

Legal Knowledge Representation Using a Faceted Scheme

Michelle Cumyn
Faculté de droit
Université Laval
Québec, Québec, Canada
michelle.cumyn@fd.ulaval.ca

Günter Reiner
Fakultät für Wirtschafts- und
Sozialwissenschaften
Universität der Bundeswehr
Hamburg, Germany
guenter.reiner@unibwh.de

Sabine Mas & David Lesieur
École de bibliothéconomie et des
sciences de l'information
Université de Montréal
Montréal, Québec, Canada
sabine.mas@umontreal.ca
david.lesieur@umontreal.ca

1 Subject Indexing in the Legal Domain

A database supports legal research by matching a user's request for information with documents of the database that contain it. Indexes are among the oldest tools to achieve that aim. Many legal publishers continue to provide manual subject indexing of legal documents, in addition to automatic full-text indexing, which improves the performance of a full-text search.

There have been numerous attempts to develop systems for the automatic subject indexing of legal texts (see Francesconi *et al.* (eds.) 2010, Part 3). Publishers continue to use human indexing (albeit computer-assisted), despite its cost, because fully automated indexing does not meet their standards. A search process that is humanly curated still delivers better results (Nevelow Mart 2013, at 43, comparing Westlaw and Lexis).

Subject indexing in the legal domain often follows a predefined sequence of search terms flowing from the relevant subject area. A well-known example is Westlaw's "Key Number System," which divides the legal order in the manner of a textbook or manual, with multiple levels, into individual areas and aspects to which it assigns approximately 100,000 searchable numbers. Because of the scheme's rigidity, the classification of documents is sometimes artificial, and representation of their content incomplete. At the opposite extreme, subject indexing that is not structured lacks consistency. What is needed is a more flexible scheme for subject indexing. Our research program investigates the use of facets to that end.

2 Introducing a Faceted Scheme

Faceted classification and indexing schemes are not new, even in law. The Indian mathematician and librarian Shiyali Ramamrita Ranganathan (1892-1972) conceived the theory and method of facet analysis. He noted the emergence of new specialized or interdisciplinary subjects, with which existing classifications were unable to cope (Satija 2017, at 292). According to Ranganathan,

every subject matter or unit of knowledge can be broken down according to five fundamental and mutually exclusive categories or "facets." Ranganathan originally described them as a "train of characteristics" and arranged them in order of "decreasing concreteness:" 1) Personality (entities or things); 2) Matter (materials or constituents of things); 3) Energy (actions or activities); 4) Space; and 5) Time (see the "Postulates" 1 and 2 of Facet Analysis, Ranganathan 1959, at 67 *et seq.*, 37).

Ranganathan's facets, which one can think of as dimensions of a topic, revolutionized the field of knowledge organization. The Classification Research Group (CRG), founded in the United Kingdom in 1952, took it up, expanding Ranganathan's original facets to 13 categories representing a "production process," and being "particularly suitable for the analysis and organization of terms in technology:" Thing/entity, Kind, Part, Property, Material, Process, Operation, Patient, Product, By-product, Agent, Space and Time (Broughton 2001, at 79). Vanda Broughton, a member of the group, explains that the "real genius of the CRG lies in its shaping of facet analysis as a generalized methodology for information retrieval" (Broughton 2011, Conclusion). The second edition (1977) of the Bliss Bibliographic Classification was built upon the principles developed by the CRG (Broughton 2001, at 74).

After participating in the CRG, Broughton sought to implement the Bliss Classification in the legal domain (Broughton 2010, at 37-40). In a similar vein, