

Using taxonomies for knowledge exploration in subject gateways

Marcin Roszkowski

National Library of Poland

Presentation plan

- Introduction to subject gateways and taxonomies
- Study overview
- Results and discussion

F.W. Lancaster (1933 -)

- “[...] it seems clear that the continued growth of network-accessible information resources will make subject analysis activities of greater importance than ever before”
 - Do Indexing and Abstracting Have a Future? *Annales de Documentacion*, (6), 137-144.



Subject gateways /overview/

- The potential of manual approach to cataloging networked resources.
- Services that provide access to Internet resources that have been reviewed, selected and described by subject specialists.
- Features:
 - Selection and quality criteria,
 - Rules for manual creation of descriptive metadata
 - Rules and tools for subject indexing

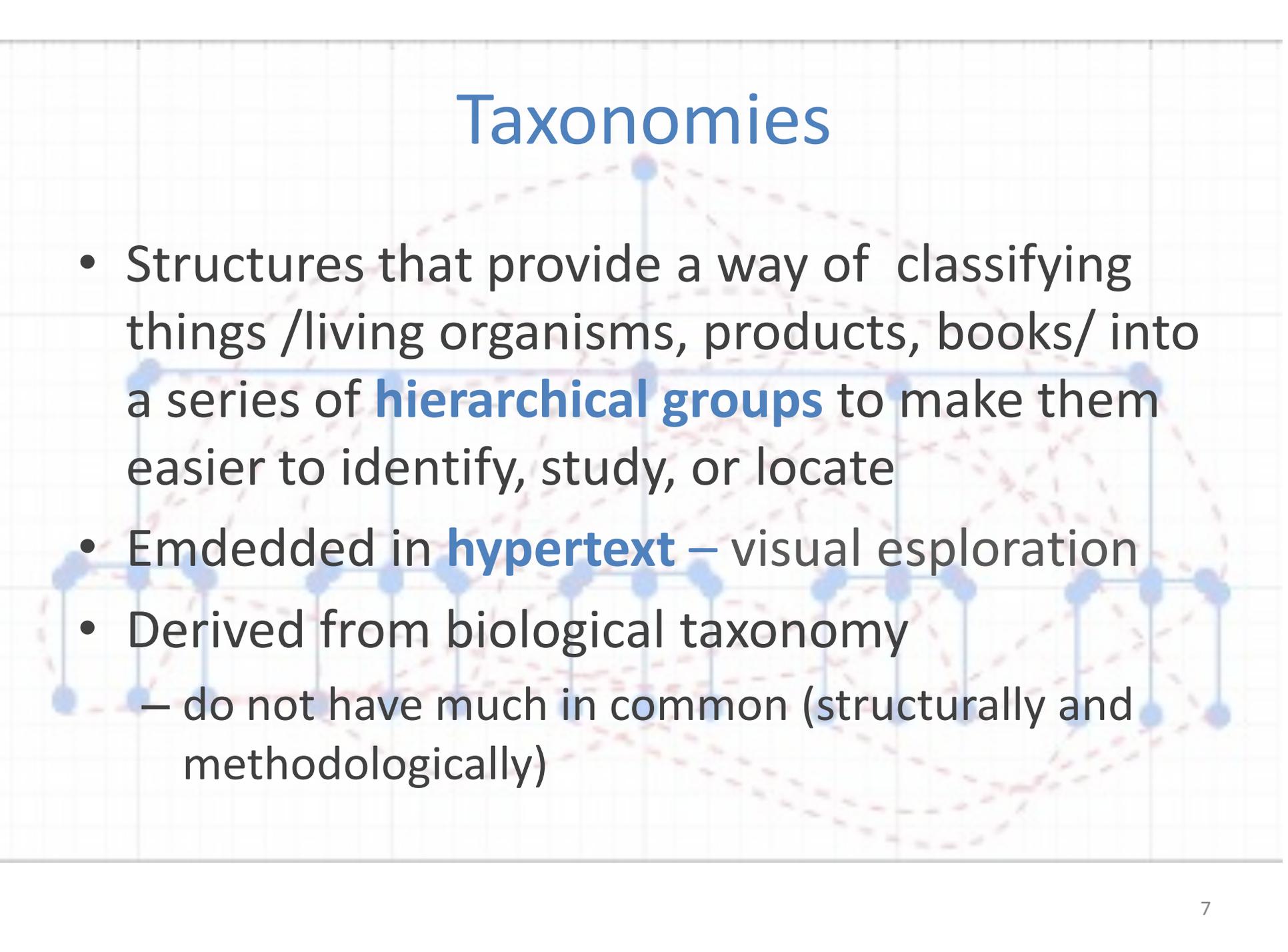
Subject gateways /overview/

- Features:
 - Rich metadata = offer more sophisticated search options than other Web indexes
 - Access to information
 - Focalized search – search engines
 - Exploratory search – browsing tools
- Output:
 - Metadata about high quality web resources from specific domain

Subject gateways

- Domain specific:
 - MathGuide – Mathematics,
 - GeoGuide - Geology,
 - The Gateway to 21st Century Skills - Education
- Multidomain:
 - Intute,
 - Infomine,
 - BUBL LINK Catalogue of Internet Resources (no longer being updated)

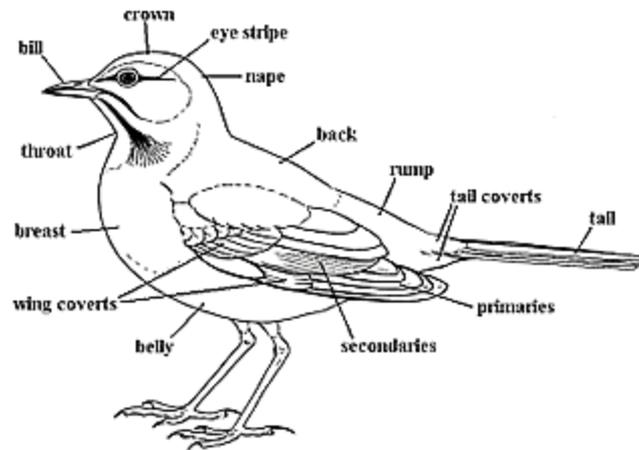
Taxonomies



- Structures that provide a way of classifying things /living organisms, products, books/ into a series of **hierarchical groups** to make them easier to identify, study, or locate
- Embedded in **hypertext** – visual exploration
- Derived from biological taxonomy
 - do not have much in common (structurally and methodologically)

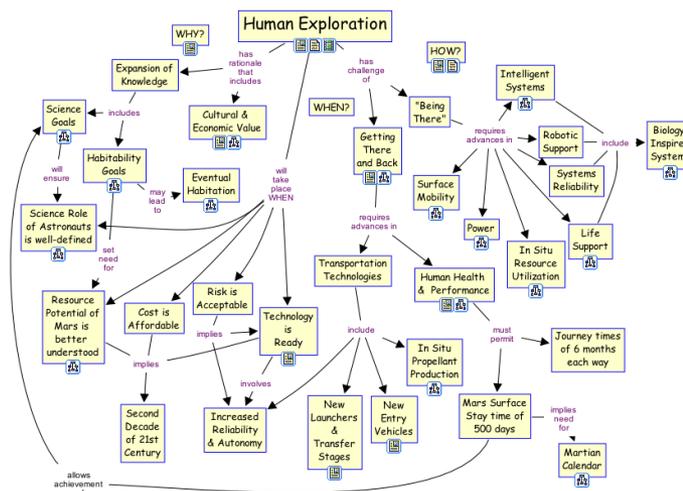
Taxonomies /function/

Identification - can help control the glut of information and identify where information should be stored by filtering, categorizing, and labeling information.



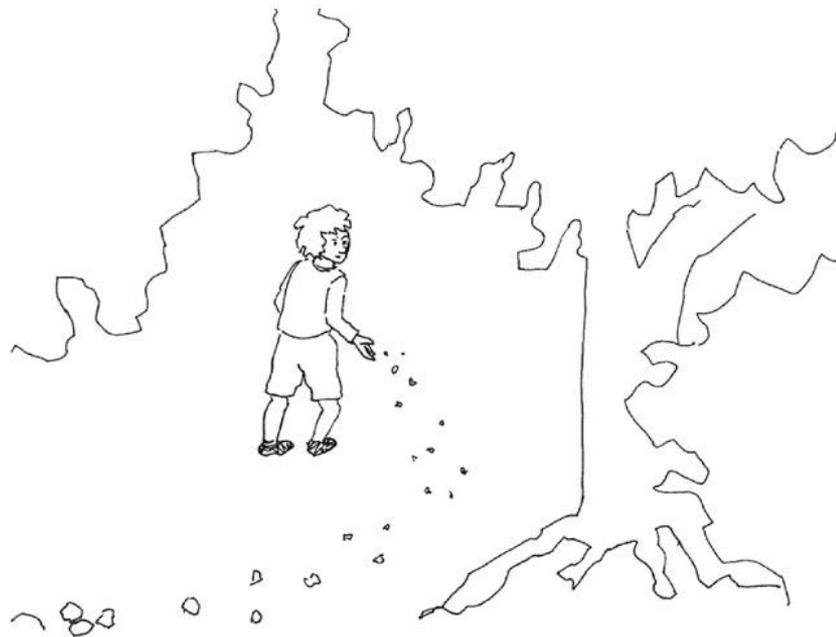
Taxonomies /function/

Discovery - additional information on a topic can be inferred by seeing where the entry is placed in context within the taxonomy and provide serendipitous guidance to the person working on the issue.



