

# Faceted classification as the basis of all information retrieval

A view from the twenty-first century

## The Classification Research Group Agenda:

---

- in the 1950s the Classification Research Group was formed to investigate the problems of managing (particularly) scientific information
- a group of practising librarians, academics, and researchers, they were generally admirers of the work of S. R. Ranganathan and the principles of faceted classification
- they developed a form of faceted classification that advanced the original theory of Ranganathan, and is particularly British in flavour
- in 1955 they published what has been regarded as the CRG manifesto with the objective: 'the need for a faceted classification as the basis of all information retrieval'.

## The view from sixty years on:

---

- to what extent has this objective been met?
- what do we understand faceted classification to be?
- is it just part of a general trend towards more structured information systems?
- what is the relationship between classification and information retrieval?
- is faceted classification just a method of building knowledge organization systems?
- or is there a sound theory underpinning that methodology?
- if so, where does the theory sit in scientific terms?

## The pervasiveness of the faceted approach:

---

- although it would be bold to say that faceted systems underpin all information retrieval, there is certainly evidence of the widespread impact of facet analysis
- Hjørland has called it 'probably the dominant approach to knowledge organization in the twentieth century'
- it is clearly influential in the design of classifications per se, and other managed knowledge organization systems
- a version of faceted classification can be seen in the organization and search tools of many websites, particularly in e-commerce
- the theory as developed by the CRG 'school of thought' has also contributed to semantic web work

## Faceted influence on subject heading lists:

---

"The first phase of the FAST development includes the development of facets based on the vocabulary found in LCSH topical and geographic headings and is limited to six facets: topical, geographic, form, period, with the most recent work focused on faceting personal and corporate names. This will leave headings for conference/meetings, uniform titles and name-title entries for future phases. With the exception of the period facet, all FAST headings will be fully established in a FAST authority file."<sup>24</sup>

# Wine.com advanced search:

## Advanced Wine Search



Sort By:  View:  [View results](#)

6150 matching products [start over](#)

**type of wine**

- Red Wine
- White Wine
- Champagne & Sparkling

**style of wine**

- Red - Light & Fruity
- Red - Smooth & Supple
- Red - Earthy & Spicy
- Red - Big & Bold

**grape varietal**

- Cabernet Sauvignon
- Chardonnay
- Sauvignon Blanc
- Pinot Gris/Grigio

**special designation**

- Magnums & Larger
- Half Bottles
- Great Bottles To Give
- Green Wine
- Screw Caps
- Kosher
- Watch The Video

**wine region**

- California
- Washington
- Oregon
- Other US
- Australia
- New Zealand
- France - Bordeaux

**specific appellation**

- Champagne
- Chile
- Coonawarra
- Austria
- Barossa
- Burgundy
- Napa Valley

**vintage**

- Non-Vintage
- 2014
- 2013
- 2012
- 2011
- 2010
- 2009

**professional rating**

- Wine Spectator
- The Wine Advocate
- James Halliday
- Connoisseurs' Guide
- International Wine Cellar
- Wilfred Wong Of Wine.Com
- Wine & Spirits

unrated

**price**

**alcohol by volume (%)**

**bottle size**

- 300ml
- 375ml
- 500ml
- 720ml

# Faceted browsing:

The screenshot displays the University of Edinburgh Library Catalogue interface. At the top, the header includes the university logo and the text "THE UNIVERSITY OF EDINBURGH INFORMATION SERVICES". Below this, navigation links for "Information Services", "Ask a Question", "Classic Catalogue", and "User Login" are visible. The main search bar contains the term "horse", and the results are sorted by "relevance".

On the left side, a faceted browsing interface is shown as a word cloud. The central word is "horse". Other words are connected to it by lines, indicating their relationship to the search term. The words are color-coded: red for "Association" and grey for "Spelling variation". The words include: Springer, force, horse, disease, cattle, pony, farrier, lameness, breeding, sheep, match, horse, foot, cure, arm, armes, run, incident, treatise, regiment, horsey, care, vhorse, Ross, and steed.

