alternative label) of a concept in the retrieval/target KOS

Automatic suggested mappings for human review, comprises:

- Keyword matches – all the same words, but can be in any order
- Stemmed keyword matches – same words, any order, but also includes plural/singular and certain grammatical variants
- Concept label phrase *within* another concept label – if the retrieval KOS concept is within the tagged KOS concept label, it's usually a good match. (The latter is longer and likely qualified, and thus more specific.)
- Combinations of above

Method for Mapping

Match Type	Tagged KOS Concept	Retrieval KOS Concept
Auto-match, needs no review		
Exact match	Information technology	Information Technology
Exact synonym match	Banknotes	Currency *altLabel* Banknotes
Auto-match + Review		
Keyword match - yes	Financing debt	Debt financing
Keyword match - no	Industry news	News industry
Stemmed keyword match - yes	Data security	Secure data
Stemmed keyword match - no	Fair trading	Trade fairs
Phrase within phrase - yes	Geothermal power plants	Power plants
Phrase within phrase - no	Computer hardware & software	Computer hardware
Multiple words within - yes	Danish language books	Danish books
Multiple words within - no	Public health education	Public higher education

Mapping Examples

Review example

	A	B	C
1	Programmable logic controllers	ok	Programmable controllers
2	Programmable logic devices	ok	PLDs (Programmable logic devices)
3	Programming (Computers)	ok	Computer programming
4	Progressivism (United States politics)	b	Progressive movement
5	Prohibited books	ok	Banned books
6	Project method in teaching	ok	Project method (Education)
7	Projectile points	ok	Projectile points (Archaeology)
8	Projection	n	Projection (Drawing)
9	Projection televisions	ok	Projection television sets
10	Prolactin	n	Prolactin test
11	Proletariat	ok	Working class
12	Prolog (Computer program language)	ok	Prolog (Programming language)
13	Promethazine hydrochloride	b	Promethazine
14	Promoters (Entertainment)	b	Promoters
15	Promotion (School)	ok	Student promotion
16	Pronghorn antelope	ok	Pronghorns
17	Propaganda, American	ok	American propaganda

Colum A:
Tagged taxonomy (from)

Column B:
Retrieval taxonomy (to)

Column C:
Human review notes:
“ok” is equivalent,
“b” second term is broader so also ok,
“n” is narrower or otherwise not acceptable.

Mapping Examples

Review example

		Candidate_ CV_Terr	CV_Terms_ Y
Makes	GVX	y	
Type of Vehicles	4 Wheel Drive	y	y
Type of Vehicles	Four Wheel Drive	y	y
Type of Vehicles	4x4	y	
Type of Vehicles	4 X 4	y	
Type of Vehicles	4x4s	y	
Type of Vehicles	4WD	y	
Type of Vehicles	All Wheel Drive	y	y
Type of Vehicles	AWD	y	
Type of Vehicles	Classic	y	
Type of Vehicles	Vintage	y	
Type of Vehicles	Antique	y	
Type of Vehicles	Commercial Vehicles	y	y
Type of Vehicles	Commercial Trucks	y	y
Type of Vehicles	Commercial Vans	y	y
Type of Vehicles	Fleets	y	
Type of Vehicles	Convertibles	y	y
Type of Vehicles	Coupes	y	y
Type of Vehicles	Diesel	y	
Type of Vehicles	Domestic	y	

Colum A:
Target/retrieval taxonomy (to)

Column B:
Source terms from search log (from)

Column C:
Auto-suggested

Column D:
Human review approves as “y” - yes

Mapping Examples

Computer Hardware & Software	N	Computer Hardware	4
Computer Hardware & Software	N	Computer Software	4
Consumer Electronics & Appliances Stores	Y	Consumer Electronics	4
Electrical & Electronic Manufacturing	Y	Electrical/Electronic Manufacturing	4
Health Care Services & Hospitals	Y	Hospital & Health Care	4
Investment Banking & Asset Management	Y	Investment Banking	4
Investment Banking & Asset Management	N	Investment Management	4
Sporting Goods Stores	Y	Sporting Goods	4
Automotive Parts & Accessories Stories	Y	Automotive	5
Biotech & Pharmaceuticals	N	Pharmaceuticals	5
Cable	N	Internet	5
Casual Restaurants	Y	Restaurants	5
Financial Analytics & Research	N	Research	5

Standards for Mapping

SKOS (Simple Knowledge Organization System)

Has a set of relation type properties for mapping:

- mappingRelation – the parent category relation-type property that includes the others:
  - exactMatch – exact match, bidirectional, in all circumstances
  - closeMatch – close match, bidirectional, in some (sufficient) circumstances or in a certain context
  - broadMatch – has broader concept in the other KOS; inverse of narrowMatch
  - narrowMatch – has narrower concept in the other KOS; inverse of broadMatch
  - relatedMatch – has related concept in the other KOS, bidirectional

➢ For directional mapping from a tagged KOS to a retrieval KOS, could use the generic mappingRelation or a combination of exactMatch and closeMatch.

