UDC: Faceted Classification Today

Faceted classification, analysis and search: some questions on their interrelations

Martin Frické,

School of Information,
The University of Arizona

6/28/2017

1: Representing concepts

Mix of lambda and predicate calculi
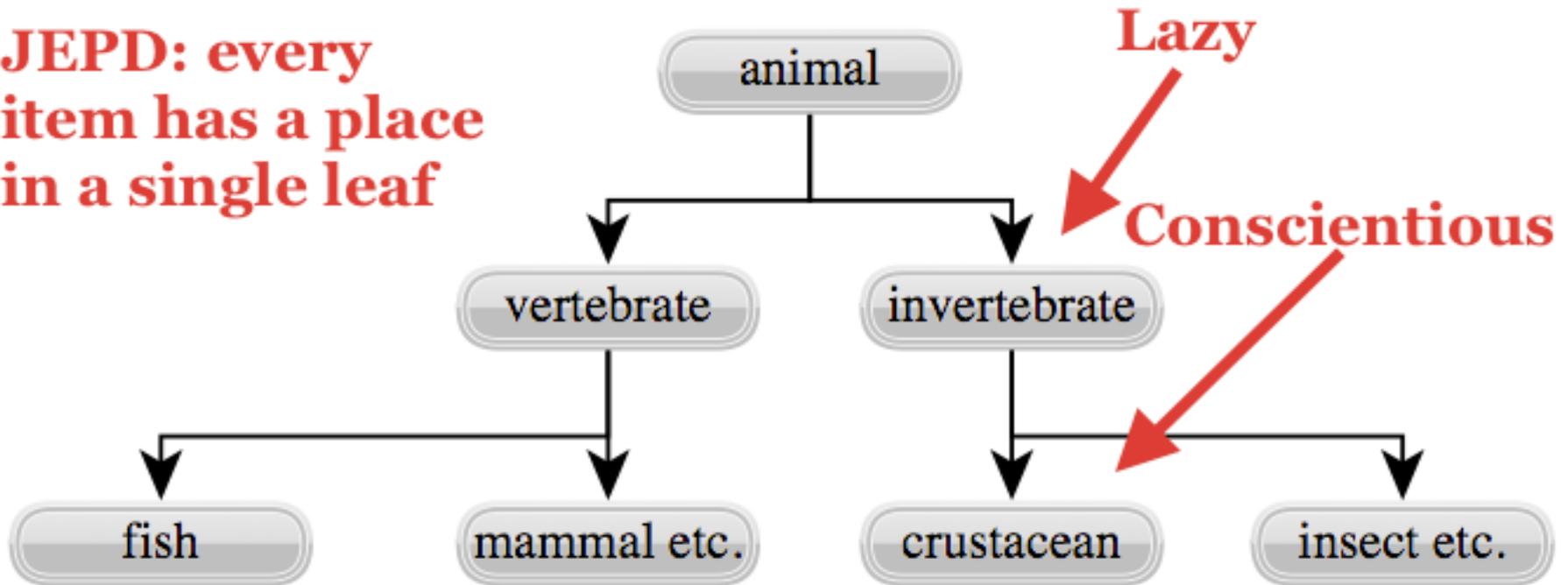
    $\lambda x.(Red(x))$

    $\lambda x.(Green(x))$

    $\lambda x.(Blue(x))$

2: Aristotelian Hierarchical Taxonomy

JEPD: every item has a place in a single leaf

Lazy

Conscientious

```
                          animal
                    ┌────────┴────────┐
                    ▼                 ▼
               vertebrate        invertebrate
              ┌────┴────┐        ┌─────┴─────┐
              ▼         ▼        ▼           ▼
            fish   mammal etc.  crustacean  insect etc.
```

3: Facets I

- z **The kinds of concepts are *facets*, and the atomic concepts within a facet are *foci*.**
- z **Foci (within a facet) *exclusive*. Also sometimes *exhaustive*.**
- z **Foci are *dependent* within a facet (thanks to *exclusivity*) and *independent* across facets (thanks to *orthogonality*).**

4: Facets II

It is the foci that should be used as components in item classification. There can be more general concepts within a facet schedule but these would be used for organization, or for directing a search, and not for the actual labeling of an item.