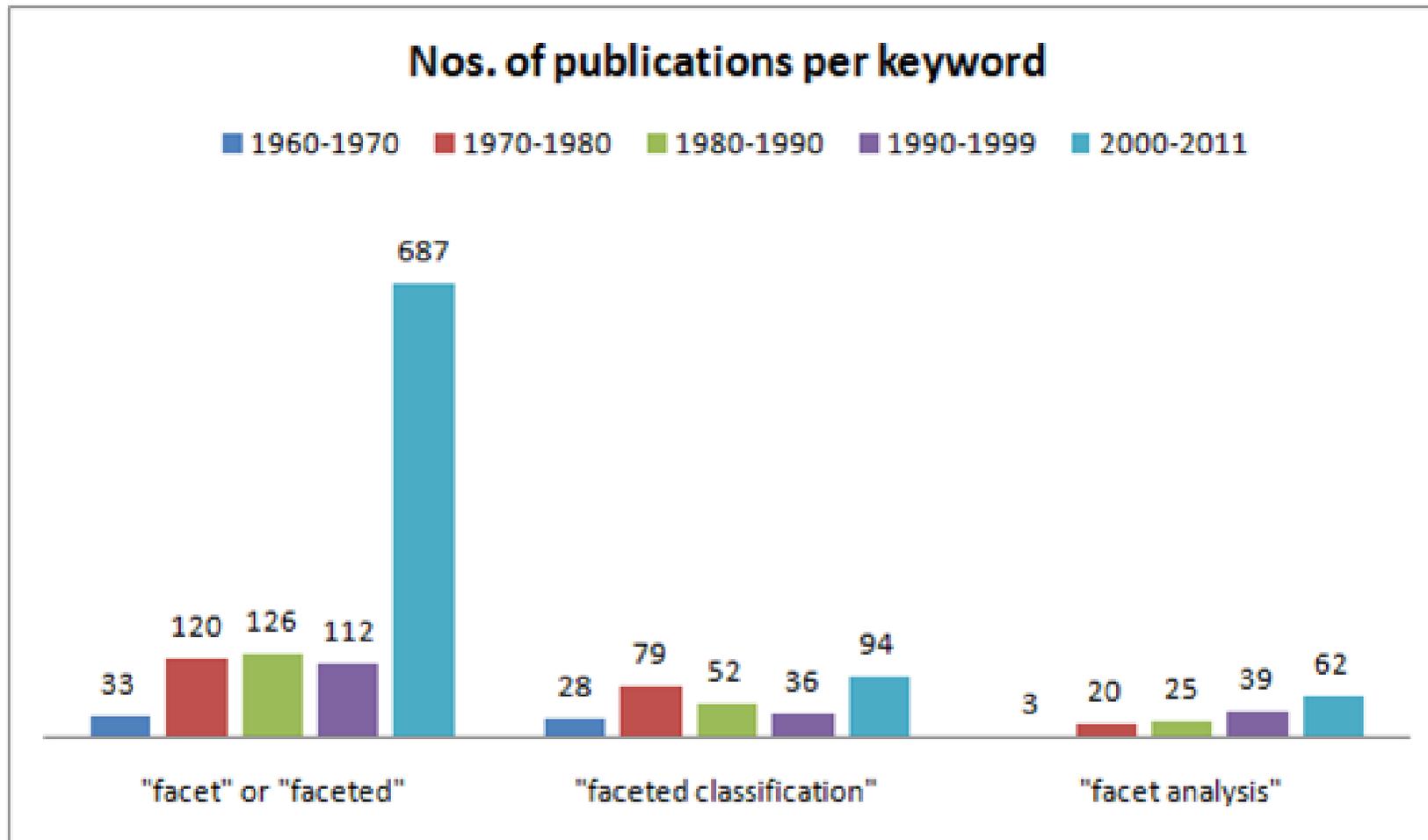
The search results on the right are as follows:

- Horse**  
Walker, Elaine, 1960-  
Book  
Publisher: London : Reaktion Books, 2008  
Series: Animal  
Subject: Horses  
Matches: Horse (7) , horses (4)  
Availability: Main Library, GT5885 Wal  
[Check availability](#)
- The horse : how to breed and rear him : race-horse, hunter, hack, troop-horse, draught-horse, shire-horse, carriage-horse, Pony**  
Day, William, 1823-1908  
Book  
Publisher: London : Richard Bentley and Son, Publishers in Ordinary to Her Majesty the Queen, 1888  
Subject: Breeding, History, Horse breeds, Horsemanship, Horses, Training  
Matches: Horse (8) , horses (3)  
Availability: Library Annex, .6361 Day  
[Check availability](#)
- Veterinary diagnostic imaging : the horse**  
Farrow, Charles S  
Book  
Publisher: St. Louis, Mo. : Mosby Elsevier, 2006  
Subject: Diseases Diagnosis, Horse Diseases, Horses, ultrasonography, Veterinary diagnostic imaging  
Matches: Horse (2) , horses (1)  
Availability: Vet Library, SF 951 Far 2006  
[Check availability](#)
- The horse : diseases and clinical management**  
Kobluk, Calvin N  
Ames, Trevor R  
Geor, Raymond J  
Book  
Publisher: Philadelphia : W.B. Saunders, 1994  
Subject: Diseases, Horses  
Matches: Horse (1) , hor (2) , horses (1)  
Availability: Vet Library, SF 951 Hor 1994