Standards for Mapping

ISO 25964-2 Information and Documentation – Thesauri and interoperability with other Vocabularies
Part 2: Interoperability with other vocabularies (2013)

- Inter-vocabulary mapping is the principal focus.
- Addresses the theory and method of various kinds of mappings.
- Addresses both one-way directional mapping, and multi-directional.
- Considers also mapping between thesauri and other kinds of vocabularies: synonym rings, classification schemes, subject heading schemes, taxonomies, terminologies, name authority lists, and ontologies.

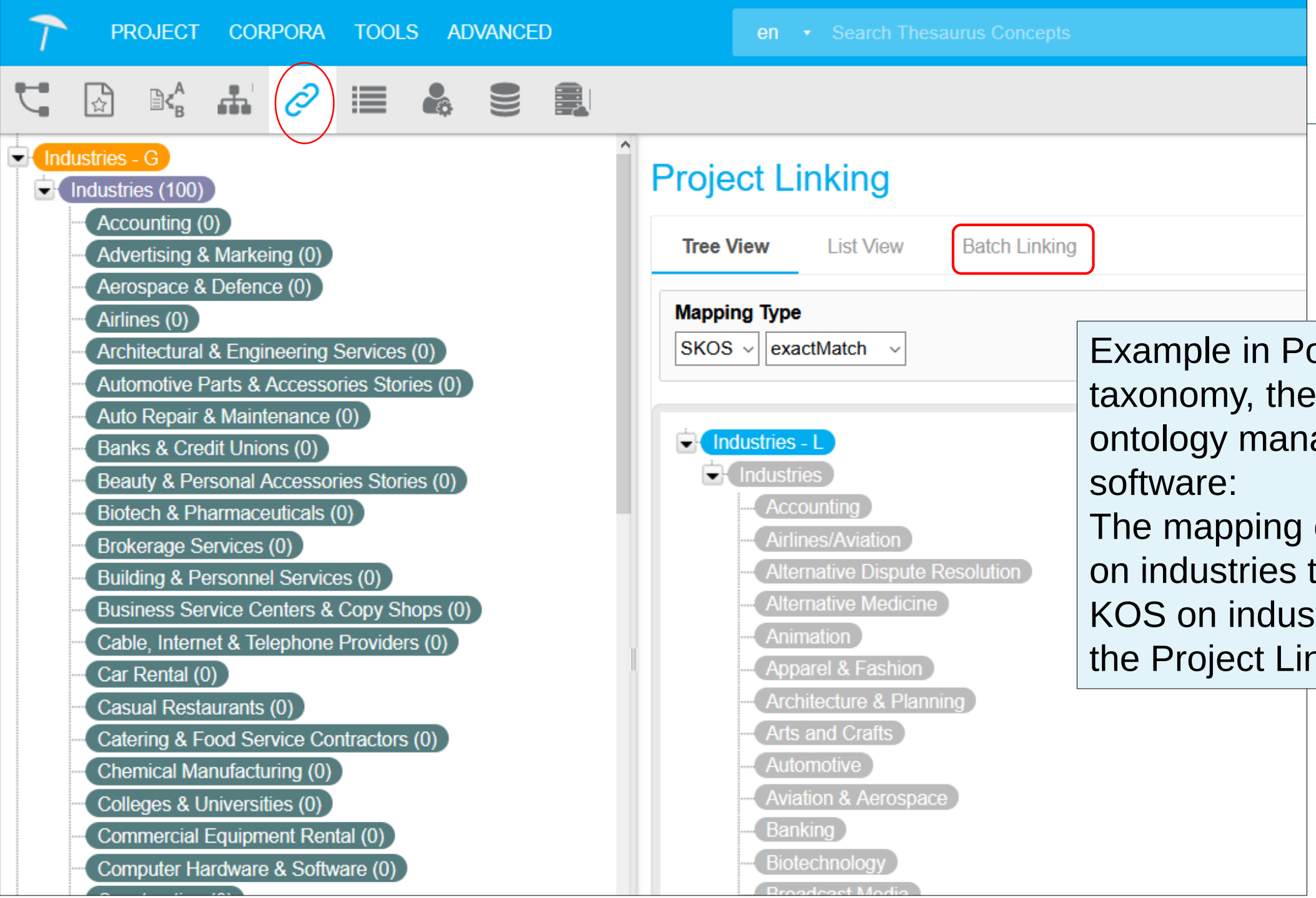
Tools for Mapping

Scripting languages (e.g. Perl), or advanced features of Excel

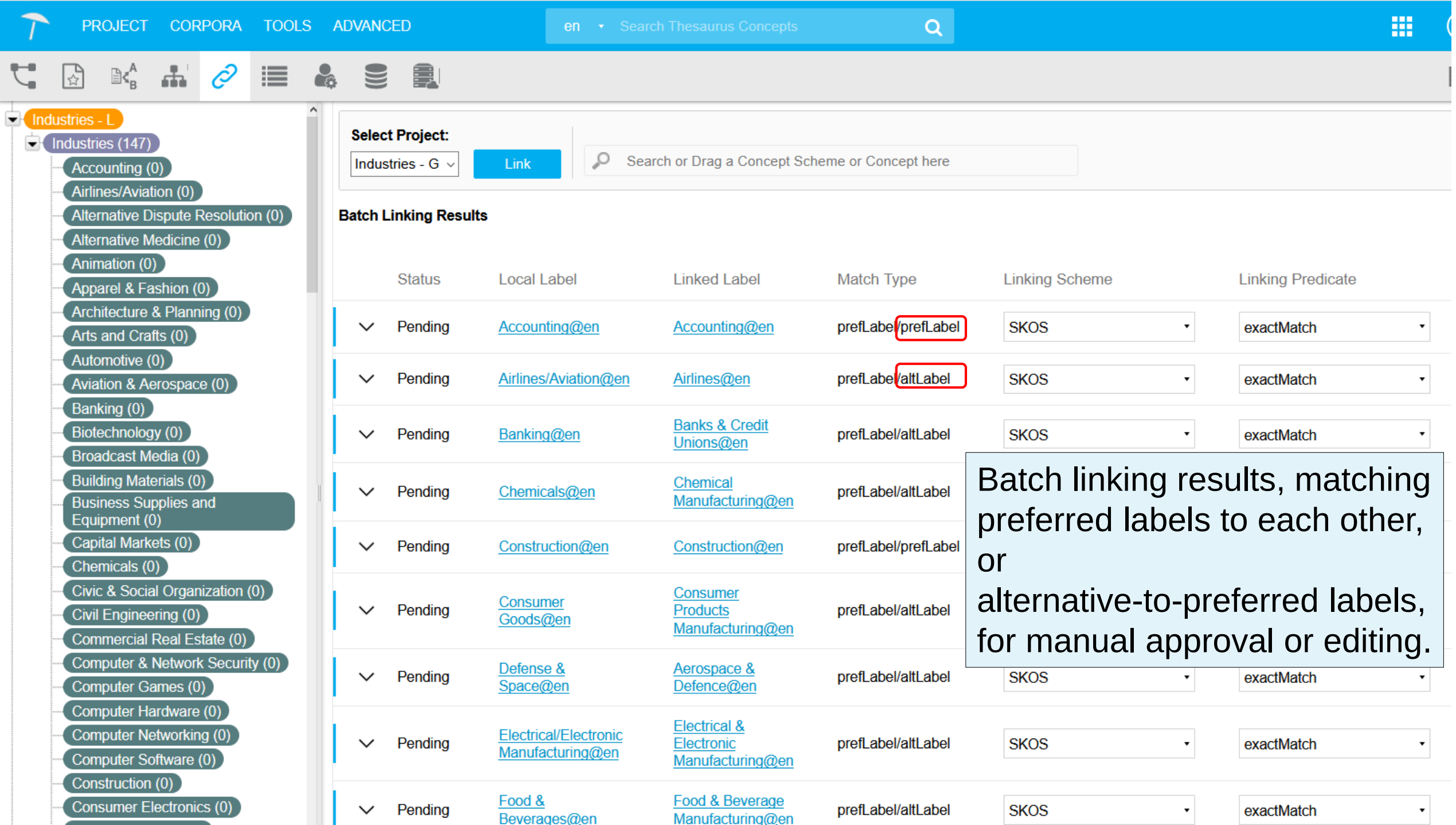
- Used if KOS management software does not have batch/auto-mapping or to enhance software mapping with additional, less-close matches