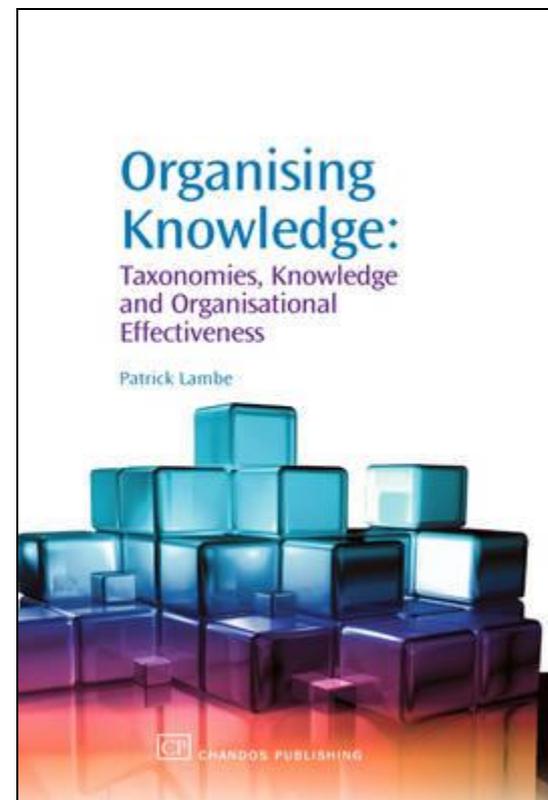
Taxonomies /function/

Delivery - can improve the retrieval process. The use of navigation paths or "breadcrumbs" based on the taxonomy's hierarchy provide context and enhance searching via free text.



Taxonomies /form/

- Taxonomies can take many forms:
 - lists,
 - hierarchies,
 - polyhierarchies,
 - matrices,
 - facets.



Study objectives

- Analysis of taxonomies in subject gateways.
- An attempt to investigate models of taxonomies.
 - Identification of the dependencies and patterns in taxonomies.

Study

- Taxonomies (browsing structures) from 20 subject gateways
- Quantitative research
 - depth of the hierarchies,
 - categories per level distribution,
 - resources per level distribution,
 - number of categories on the first level of division.
- Qualitative research
 - ways of taxonomy exploration, principles of division, taxonomy structure.

Results

- Development of structural models (3) of taxonomies in subject gateways
 1. „Flat” model (non-hierarchical)
 2. Static hierarchical model
 3. Dynamic hierarchical model

„Flat” model

- Categories in alphabetic order
- No semantic relationships between categories
- Categories represent both subjects and form of resources (often in one listing)
- Form-oriented categories often have some kind of apposition
 - [resource type]

Subject links	A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
About us	Abbreviations [2]
A-Z of services	Abbreviations [Publication Type] [2]
Internet catalogue	Abdomen [9]
Advanced search	Abdominal Injuries [2]
Browse	Abdominal Muscles / anatomy & histology [1]
New resources	Abdominal Muscles / physiology [1]
Suggest a site	Abdominal Neoplasms / case studies [1]
Internet training	Abdominal Pain [2]
Virtual Training Suite	Abdominal Pain / case studies [1]
Support materials	Abnormalities [11]
Intute events	Abnormalities, Multiple [1]
Additional services	ABO Blood-Group System [1]
Hot topics	Abortion, Habitual [1]
Timelines	Abortion, Habitual / diagnosis [1]
MedHist events	Abortion, Habitual / immunology [1]
Psci-com events	Abortion, Induced [6]
Psci-com books and articles	Abortion, Induced / prevention and control [1]
	Abortion, Legal [4]
	Abortion, Spontaneous [3]
	Abortion, Threatened [1]
	Abruptio Placentae [1]
	Abscess / radiology [1]
	Academies and Institutes [222]
	Access to Information [2]
	Accident Prevention [19]

Categories and appositions in the Intute: health & life sciences flat taxonomy

Browse Subjects for 'A'

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#) [0-9](#)

“AB - AG”

[AB - AG](#) · [AG - AQ](#) · [AQ - AZ](#)

 [Expert records](#)

 [Add Robot-selected Resources?](#)

subjects under 'A'

- [ABILITY -- SOCIAL ASPECTS -- UNITED STATES \(1\)](#)
- [ABORIGINAL AUSTRALIANS \(1\)](#)
- [ABSENTEEISM \(LABOR\) -- NORTH CAROLINA -- ECONOMETRIC MODELS \(1\)](#)
- [ACADEMIC ACHIEVEMENT -- ECONOMETRIC MODELS \(5\)](#)
- [ACADEMIC ACHIEVEMENT -- ECONOMIC ASPECTS -- UNITED STATES \(1\)](#)
- [ACADEMIC ACHIEVEMENT -- ISRAEL -- ECONOMETRIC MODELS \(1\)](#)
- [ACADEMIC ACHIEVEMENT -- MATHEMATICAL MODELS \(1\)](#)
- [ACADEMIC ACHIEVEMENT -- NORTH CAROLINA \(1\)](#)
- [ACADEMIC ACHIEVEMENT -- NORTH CAROLINA -- ECONOMETRIC MODELS \(1\)](#)
- [ACADEMIC ACHIEVEMENT -- SOCIAL ASPECTS -- UNITED STATES \(1\)](#)

