Modelling a folksonomy with the postulational approach to facet analysis

Elise Conradi
National Library of Norway
ISKO UK 2011

Two intentions

1. Analyze the use of **the postulational approach to facet analysis** by applying it to a folksonomy

2. Find out what types of structures can be found in a **broad folksonomy** through the process of facet analysis

Four research questions

1. How does one apply the postulational approach to facet analysis to a folksonomy?

2. Where do the results of a facet analysis of this type of data correspond to faceted classification theories and where do they depart?

3. What kinds of facets and conceptual categories can be identified in the folksonomy chosen for this research and how are these characterized?

4. What types of challenges and problem areas exist in a facet analysis of this type of data?

The Folksonomy

- LibraryThing: 940,000 members, 45 million, nearly 5 million individual works, over 58 million tags describing books.
- 76 of the 250 top books yielded in a Tagmash of history and non-fiction
- On average 2070 LibraryThing members per book (between 514 and 10,081)
- 45 most used tags representing each book
- 107,341 instances of 1,275 unique tags

Example: 2 of 76 books

1_Guns, Germs and Steel / Jared Diamond
Agriculture(43) anthropology(581) archaeology(48) biology(79) Civilization(169) cultural studies(19) culture(128) culture diffusion(18) development(21) disease(35) ecology(55) economics(57) environment(62) epidemiology(19) ethnology(48) evolution(114) geography(149) germs(14) history(1,452) jared diamond(17) Natural History(22) nf(15) non-fiction(808) own(52) paperback(15) politics(38) popular science(24) prehistory(16) pulitzer prize(68) read(105) Science(454) social evolution(49) Social History(33) social science(42) societies(13) society(95) sociology(261) tbr(25) technology(70) unfinished(14) unread(102) war(29) WishList(23) World(36) World History(145)

3_The Devil in the White City/ Erik Larson
1893(12) 19th Century(44) 2005(13) 2006(20) 2007(18) America(23) American(25) American history(94) architect(12) architecture(161) Audiobook(14) Biography(32) book club(31) borrowed(13) chicago(525) Chicago history(17) Chicago World's Fair(41) columbian exposition(26) crime(150) Daniel Burnham(12) Ferris Wheel(18) fiction(84) historical(43) historical fiction(51) history(637) Illinois(21) library(13) murder(147) mystery(59) nf(13) non-fiction(626) Novel(17) own(29) read(98) serial killer(177) tbr(30) Thriller(19) to read(14) true crime(191) united states(18) unread(51) us history(20) world fair(13) world's columbian exposition(20) world's fair(201)
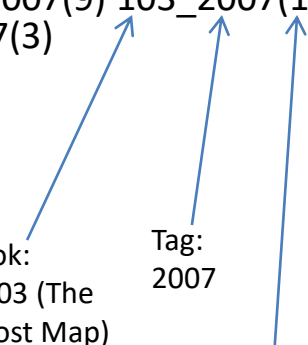
Systematization of data

showing 8 of 1,275 distinct tags:

- 3_2005(13) 14_2005(11) 98_2005(5) 131_2005(4) 244_2005(5)

- 3_2006(20) 8_2006(6) 12_2006(5) 14_2006(13) 19_2006(3) 32_2006(4) 42_2006(5) 45_2006(5) 71_2006(3) 81_2006(2) 94_2006(7) 98_2006(5) 103_2006(4) 134_2006(3) 165_2006(11) 244_2006(5) 249_2006(2)

- 3_2007(18) 8_2007(10) 12_2007(5) 14_2007(12) 19_2007(3) 27_2007(5) 32_2007(4) 42_2007(6) 45_2007(11) 46_2007(4) 70_2007(3) 71_2007(5) 85_2007(4) 94_2007(7) 97_2007(9) 103_2007(13) 121_2007(3) 134_2007(7) 146_2007(19) 165_2007(11) 244_2007(3) 249_2007(3)

- 71_2008(3) 97_2008(2) 146_2008(5) 221_2008(3)

- 131_21st Century(2)

- 63_900(3)

- 63_@(2)

- 165_A.J. Jacobs(4)

Book: #103 (The Ghost Map)

Tag: 2007

Frequency: 13

Two facet analyses

Universe (domain)

- The classificationist performs a facet analysis of a universe in the creation of a faceted classification.