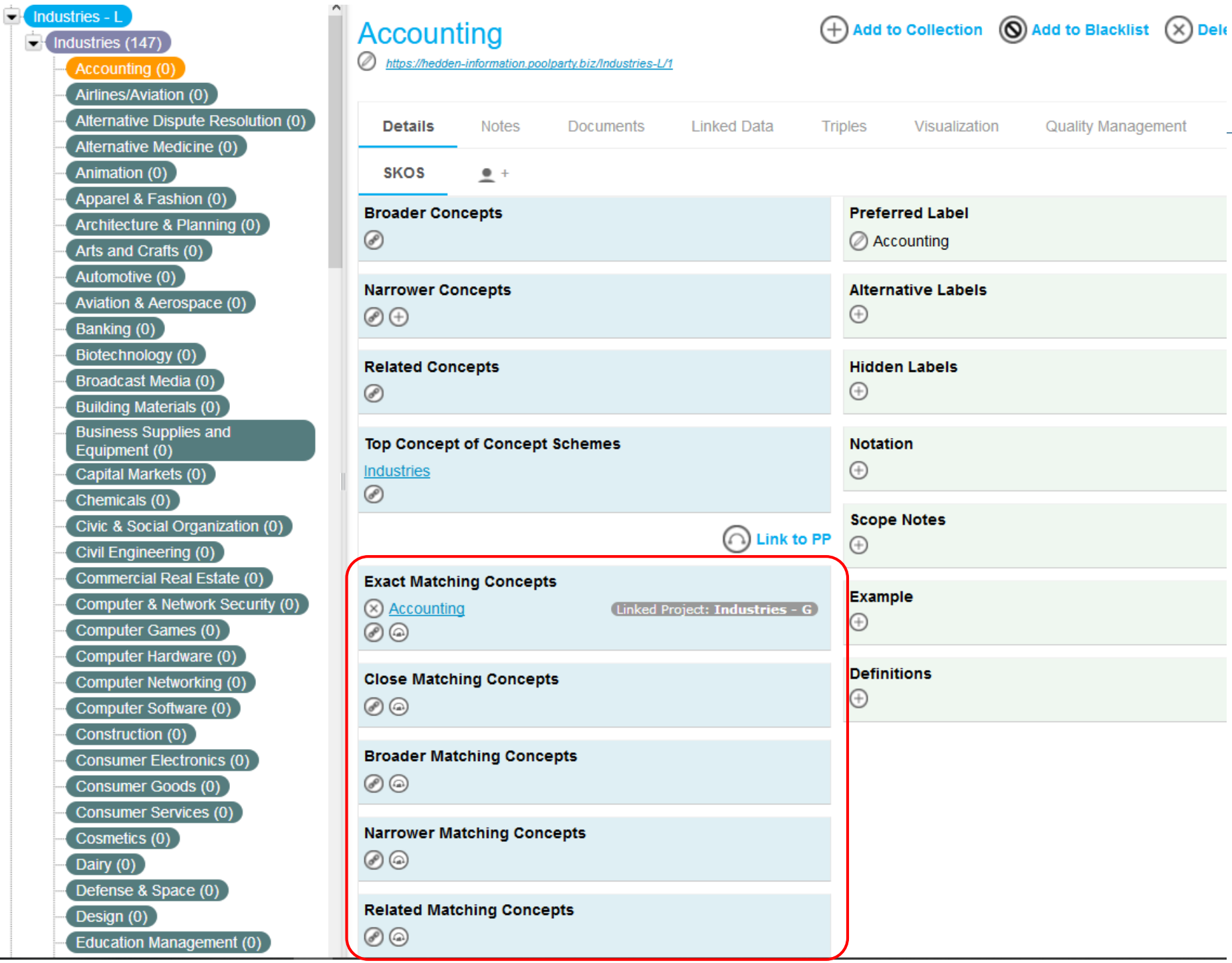
KOS management software feature (PoolParty, Synaptica, Semaphore)

- SKOS-based KOS management software supports mapping relationships between concepts in different vocabularies\
- KOS management software may also have batch/auto-mapping feature for exact and close matches.
- Maintaining mapping relations in a KOS management software supports ongoing maintenance, in case changes occur with concepts.



Example in PoolParty taxonomy, thesaurus, and ontology management software:
The mapping of one KOS on industries to another KOS on industries, using the Project Linking feature.