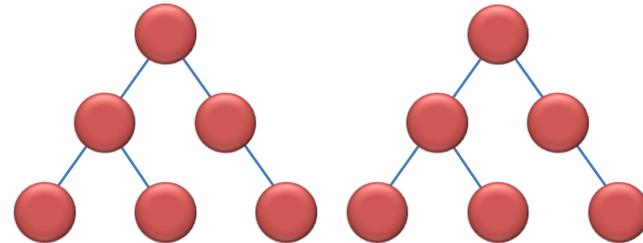
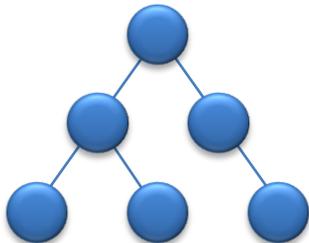
Compound subject headings (LCSH) as names for categories in Infomine gateway.

Hierarchical models

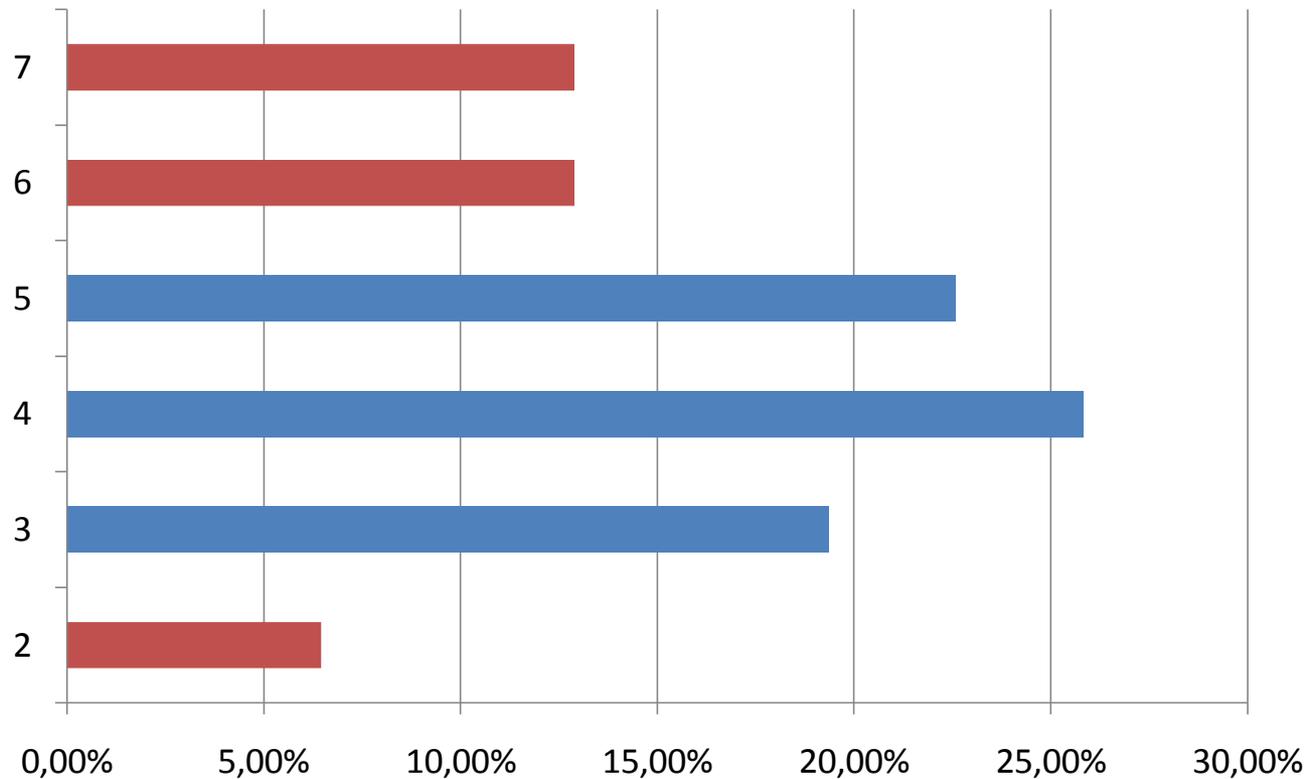
- Developed on the basis of the scope of the hierarchical relationships as a way for expressing the taxonomy structure
 - Static model
 - Dynamic model