- The classificationist is guided by Ranganathan's Canons of Classification while constructing the scheme.

Entities to be classified

- The classifier performs a facet analysis on the entities to be classified according to the faceted classification.

- The classifier is guided by Postulates to ensure consistency in classifying. These are based on the faceted classification scheme.

How to perform a facet analysis on a folksonomy?!

Classificationist facet analyzes the universe, according to the Canons of Classification

Classifier facet analyzes the entities in the universe under the guidance of Postulates.

6_civil rights(6)

4_civil war(10) 121_civil war(153) 146_Civil War(29) 156_civil war(10)

244_civil war(8)

121_civil war reenactors(4)

1_Civilization(169) 22_civilization(28) 58_civilization(28)
174_Civilization(9)

6_class(8) 100_class(2)

21_classic(80) 52_classic(41) 82_Classic(23) 84_classic(20) 122_classic(8)
156_classic(16) 182_classic(4) 217_classic(11)

21_classical(38) 52_classical(34) 82_Classical(19) 84_Classical(13)

21_Classical Greece(9) 52_classical greece(7)

21_Classical History(23) 52_Classical History(22) 82_classical history(17)
84_classical history(14)

21_classical literature(27) 52_Classical Literature(18) 82_classical
literature(19)

21_classical studies(18) 52_classical studies(12) 82_classical studies(15)
84_classical studies(6)

21_Classics(281) 52_classics(187) 82_Classics(131) 84_classics(64)
156_classics(4) 217_Classics(8)

54_clinton(4)

9_clocks(34)

81_Coal(3)

Seven Canons of Classification

1. The Canon of Differentiation
2. The Canon of Relevance
3. The Canon of Ascertainability
4. The Canon of Permanence
5. The Canon of Concomitance
6. The Canon of Exhaustiveness
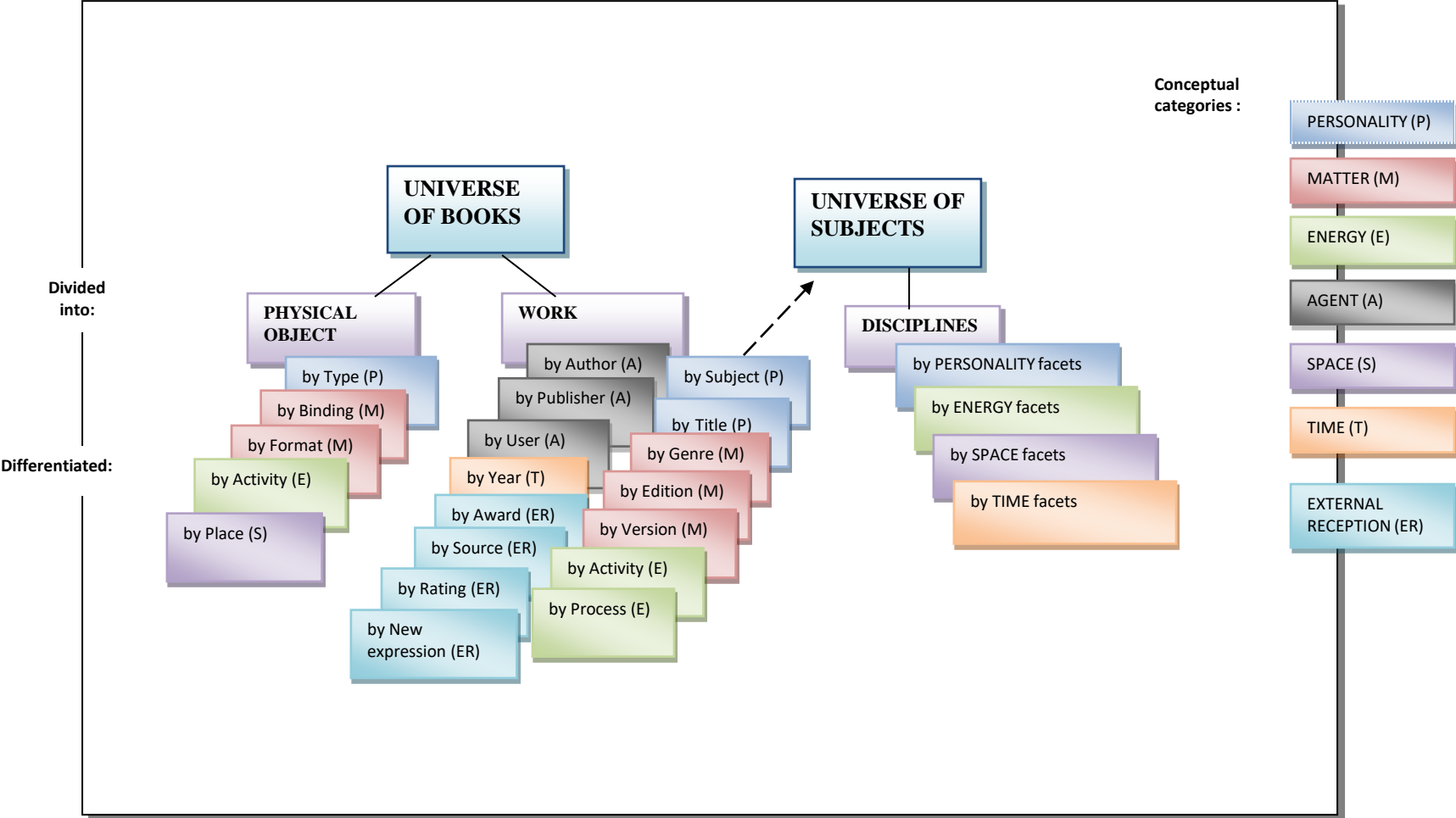7. The Canon of Exclusiveness

Three Postulates

1. The Postulate of Fundamental Categories

2. The Postulate of Basic Facet

3. The Postulate of Isolate Facet

The Postulational Approach

1. Look for conceptual categories to which all the facets in the universes to be classified belong.

2. Look for explicit or implicit basic facets. These represent classes in the universes to be classified.

3. All the explicit or implicit facets found will belong to one and only one of the conceptual categories found. By extension, each tag in the user-generated metadata will belong to one and only one facet.

Results: Ontological model

Conceptual categories :

- PERSONALITY (P)
- MATTER (M)
- ENERGY (E)
- AGENT (A)
- SPACE (S)
- TIME (T)
- EXTERNAL RECEPTION (ER)

UNIVERSE OF BOOKS

Divided into:

PHYSICAL OBJECT
- by Type (P)
- by Binding (M)
- by Format (M)
- by Activity (E)
- by Place (S)

WORK
- by Author (A)
- by Publisher (A)
- by User (A)
- by Year (T)
- by Award (ER)
- by Source (ER)
- by Rating (ER)
- by New expression (ER)
- by Subject (P)
- by Title (P)
- by Genre (M)
- by Edition (M)
- by Version (M)
- by Activity (E)
- by Process (E)

UNIVERSE OF SUBJECTS

DISCIPLINES
- by PERSONALITY facets
- by ENERGY facets
- by SPACE facets
- by TIME facets

Differentiated:

RESULTS: The Universe of Books

Universe	Category	Facet	Examples of tags
Universe of Books	Basic Facet (0%)	By aspect (0%)	*Work* (implicit), *Physical Object* (implicit)
	Personality (70.67%)	By subject (70.54%)	See Table 2: Universe of Subjects
		By type (0.12%)	*audiobook, library book*
		By title (0.01%)	*the histories*
	Matter (22.4%)	By genre (21.89%)	*historical, mystery, non-fiction*
		By binding (0.26%)	*hardcover, paperbook*
		By version (0.12%)	*translation*
		By format (0.07%)	*audio, mp3*
		By edition (0.06%)	*first edition*
		By series (0.01%)	*Hinges of History*
	Energy (4.95%)	By activity (work) (3.74%)	*read, tbr, unread*
		By activity (object) (1.17%)	*borrowed, own, owned, wishlist*
		By process (0.03%)	*illustrated, made into movie, translated*
	Agent (1.11%)	By author (0.76%)	*gibbon, Albert Manguel, Australian author*
		By publisher (0.28%)	*folio, folio society, penguin classics*
		By user (0.07%)	*book club, adult, teen, ya*
	Space (0.17%)	By place (0.17%)	*library, box 2, storage*
	Time (0.36%)	By year written or published or by year read (0.36%)	*100s, 1984, 2006, 2007*
	External Reception (0.33%)	By source (0.15%)	*Comedy Central, daily show, npr, This American Life*
		By award (0.11%)	*pulitzer prize, national book award*
		By new expression (0.05%)	*film, movie, Hinges of History*
		By rating (0.02%)	*favorite, staff pick*

RESULTS: The Universe of Subjects

Universe	Category	Facet	Examples of tags
Universe of Subjects	Basic Facet (52.82%)	By discipline (52.82%)	*biology, history, literature, religion*
	Personality (16.16%)	By person	*aaron burr, Rasputin, sickert, us president, serial killer*
		By group	*women, American Indians, secret societies, marine corps, merovingians*
		By belief-system	*atheism, Judaism, Darwinism, communism*
		By entity	*animals, Mayflower, theory, codes, map, television, culture, books, cod, theory*
	Energy (12.72%)	By energy facets	*cultural diffusion, crime, evolution, murder, politics, reading, shelving*
	Space (14.94%)	By place	*america, boston, college, sea, the west, world*
	Time (3.36%)	By time	*19th century, 1990s, antiquity, dark ages, renaissance*

Facets of History

- History "by Time": *14th century history, 18th century history, 20th century history, ancient history, classical history, colonial history, early modern history, medieval history, modern history, pre-columbian history, pre-contemporary history*

- History "by Place": *african history, american history, asian history, australian history, belgian history, british history, california history, chicago history, chinese history, commodity history, english history, european history, french history, german history, history—us, israeli history, japanese history, london history, maritime history, middle eastern history, russian history, spanish history, texas history, us history, western history, world history*

- History "by Time" and "by Place": *ancient greek history, ancient roman history*

- History "by Energy": *history—wwii, history of reading*

- History "by Group": *family history, jewish history, indian history, military history, native american history, naval history, royal history, tudor history, women's history*

- History "by Entity": *book history, church history, commodity history, culinary history, cultural history¸ food history, history of ideas, history of life, history of medicine, history of sexuality, intellectual history, library history, medical history, social history, urban history*

Theoretical Implications

- Shown how the postulational approach to facet analysis can be used on a folksonomy

- Shown that it is possible to discern facets and categories in tags representing a smaller domain (the Universe of Books)

- Shown that it is difficult to discern facets in tags representing a larger domain (the Universe of Subjects)

Seven Canons of Classification

1. The Canon of Differentiation ✔
2. The Canon of Relevance ❓
3. The Canon of Ascertainability ✗
4. The Canon of Permanence ✗
5. The Canon of Concomitance ✔
6. The Canon of Exhaustiveness ❓
7. The Canon of Exclusiveness ✔✗

Practical Implications

- Identified a user-dimension in faceted classifications and discussed the implications thereof

- Revealed which facets and categories are most popular in the dataset and discussed the use of folksonomies as literary warrant.