PROJECT
CORPORA
TOOLS
ADVANCED
en
Search Thesaurus Concepts
Industries - L
Industries (147)
Accounting (0)
Airlines/Aviation (0)
Alternative Dispute Resolution (0)
Alternative Medicine (0)
Animation (0)
Apparel & Fashion (0)
Architecture & Planning (0)
Arts and Crafts (0)
Automotive (0)
Aviation & Aerospace (0)
Banking (0)
Biotechnology (0)
Broadcast Media (0)
Building Materials (0)
Business Supplies and Equipment (0)
Capital Markets (0)
Chemicals (0)
Civic & Social Organization (0)
Civil Engineering (0)
Commercial Real Estate (0)
Computer & Network Security (0)
Computer Games (0)
Computer Hardware (0)
Computer Networking (0)
Computer Software (0)
Construction (0)
Consumer Electronics (0)
Select Project:
Industries - G
Link
Search or Drag a Concept Scheme or Concept here
Batch Linking Results
Status
Local Label
Linked Label
Match Type
Linking Scheme
Linking Predicate
Pending
Accounting@en
Accounting@en
prefLabel/prefLabel
SKOS
exactMatch
Pending
Airlines/Aviation@en
Airlines@en
prefLabel/altLabel
SKOS
exactMatch
Pending
Banking@en
Banks & Credit Unions@en
prefLabel/altLabel
SKOS
exactMatch
Pending
Chemicals@en
Chemical Manufacturing@en
prefLabel/altLabel
Pending
Construction@en
Construction@en
prefLabel/prefLabel
Pending
Consumer Goods@en
Consumer Products Manufacturing@en
prefLabel/altLabel
Pending
Defense & Space@en
Aerospace & Defence@en
prefLabel/altLabel
SKOS
exactMatch
Pending
Electrical/Electronic Manufacturing@en
Electrical & Electronic Manufacturing@en
prefLabel/altLabel
SKOS
exactMatch
Pending
Food & Beverages@en
Food & Beverage Manufacturing@en
prefLabel/altLabel
SKOS
exactMatch
Batch linking results, matching preferred labels to each other, or alternative-to-preferred labels, for manual approval or editing.



Concept details
Advanced SKOS view displays the various SKOS mapping types.

Mapping Case Study

Regulatory information database vendor Wolters Kluwer Financial Services wanted to map its new Regulatory Change taxonomy to the internal taxonomy of a leading bank client of theirs, so that the client could retrieve both its internal content and the subscribed regulatory change content with a single taxonomy.

Mapping Case Study

Issue: Initial mapping was done before the new Wolters Kluwer regulatory change taxonomy was completed, since it was desired to have mapping also serve to enrich the taxonomy with new terms.

Problems:

- Concepts and their labels were not yet finalized in the retrieval taxonomy, so mapping would be postponed or might have to be redone.
- A change in a label is OK, but a change in the meaning of a concept impacts mapping.

Solution:

- Using a KOS management software tool that automates mappings saves time in doing mappings, so doing mapping twice at different stages in the project is OK.

Mapping Case Study

Issue: It was desired to have the mapping go in both directions.

Problems:

- Only exact matches would work in both directions, but many mappings are not exact, but slightly narrower-to-broader.
- Mappings could be done twice, once in each direction, but that's more work.

Solutions:

- Using SKOS designated broadMatch and narrowMatch, in addition to exactMatch, preserves narrower-to-broader distinctions, and the mappings function in both directions.
- Using a KOS management software tool that automates mappings of exact matches and close matches saves time in doing mappings, so a mapping in the other direction can also be done to check quality of initial mapping.

Mapping Case Study

Issue: A very low number of automated matches were initially achieved.

Problems:

- Scope of taxonomies do not match.
- Terms that could be mapped were not because they were not similar enough.
- Synonyms/alternative labels were very few in one taxonomy and not complete in the other.

Solution:

- Adding more alternative labels to concepts in both vocabularies support automated matching, and the automated matching is run again.

Examples that did not automatically match, but should have:

Commercial Accounts <-> Business Deposit Accounts
(The latter had more specific types only, as examples, for alternative labels.)

Beyond Equivalency Mapping: Other KOS Linking

Other cross-taxonomy relationships with other functions

- Relationships across taxonomies, that are “related term” types of relationships, not equivalence type
- No automatic way to create them, done term-by-term
- Could use SKOS relatedMatch relationship
- When a user selects a concept, it does *not* retrieve content tagged to both concepts in both taxonomies.
- Relationships (directly or indirectly) must display to the end user.
- Relationships can be generic “related term” or customized/semantic.
- Example: Products Taxonomy concepts related to Interests Taxonomy concepts

Questions/Contact

Heather Hedden

Taxonomy Consultant
Hedden Information Management
Carlisle, MA USA
+1 978-467-5195
www.hedden-information.com
accidental-taxonomist.blogspot.com
www.linkedin.com/in/hedden
Twitter: @hhedden