On the right side of the results, there are several filters:

- Select Location**: Showing results from: All locations
- Refine**
  - Format**
    - Book (4,458)
    - E-Book (1,814)
    - Music score (50)
    - Journal (49)
    - Music recording (26)
    - 3 more...
  - Publication date**
    - Last 5 years (458)
    - Last 10 years (858)
    - Last 20 years (1,532)
    - Last 50 years (2,144)
    - Older than 50 years (2,454)
    - Before 1850 (2,095)
  - Author**
    - England and Wales. Parliament (134)
    - Markham, Gervase, 1568?-1637 (75)
    - England and Wales. Sovereign (1625-1649 : Charles I) (43)
    - Wesley, John, 1703-1791 (33)
    - Charles I, King of England, 1600-1649 (30)
    - 4,250 more...
  - Subject**
    - Horses (1,123)
    - History (793)
    - Great Britain (620)
    - Diseases (327)
    - Veterinary medicine (166)

At the bottom right, there is a vertical text: "powered by Aquabrowser Librar".



**Figure 1: Representation of facets in the LIS literature**

## The versions of faceted classification:

---

- there are several different models of faceted ‘classification’
- what might be called the ‘classical’ Ranganathanian or (modified) CRG version
- the CRG did not have a single model (Farradane’s approach was somewhat different)
- Spiteri has produced a model that reconciles the differences between these two essentially similar approaches
- the e-commerce version is simpler and flatter in terms of the facet structure, and although there is a body of commentary, there is not much theoretical basis
- ‘faceted browse’ as used in discovery tools is very similar in nature to e-commerce
- the web enabled model managed in SKOS has more difficulty in representing the nuances of the fully fledged ‘classical’ model

## What was novel about faceted classification?

---

- classification theory in the early twentieth century (ignoring Otlet and the UDC)
- based on practical needs of collections and users
- philosophical and conceptual basis is in pragmatism (Bliss after John Dewey)
- much use is made of traditional logic to establish class relationships
- notions such as sub- and super-ordination, class membership based on attributes
- relationships are very precisely addressed, but they tend to concentrate on hierarchies

## The faceted classification:

---

- the faceted classification is more structured in design
- there is a more scientific and mathematical feel about it
- there is a greater sense of regularity in the structure
- it avoids pre-coordination
- there is an underlying theory/methodology that provides a model for building classifications in a standard manner
- this can be applied to all subject domains
- there is some sense of a general theory of library or information science that is independent of the specific needs of users or collections in subject domains
- overall it is less context dependent and less pragmatic in approach
- it has all the appearance of a general theory

# What did the CRG think the faceted classification had to offer?

---

According to the manifesto:

- the display of useful generic relationships;
- full and accurate cross-referencing;
- accurate application of principles of division;
- a clear citation order;
- established rules for compounding;
- an appropriate notation.

Why would we think the faceted classification is particularly good for information retrieval?

---

- In the late 1990s there was a tranche of papers promoting faceted classification as the answer to search and retrieval on the world wide web
- what aspects of faceted systems did they promote?
- largely design and construction features
- rigorous analysis
- logic of the structure
- management of hierarchical relationships

# What would we regard today as advantages of the faceted approach?

---

- largely end-user features
- providing a map of the domain
- to a good degree intuitive to use
- the capacity to manage complex content
- to visualise complex content in a way that supports browse and search
- it makes a very good basis for a visual display
- support for query formulation
- support for query modification
- compatibility with automatic search

## Classification and information retrieval: what did the CRG understand the relationship to be?

---

- it's clear that from the outset, some CRG members regarded the two as synonymous
- papers published by Vickery and Foskett conflate the two ideas
- in a pre-machine age faceted classification had several advantages over an enumerative scheme
  - complex content could be more easily expressed
  - citation order provided a way of dealing with the placing of that content more consistently and predictably
- but mechanised information retrieval offered alternatives to linear arrangement

## Use of classification in mechanized retrieval systems:

---

- In other environments classification was being used in conjunction with the development of mechanized systems
- Seminars on UDC in mechanized retrieval (1969, 1970, 1976)
- Vickery's 1958 international conference paper also compares the CRG work with some other retrieval tools

