

Indexing and subject organisation

Madely du Preez

Dept of Information Science

University of South Africa (UNISA)

LIASA IGBIS WORKSHOP 2018: 16-18 August, Centurion Lake Hotel.



Menu

- ▶ Subject organisation
- ▶ Indexing
 - ▶ Why index
 - ▶ Indexing languages
- ▶ Thesauri
- ▶ Taxonomies
- ▶ Folksonomies
- ▶ Subject cataloguers as indexers
- ▶ Indexing training



Information organisation and subject organisation

- Information organisation
 - Description of an entity
 - Preparing the representation of the entity
- Subject organisation
 - Presenting information to ensure easy access
 - Helpful arrangement:
 - Arrangement is based on user's needs and interest
 - Accurate specification
 - Based on the nature of knowledge



Subject organisation

- ▶ Objective
 - ▶ Provide access to the content of an entity
 - ▶ Support information retrieval
 - ▶ Grouping of entities
- ▶ Information systems
 - ▶ Sets requirements for subject organisation:
- ▶ Notational
- ▶ Verbal



Indexing



- **Index**: a **systematic guide** to the **contents** of a document
- **Indexing terms** are terms that **describe** the **intellectual contents** of a document. They are also referred to as **key terms, subject terms, descriptor terms** or **keywords**
- **Indexing** is the **process** of **creating** an index to the contents of a document or other entity



Why index?

- Indicate importance aspects
- Obtain access through **terminology of a document**
- Obtain access through terminology a **user might prefer**



Indexing languages

- **Natural language**

- The language of the user or the language of the author

- **Controlled vocabularies**

- controlled vocabularies **contain indexing terms** used for the description of an entity's contents (Mamassion 2010)
- “controlled and dynamic vocabulary of **terms** that **share semantic** and **generic relationships**, and that are **applied** in a particular **field of knowledge**” (Currás 2010, p. 72)
- A **predetermined list of words** that can be used to organise as well as retrieve information



Natural language

- ▶ When do we use Natural language?
 - ▶ Books: Print and e-books
 - ▶ Websites
 - ▶ Indexing databases
- ▶ Problems associated with natural language:
 - ▶ Synonyms
 - ▶ Homonyms



Controlled vocabularies

- **Thesauri:** controlled vocabularies that contain index terms that are used to describe the contents of a document
- **Subject headings:** systematic list of terms that describe a given subject matter
- **Taxonomies:** The science of classifying things; a classification into ordered categories
- **Folksonomies:** A classification system derived from user-generated electronic tags or keywords that annotate and describe online content. Also known as social-tagging

Thesauri

- ▶ Emerge from the use of **concepts taken from existing documents** in a collection
 - ▶ Printed documents; photo collections; art collections; legal documents, etc.
- ▶ Used to **categorise information** sources according to **subject** – narrow searches
 - ▶ Keyword search for “Ebola virus” + Thesaurus term “Economy” will not retrieve medical related information on the virus.
- ▶ Conditions a thesaurus must fulfil:
 - ▶ **Specialised language**
 - ▶ Conversion of thesaurus terms into keywords
 - ▶ Include **hierarchical and non-hierarchical relationships** among terms
 - ▶ Terminology
 - ▶ Allow for the updating of the thesaurus



Taxonomies

- ▶ Knowledge classification systems
- ▶ Their structures resemble “trees”
- ▶ Allows for the **graphic representation** of data, e.g. the organisation of concepts within a field of expertise
- ▶ Construction involves: identification, comparison & grouping of elements
- ▶ **Supports information retrieval** from the Internet through synonym rings or browsable taxonomies



Folksonomies

- ▶ **Classification scheme** built by a broad **community of users**.
- ▶ **Other terms**: collaborative tagging; social classification; social indexing; social tagging
- ▶ Are part of **social software applications** such as social bookmarking (e.g. annotate photographs)
- ▶ Some websites include “**tag clouds**” – represents text data to depict **keyword metadata** (tags) on websites
- ▶ Folksonomies evolve when **users create** or store **content** on a particular site
- ▶ “**Like**” postings and tagging content create **tag clouds**
- ▶ **Tag clouds** pinpoint **identifiers** and the **frequency of use** on a folksonomy site.

Characteristics

Thesauri	Subject headings	Taxonomies	Folksonomies
<p>Alphabetical Term relationships:</p> <ul style="list-style-type: none"> • Hierarchical • Equal <p>Includes references</p> <p>Restricted list</p> <p>Collection specific:</p> <ul style="list-style-type: none"> • General (ISAP) • Subject (ERIC) • Specialised (INIS) 	<p>Alphabetical Term relationships</p> <ul style="list-style-type: none"> • Hierarchical • Equal <p>Includes references</p> <p>May add terms</p> <p>Library catalogues</p>	<p>Classification system based on defined classes</p> <p>Hierarchical</p> <p>Term relationships:</p> <ul style="list-style-type: none"> • Hierarchical <p>May add terms</p> <p>Various subject fields:</p> <ul style="list-style-type: none"> • Natural science • Education • Safety 	<p>User defined metadata collections</p> <p>Classification scheme</p> <p>Automatic</p> <p>No hierarchical relationships</p> <ul style="list-style-type: none"> • Broad: Multiple users tag terms • Narrow: content creator tags using limited number of terms • Reflects users' vocabulary • Subject to manipulation