Static model

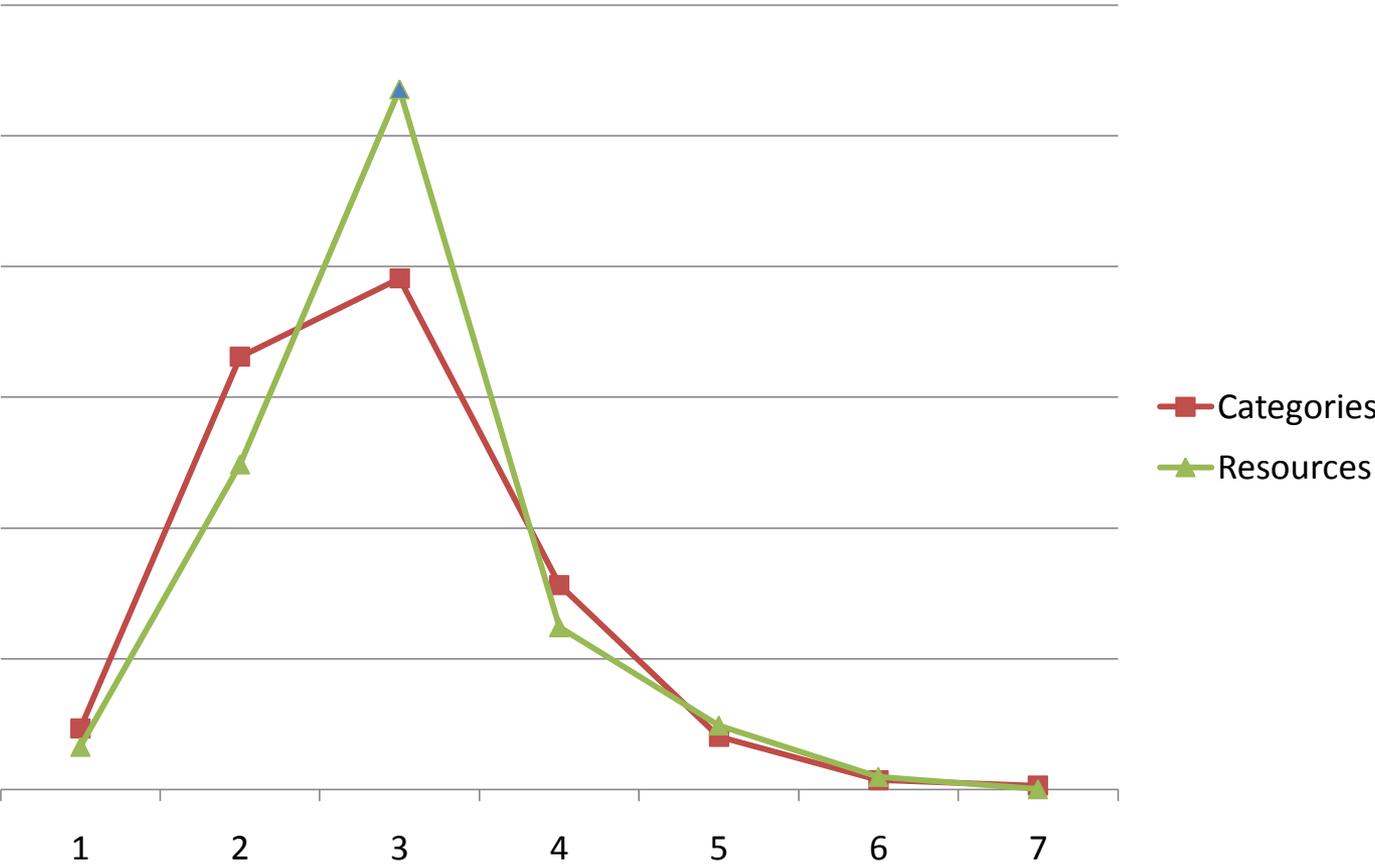
- Mono- and polyhierarchical structures
- The use of **broader/narrower** semantic relationship for organization of categories
- No logical division (no mutually exclusive and exhaustive)
 - no „is-a” relation