Perry [42]	Andrews & Newman [6]	Vickery
<b>Analytic relations</b>	<b>Modulants</b>	<b>Facets</b>
Class inclusion Material of composition Whole-part Process Agent Patient or product Attributive Negative	Process Apparatus Product Starting-material Intermediate Condition Made-from Combination-including	Substance, product Organ or part Constituent Property Patient Action, operation or process Agent Apparatus

## The division between classification and information retrieval:

---

- it seems that some members of CRG did not see the necessary connection between faceted classification and retrieval in an electronic context
- there was a bifurcation between what might be loosely described as the 'library scientists' and the 'information scientists'
- at a fairly early stage Vickery left the Group, and the emphasis on that electronic dimension diminished
- the library scientists continued to explore the theory of faceted classification with the broad objective of creating a 'new British classification scheme'
- interestingly, a similar split occurred in the United States

## Facet analysis as a tool building methodology:

---

- this is the area of greatest influence
- introducing better (i.e. more logical) structure to a whole range of conventional information management tools
- improves user understanding in terms of consistency and predictability
- a tested and proven methodology
- a generalised methodology that can be applied equally well to different domains, subject or otherwise
- a generalised methodology that can produce different kinds of KOS
- a methodology that is included in the international standard for structured vocabularies
  
- the model of a faceted KOS that we have nowadays is more sophisticated than the originals

## How reliable is the theory attached to that methodology?

---

- conventionally, facet analysis has been regarded as rationalist in approach
- its claims of intellectual and logical rigour would tend to reinforce that view
- with hindsight it's quite crude and remarkably full of holes ....
- Ranganathan does not introduce the idea of fundamental categories at all until Edition 3 of the Colon Classification
- (although there is a rudimentary sense of facets of different subjects)
- he never gives an adequate exposition of the categories, and adopts them more or less intuitively (Foskett)
- he is never able to clearly define P, which is 'too elusive' and 'ineffable'

**T EDUCATION**

*Divisions based on  
Educand or E  
Characteristic*

- 1 Pre-secondary
- 12 Pre-school child

*Divisions by the  
Chronological Device  
(Illustrative)*

- 2J92 Commenius, object teaching
- 2L12 Rousseau's school
- Kant's school
- Pestalozzi's school
- Fichte's school
- Herbart's school
- Froebel's kindergarten
- Mann's school
- Manual training
- Montessori school
- Project method
- Dalton plan

*Divisions based on  
Educand or E  
Characteristic*

- 1 Pre-secondary

- 55 Female
- 6 Abnormal
- 61 Genius
- 62 Idiot
- 63 Insane
- 65 Criminal
- 67 Deaf and dumb
- 68 Blind
- 7 Backward classes

*Divisions based on  
Problem or P  
Characteristic*

- 1 Nomenclature
- 2 Schools of Education

*to be divided by The  
Subject device with the  
following modifications--*

In the case of Philology omit the stage number and instead of the language numbers substitute the following :

- 1 Mother Tongue
- 5 Foreign language
- 8 Classical language)
- 4 Organisation
- 41 Admission, fees, scholarship, term
- 411 Admission

- 412 Fees, scholarship 46 Finance
- 4125 Research fellowship 47 Educational measurements
- 413 Term, holidays, attendance 471 Intelligence tests
- 414 Classes, Grades *Subdivisions by the Chronological Device (Illustrative)*
- 415 Time-table
- 417 Residence

*Divisions by the  
Chronological Device  
(Illustrative)*

- 2J92 Commenius, object teaching
- 2L12 Rousseau's school

- 45 activities 7 Students' social life and organisation
- 45 State, community and education 71 International students' organisation
- 451 Board of Education
- 4515 Compulsory education 8 Special topics
- 452 Community and education 81 Medium of instruction
- 4528 Parent associations 83 Study methods
- 85 Visual instruction

# T EDUCATION

T [E]: [P]

<i>Divisions based on the train of Educand or E</i>		7	<b>Backward classes</b>
<i>Characteristic</i>		9	<b>Other classes</b>
1	<b>Pre-secondary</b>	<i>To be divided by the Subject Device</i>	
13	Pre-school child	<i>(illustrative)</i>	
15	Elementary	9Y131	Rural community
2	<b>Secondary</b>	9Y152	Aristocracy
25	Intermediate	9Y154	Military classes
3	<b>Adult</b>	<i>Divisions based on the train of Problem or P</i>	
31	Literate	<i>Characteristic</i>	
35	Foreigner	1	<b>Nomenclature etc.</b>
38	Illiterate	2	<b>Schools of Education</b>
4	<b>University</b>	<i>Divisions by the Chronological Device</i>	
42	Pass	<i>(illustrative)</i>	
43	Honours	2J92	Commenius. object
45	Post-graduate		
48	Research		