5: Directed Acyclic Graphs (DAG)
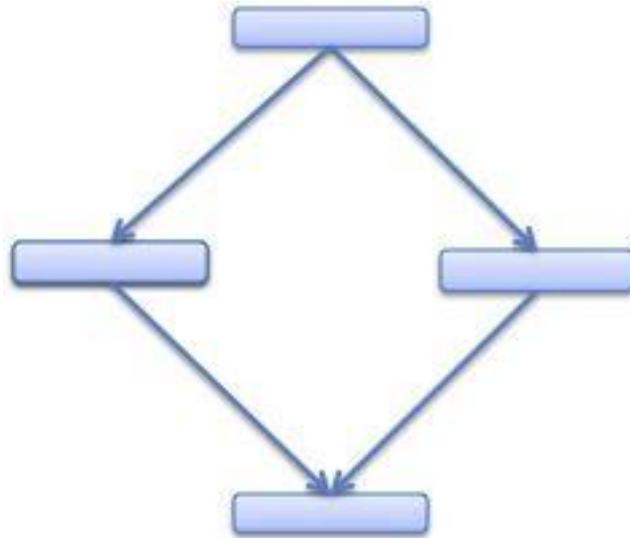
Nodes may have several children, nodes may have several parents, but no node is an ancestor of itself.
Contrast with hierarchy (or tree), in which no node can have more than one parent.

6: Faceted classification

z **Ordinarily polyhierarchical**

z **Otherwise:**

  y **Which goes higher?**

  y **Distributed relatives and repetition**

7: Faceted search I

z **Descriptor e.g. λx.(Female(x))**

z **Often iterative**
     λx.(Female(x))
     λx.(Female(x)&Middleaged(x))

z **Filtering or narrowing**

8: Faceted search II

"Faceted search is a prominent approach for querying document collections where users can narrow down the search results by progressively applying filters, called facets. A facet typically consists of a property (e.g., 'gender' or 'occupation' when querying documents about people) and a set of possible string values (e.g., 'female' or 'research'), and documents in the collection are annotated with property-value pairs. During faceted search users iteratively select facet values and the documents annotated according to the selection are returned as the search result." (Arenas et al., 2014).

9: Faceted search III

Individual facets can have a structure:

"Each facet can be developed/expanded using its own logic and warrant and its own classificatory structure. For example, the Period facet can be developed as a timeline; the Materials facet can be a hierarchy; the Place facet a part/whole tree, and so on." **(Kwasnik, 1999: 39-40)**

Note here about nature of links between concepts

10: Links between concepts

z **"One could construct the tree of our knowledge by dividing it into natural and revealed knowledge, or useful and pleasing knowledge, or speculative and practical knowledge, or evident, certain, probable, and sensitive knowledge, or knowledge of things and knowledge of signs, and so on into infinity…. we are too aware of the arbitrariness which will always prevail in such a division to believe that our system is the only one or the best." (d'Alembert 1751)**

z **Ranganathan talks about a knowledge classification system setting out to be 'helpful'. And, of course, what is helpful may change from context to context.**

11: Ersatz Faceting I

- **one domain (Ranganathan 'simple superimposed classes'),**
- **selection by a single focus, or combination of foci from different facets, merely identifies a subset, subclass, or generating subconcept, of the domain,**

12: Ersatz Faceting II

- **permuting, symmetric, (order indifference of filtering operations),**
- **all the Boolean operations are sound, provided that the mutual exclusivity of foci within a facet is respected.**
- **[from a logical point of view, ersatz faceting ANDs the conditions or attributes]**

13: Ersatz Faceting III

Ersatz faceting is the standard view.

There are formal accounts of faceted classification, using set theory, formal concept analysis, lightweight ontologies, and (mathematical) category theory, etc. All of these use one-place properties or attributes, exclusive foci and orthogonal facets i.e. they all model ersatz faceting.

14: Organizing Amazon I: Goods

- **One obvious possibility is to use ersatz faceting, faceted search, Boolean search, free search, natural language search, and thesaurus support.**
- **Goods facet, Price facet, Gift-for-whom facet, etc.**
- **In sum, ersatz faceting seems adequate for websites, and similar, which give the User selection by progressive filtering via orthogonal properties.**

15: Organizing Amazon II: Services

- **Computers aaS (as a service)**
- **Computers aaS are not computers**
- **Sorted domain**

16: Interlude: Sorted domains I

- **Domains of different kinds of things ('sorts')**
- **Consider a simple (conceptual) library. It has patrons (one sort) and books (a second, different sort).**
- **Now, patrons borrow books, but books do not borrow patrons. Sorts sort (sorry) this out for you. But notice that the 'borrow' relation cannot be used willy nilly in a sorted domain.**

17: Interlude: Sorted domains II

- **A universal domain can fully or partially divided into a sorted domain by use of attributes or conditions (e.g. ... is a book, ... is a patron)**
- **That is not the problem. The problem is that with ersatz faceting we can use any of the facets, in any order, anywhere we like, but when there are sorts this cannot be done**

18: Amazon III: Sorted domain

- **Amazon's goods and services need a sorted domain**
- **Ersatz faceting uses a single domain**
- **Therefore, Ersatz faceting is not good enough for Amazon as a whole**

The argument here is that sorts mean that there cannot be a single faceted classification for everything.