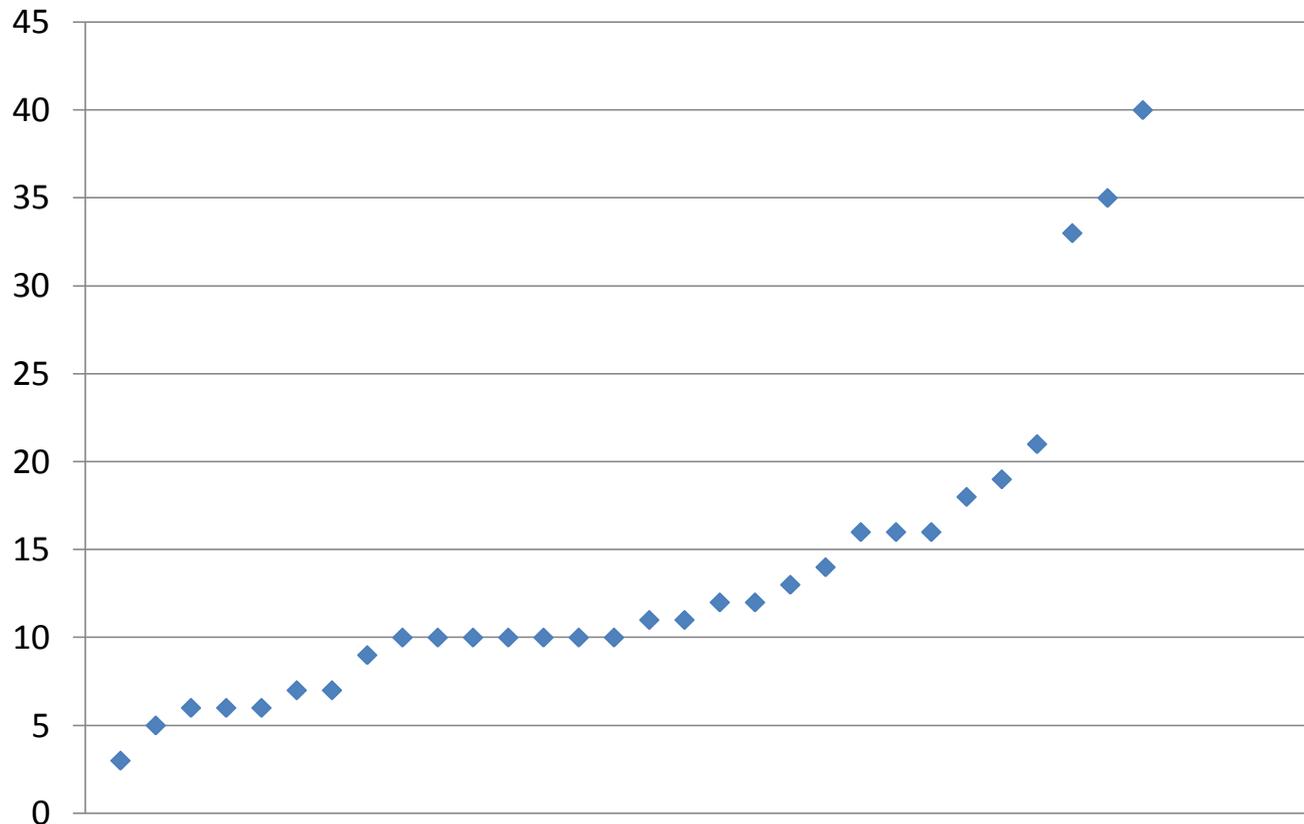
Depth of the hierarchies



Correlation between resource and category per level distribution



First Cut - number of categories on the first level of division



Prof Babara Kwasnik: „this determines the shape and eventually the representational eloquence of the hierarchical structure

Static model /construction/

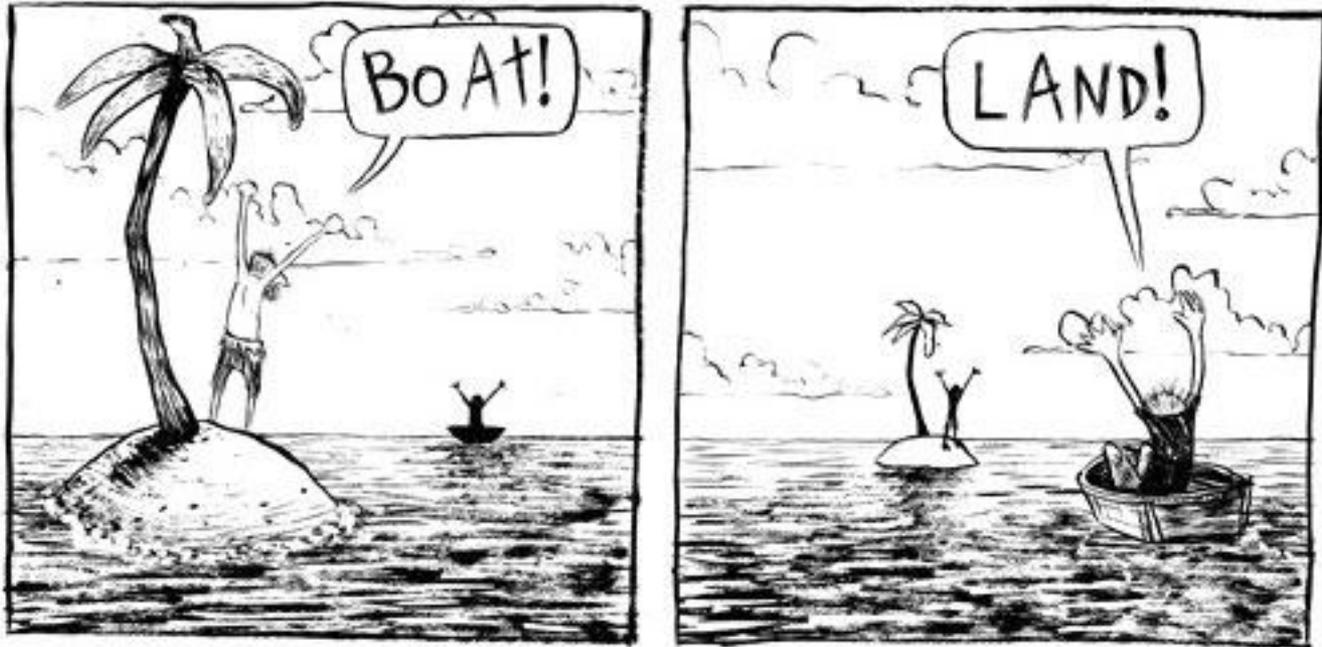
- Literary warrant
 - introduced by E. Wyndham Hulme (1911) as a means of class determination,
 - the principle prescribes that concept/category usage should be empirically derived from collection of documents containing the concepts to be named
 - no empty category (without a posting)
 - not applied consistently