Controlled vocabulary presentation of indexing terms

Thesauri	Subject headings	Taxonomies	Folksonomies
<p>Indexing terms: Bread; Fat cakes</p> <p>Thesaurus entry: BREAD BT Staple food RT Fat cakes</p> <p>FAT CAKES RT Bread</p>	<p>Catalogue entry 1. Citrus fruits – Limpopo Province</p> <p>SH List entry Citrus fruits (may subdiv geog.) UF Citrus BT Fruit NT Lemons Limes Oranges</p>	<p>Anseriformes</p> <p>Organisms: Ducks, Geese, Swans</p> <p>Information: Kingdom: Animalia Phylum: Chordata Subphylum: Vertebrata Class: Aves Order: Anseriformes</p>	<p>Elements of Chemistry / A L Lavoisier</p> <p>Librarything.com tags:</p> <p>Science (8) Chemistry (4) Nonfiction (4) Great books (2) Attic-V (1) Britannica Great Books (1) Chemical Analysis (1)</p>

Compilation of controlled vocabularies

Thesauri	Subject headings	Taxonomies	Folksonomies
Individuals, teams Indexing terms Used for specific collections Can use software	Editorial team	Automated •Web mining •Website headers •Hyperlinks •Key names, concepts data •Categories (Topics) & User tags •Applies machine learning techniques	Automated Authors &/ users •Social tagging •Collaborative tagging •Cooperative classification •Social indexing •Crowd-sourced (meta) data •Creates grass roots, bottom up classification structure

Used for

Thesauri	Subject headings	Taxonomies	Folksonomies
<p>Control formation of indexing terms Assists</p> <ul style="list-style-type: none">• Indexers• Users <p>Combined use with natural language Information searches are more specific Narrows information searches</p>	<p>Consistent description of entities Supports information retrieval</p>	<p>All users use taxonomies – without knowing Search engines:</p> <ul style="list-style-type: none">• Improve relevance <p>Blogs: grouping of posts Grouping of websites</p>	<p>Knowledge acquisition Finding & re-finding content in a group (e.g. FB) Discover info & people using creator tags Marketing:</p> <ul style="list-style-type: none">• Amazon.com (requests to evaluate books)



What does indexing involve?

- ▶ **Conceptual analysis** – identifying the aboutness of the entity
- ▶ Identification of **indexable concepts**
- ▶ **Translation of concepts** into indexing terms
- ▶ Consider **users' needs**
- ▶ Consider the **systems' requirements**
 - ▶ **Specificity of indexing**, i.e. the use of descriptive terms for a series of related terms versus the specific term (e.g. Citrus fruit versus oranges, lemons, pomelos, etc.)
 - ▶ **Exhaustivity of indexing**, i.e. the extent to which contents is reflected



Subject cataloguers as indexers

- ▶ **Subject cataloguers are indexers** because
 - ▶ They are focused on identifying concepts
 - ▶ Organising and describing information in a structured manner
 - ▶ Know how to use controlled vocabularies when describing information
- ▶ **Natural language sets different requirements to subject organisation**
 - ▶ Controlled vocabularies are not used. The author's terminology is used
 - ▶ More indexing terms are assigned than when assigning subject headings
 - ▶ Consider aboutness and avoid over indexing
 - ▶ Different skills are required to do book indexing, journal indexing for a database, and web indexing
- ▶ **New career opportunities are available – including freelance work**



Learning how to index?

- ▶ Library and information science schools teach database indexing as part of their curricula
- ▶ The Association of Southern African Indexers and Bibliographers (ASAIB)
 - ▶ Indexing training workshops for: basic indexing principles, book indexing, journal indexing, e-pub indexing, thesaurus construction
 - ▶ Customised indexing training for the indexing of specialised types of documents, e.g. legal documents, standards, etc.
 - ▶ All workshops comply with international indexing standards
- ▶ Society of Indexers
 - ▶ Online indexing courses



Learn more about ASAIB

- ASAIB web address: www.asaib.org.za
- Provides information on workshops and conferences
- A directory for freelance indexers



Resources



- ▶ Currás 2010
- ▶ Folksonomy. 2018. <https://en.wikipedia.org/wiki/Folksonomy> (accessed 9 August 2018).
- ▶ Hedden, H. 2010. Controlled vocabularies, thesauri, and taxonomies. In J. Perlman & EL Zafran (eds.), *Index it right: advice from the experts* (pp. 135-154). Medford, NH: Information today.
- ▶ Hedden, H. 2014. Tagging versus indexing. *The Indexer* 32(2): 81-82
- ▶ Mamassion, L. 2010. Through the looking glass: a freelance perspective on database indexing. In J. Perlman & EL Zafran (eds.), *Index it right: advice from the experts* (pp. 135-154). Medford, NH: Information today
- ▶ Macgregor, G. & McCulloch, E. 2006. Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review* 55(5): 291-300.
- ▶ Taxonomy. 2018. <https://www.wpbeginner.com/glossary/taxonomy/> (Accessed 9 August 2018).
- ▶ Thesaurus. 2018. [https://en.wikipedia.org/wiki/Thesaurus_\(information_retrieval\)](https://en.wikipedia.org/wiki/Thesaurus_(information_retrieval)) (Accessed 9 August 2018).