19: Amazon IV: Books

- **Amazon also sells books ('No ****, Sherlock')**
- **Subject tagging or classification**
- **Svenonius 'Referential Semantics'**
- **Subject classification wider**

    Nouns, noun phrases. Nominals, Statements
    Aspects & very complex concepts
    Multiple tag tagging

20: Topics as aspects

λx.(Habitat(x)&Rabbits(x))

/* 'habitats which are also rabbits' */

Rather, it needs a function:

λx.(x=habitatOf(rabbits))

This is not now ersatz faceting—the collective habitats of rabbits is not a narrowing or filtering of rabbits.

21: Facet analysis I

- **Familiar to this audience**
- **Computer scientists have also been very active in this area**
- **language, not concepts. Selective domain analysis is done on titles, contents and communities, because the classification is domain dependent. The terms are clustered. Prominent clusters become facets. Exemplars from the facets are chosen as foci.**

22: Facet analysis II

Remark: the order of the facets in a compound concept('citation order'), and the order of foci within the facets ('filing order')— not so important. Shelving books is no longer central.

"The Dewey Decimal Classification, in whatever edition, and the shelf of drab paper-bound volumes that are the classification of the Library of Congress, are not a gate though which the mind is led into the recorded world of the human adventure, they are only an address-book for the library stacks." (Jesse Shera 1965, p. 134)

23: Facet analysis III

- **Computer scientists typically use only conjunction on the foci (Term Composition Algebra (Tzitzikas, et al. 2004) i.e. ersatz faceting).**
- **Typical library science facet analysis would permit non-conjunctive types e.g. 'human growth'**
- **Researchers in librarianship have been very sophisticated in what they have devised (Farradane, 1950, 1963; Moss, 1964; Gardin, 1965, 1969; Perreault, 1965, 1969; Austin, 1984).**

24: Facet analysis IV

They have faceted down to foci but allow simple relations, which can be functions, between foci e.g. 'explosions causing injuries'.

There is a richness here that is not present in ersatz faceting. There are functions, relations and basically the full range of logical connectives. But this comes at a price as far as faceting is concerned.

25: Facet analysis V

The relations themselves between foci need organizing. Here are some relations:

     ... causing ..., ... correlating with ...
     ... preventing ..., ... giving ...
     ... donating ..., ... loving ...

     or, more generally, for example, constructions expressed by transitive
     verbs or comparative adjectives.

How are these to be organized or faceted? It is an open question.

26: A positive suggestion

Using functions as the atomic components, and function applications and function composition as the constructors e.g. formal linguistics, Fregean approach to concept analysis, functional programming in computer science such as that of Haskell (Haskell.org, 2017). Haskell has a type system and mechanisms for combining functions and types.

27: Conclusion

- **Librarian faceted classification and Computer science faceted search somewhat mutually in stress**
- **Computer science tends to stick with ersatz faceting, which is not enough**
- **Librarians do allow much richer constructions in their subject heading languages but a) how are these to be organized and faceted? and b) how should it fit with faceted search as an iterative filtering**

28: The future I [Right now]

- **Perfect absolute permanence of digital recordings**
- **Perfect integrity, tamper proof**
- **Perfect date, time, and authorship assurance**
- **Perfect swift, censor proof, delivery**

29: The future II [Right now]

Decentralized, Distributed:
- Cryptographic hashes (perfect integrity, changes transparent)
- Content based indexing
- Cryptographic data structures (Merkle trees, DAGs, the blockchain)
- Distributed Hash Tables
- Distributed 'swarm' delivery
- (Smart) contracts, game theory, and cryptoeconomics

30: The future III [Right now]

Call me Ishmael. Some years ago - never mind how long precisely - having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world. It is a way I have of driving off the spleen, and regulating the circulation. Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypos get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people's hats off - then, I account it high time to get to sea as soon as I can. This is my substitute for pistol and ball. With a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship.

bd603c18925076dedf70f2a6bd0a6eae
88b5bb93472fa8fc55bc3a919922dad6

31: The future IV [Right now]

bd603c18925076dedf70f2a6bd0a6eae
88b5bb93472fa8fc55bc3a919922dad6

b4127b6ed6d3e20be5a355235df6a848
071045eedc46e329223dcd84f0ce1a98

Infrastructure (content base indexing, chunking, deduplication).

Inter Planetary File System. Filecoin

32: The future V [Partial agenda]

Attach meaningful names to the numbers. Distributed name space control. Mutable, immutable.

Indexing will make a comeback (indexing essentially subject classification—> faceted classification) Not easy to do real time text in text search. Buy the text. Pre-emptive.

33: The future VI [Partial agenda]

There are deep intellectual challenges for facet analytical theorists.

I commend these challenges to you.

Thank you!