Static model /construction/

- “dummy categories”
 - used as empty intermediaries in specific branch expansion
 - allow for creation of consistent structure
 - show to the user that specific topic/concept is included in the scope of taxonomy but currently there is no posting to collection
 - **Category_A**
 - **Dummy Category**
 - **Category_B**

Static model /construction/

- User and viewpoint warrant

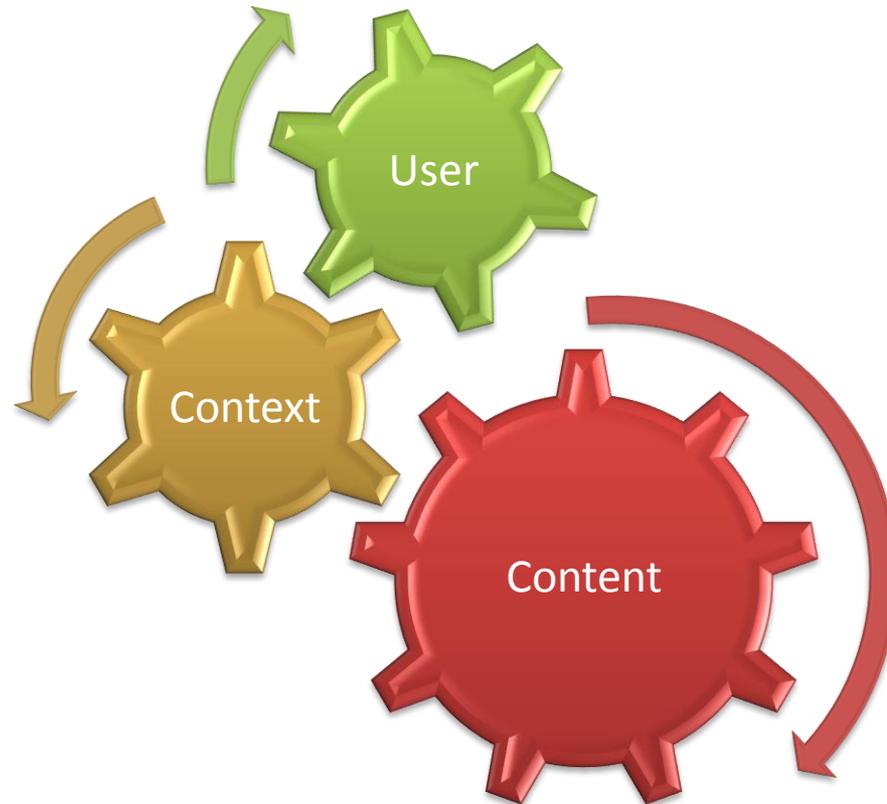


Problems of inconsistency in taxonomies

Animals:

- A. animals that belong to the Emperor
- B. embalmed animals
- C. animals that are trained
- D. suckling pigs
- E. mermaids
- F. fabulous animals
- G. stray dogs
- H. animals included in the present classification
- I. animals that tremble as if they were mad
- J. innumerable animals
- K. animals drawn with a very fine camelhair brush
- L. other animals
- M. animals that have just broken a flower vase
- N. animals that from a long way off look like flies

Static model /construction/



Dynamic model

- Faceted approach to knowledge organization
- Analytic-synthetic methods for concept categorization and categories organization
- Facets correspond to different views on organized set of concepts
- Each facet contains categories representing one specific dimension

The case of



- Coverage : web resources for education in the USA
- Taxonomy:
 - a six dimensional categorization of concepts represented by categories
- Facets:
 - Subject, Resource Type, Educational Level, Mediator, Beneficiary, Price Code

Gateway to 21st Century Skills



- start
- learning & teaching
- leading & managing
- partnering
- asn
- about
- help

zaloguj dołącz szukaj

jesteś w: start

nawigacja

- Start
- Learning & Teaching
- Leading & Managing
- Partnering
- ASN
- About
- Help
- portal_seamarksearch
- browse



Great Public Schools
for Every Child



1 to 10 of 3998 items | [Next](#)

[Enlargement of Figures](#)

Description	In this activity you will learn how to make geometrical figures larger.
Medium	text/HTML
Type	Activity
Grade Level	6 , 7 , 8 , 9 , 10 , 11 , 12

[view full record](#)

[Stowaway Adventure: Real Time Data Activities](#)

Description	These six interdisciplinary activities in support of The Stowaway Adventure Project ask students to chart a ship's course, determine ports of call, calculate distance covered and ship speed, determine how weather conditions affect the ship's progress, and more.
Medium	text/HTML
Type	Activity
Grade Level	6 , 7 , 8 , 9 , 10 , 11 , 12 , 5

[view full record](#)

[Money In the Bank](#)

Medium	text/HTML
---------------	-----------

Search the GEM Catalog

find

in

Search in Results
 Start Over

[Go](#) [Help ?](#)