## CRG style faceted classification

---

- there's no single coherent statement of CRG theory
- the nearest equivalents are the *Faceted classification in .....* series, and the Introduction to the Second Edition of Bliss's Bibliographic Classification (BC2)
- BC2 is the most comprehensive in scope
- close scrutiny shows that does not include much that is conceptual
- it's really a detailed account of the methodology
- like the early papers on the CRG approach, it tells you *how* facet analysis works, but not *why*
- It would be fair to say that the work concentrated on developing the methodology rather than testing the theory

## Bliss's Bibliographic Classification 2<sup>nd</sup> edition

---

- apart from CC, BC2 is the only existing general faceted classification scheme
- it is the manifestation of CRG style facet analysis
- like CC, the early drafts show lots of anomalies

# BC2 draft schedule for Music and its source, British Catalogue of Music

B	Female or children's voices		
D	Soprano	AF	Female or Children's voices
F	Mezzo-soprano	<b>AFB</b>	<b>Boys' voices</b>
H	Alto	AFL	Soprano
		AFN	Mezzo-soprano
		AFQ	Alto
J	Changing voices	AFS	Changing voices
L	High voices	AFT	High voices
M	Middle voices	AFV	Middle voices
N	Low voices	AFX	Low voices
P	Male voices	AG	Male voices
		<b>AGG</b>	<b>Alto voice. [Countertenor].</b>
		<b>AGGC</b>	<b>Castrato voice</b>
Q	Alto	AGH	Tenor
R	Tenor	AGN	Baritone
S	Baritone	AGX	Bass
T	Bass	AHN	Falsetto
		AHT	Whistle
V	Falsetto	AHY	Speaker

## Categories as the basis of content analysis and modelling:

---

- the use of categories as an analytical method predates Ranganathan by some decades (Brown, Kaiser)
- categories are also a common feature of other modelling methodologies
- soft systems theory
- grounded theory
- many information retrieval systems of the mid-twentieth century (faceted and otherwise) use a very wide variety of categories

## Vickery's comparison of categories:

---

Perry [42]	Andrews & Newman [6]	Vickery
<b>Analytic relations</b>	<b>Modulants</b>	<b>Facets</b>
Class inclusion	Process	Substance, product
Material of composition	Apparatus	Organ or part
Whole-part	Product	Constituent
Process	Starting-material	Property
Agent	Intermediate	Patient
Patient or product	Condition	Action, operation or process
Attributive	Made-from	Agent
Negative	Combination-including	Apparatus

Andrews, D.D. and Newman, S.M. (1956) Storage and retrieval of contents of technical literature, non-chemical information, 15 May 1956, US Patent Office, Research and Development Report (Report)

Perry, J.W.; Kent, A. and Berry, M.M. (1956) Machine literature searching, Interscience: New York (Book).

## Categories in soft systems theory:

---

<b>Soft systems elements</b>	<b>Facet categories</b>
C = customers	patients
A = actors	agents
T = transformations	actions/operations
W = weltanschauung (world view)	discipline
O = owner	entity/organization
E = environmental constraints	properties/context

## Categorization as a theory building methodology:

---

- categorization is very typical of content analysis methodologies developed from the 1960s onwards
- grounded theory is perhaps the most advanced of these
- it was developed to give a proper 'scientific' foundation to qualitative research methodologies
- the principle feature of grounded theory is that the evidence base and analysis precedes the theory
- it is a 'theory building' methodology

## Facet analysis as theory building:

---

- it is suggested that the theory of faceted classification emerges from the practice of designing classifications, rather than the reverse
- both Ranganathan and the CRG were slow to develop a complete theory
- the model faceted classification comes quite late on in the history
- consequently, facet analysis can be regarded as a more flexible approach than is sometimes perceived
- facets can be continually discovered and re-discovered from the analysis of the domain if that is regarded as a text
- this may well be what is happening on a more intuitive basis with many recent manifestations of faceted KOS

## Some conclusions:

---

- facet analysis is very influential on all kinds of current knowledge organization and retrieval tools
- there are quite different models of what constitutes a faceted system
- the role of classification in information retrieval is also differently understood
- in broad terms, facet analysis provides a sound methodology for building a KOS
- the general idea of categorical analysis predates Ranganathan and the CRG
- what they did was innovative in its day, but now seems poorly formulated
  
- the ideas about why faceted classification is useful change over time
- the conceptual framework for facet analysis can also shift
- facet analysis is an evolving approach to knowledge organization