[Browse the Catalog](#)

3998 items matching
subject is 'mathematics'

Refine by Subject

- Top Term
- [arts](#) 232
- [educational technology](#) 320
- [foreign languages](#) 87
- [health](#) 297
- [language arts](#) 538
- [mathematics](#) 3998



Gateway to 21st Century Skills

start learning & teaching le

jesteś w: start

nawigacja

- Start
- Learning & Teaching
- Leading & Managing
- Partnering
- ASN
- About
- Help
- portal_seamarksearch
- browse

Type

- [Activity](#) 1195
- [Artifact](#) 1
- [Best practice](#) 7
- [Catalog record](#) 6
- [Collection](#) 266
- [Curriculum](#) 19
- [Curriculum support](#) 237
- [Data set](#) 6

Grade Level

- [1](#) 804
- [10](#) 1488
- [11](#) 1446
- [12](#) 1439
- [2](#) 852
- [3](#) 1069
- [4](#) 1162
- [5](#) 1428
- [6](#) 1926
- [7](#) 1929
- [8](#) 2019
- [9](#) 1602
- [adult/continuing education](#) 103
- [all](#) 20
- [community college](#) 190
- [higher education](#) 387
- [kindergarten](#) 540
- [preschool education](#) 242
- [unspecified](#) 177
- [vocational education](#) 50

Mediator

- [College/University Instructors](#) 211
- [Elementary school teachers](#) 694
- [General public](#) 138
- [Middle school teachers](#) 788
- [Parents](#) 493
- [Secondary school teachers](#) 518
- [Students](#) 535
- [Teachers](#) 2524
- [Teaching professionals](#) 195

[Stowaway Adventure: Real Time](#)

Description	These six interdisciplinary act... Project ask students to chart... calculate distance covered an... conditions affect the ship's pr...
Medium	text/HTML
Type	Activity
Grade Level	6 , 7 , 8 , 9 , 10 , 11 , 12 , 5

[Money In the Bank](#)

Medium	text/HTML
---------------	-----------



Great Public Schools
for Every Child



Gateway to 21st Century Skills



start learning & teaching leading & managing partnering asn about help

zaloguj dołącz szukaj

jesteś w: start

nawigacja

- Start
- Learning & Teaching
- Leading & Managing
- Partnering
- ASN
- About
- Help
-

1 to 8 of 8 items

[Illuminations Mathematics Standards](#)

Description	Principles and Standards for School Mathematics has been produced by the National Council of Teachers of Mathematics. Principles and Standards for School Mathematics describe what mathematics students should learn and provide guidelines for teaching mathematics, as well as for assessing student progress. Information about the standards are found on this page from Illuminations, the National Council of Teachers of Mathematics website and member of the MarcoPolo consortium.
Medium	image/gif, image/jpeg, text/HTML
Type	Curriculum
Grade Level	6 , 7 , 8 , 9 , 10 , 11 , 12 , kindergarten , 1 , 2 , 3 , 4 , 5
view full record	

[Home Energy Use](#)

Description	Students perform "personal energy audits" to measure the amount of energy that they, and their family, use in their daily lives. The students record the items in each room of their homes that use energy. In the second part of the lab the students calculate the total amount of energy used in their homes on an average day. Discussion then takes place about ways to limit the amount of energy use in homes.
Medium	application/pdf
Type	Curriculum

Search the GEM Catalog

find

in

Search in Results

Start Over

[Help ?](#)

8 items matching
 subject is ' **mathematics** '

type is ' **Curriculum** '

level is ' **6** '

Refine by Subject

- Top Term
- [arts](#) 2
- [educational technology](#) 1
- [foreign languages](#) 1
- [health](#) 1
- [language arts](#) 5
- [mathematics](#) 8



Great Public Schools for Every Child



Conclusions

- Different models of taxonomies in subject gateways
- Use of controlled vocabularies
 - for naming categories (LCSH, many thesauri)
 - adaptation of library classification (eg. DDC, NLM Classification)
- Classification Research Group Manifesto (1955)
 - „The need for a faceted classification as the basis of all methods of information retrieval”

Conclusion

- V. Broughton argues :
 - “facet analysis is significant for the clarity of the light it shines upon the relationships between objects and entities, and abstract concepts and their associated labels. It gives a rational, scientific, methodology for the construction of systems.
- **There is a need for faceted taxonomies as the basis of exploratory access to information**
 - Recommended projects: Dynamic Taxonomies and Faceted Search





Thank You !



Marcin Roszkowski

National Library of Poland

Image sources

- <http://www.wallstory-murals.co.uk/Level%201/Article%20Pics/Wallpaper%20Tree/Wallpaper-tree.jpg>
- <http://www.libsci.sc.edu/bob/isp/lancaster2.htm>
- <http://www.birding.com/idbody.asp>
- <http://www.kenston.k12.oh.us/khs/articles/img/bread-crumbs.jpg>
- <http://twentytwowords.com/wp-content/uploads/funny-comic-desert-island-boat-land.jpg>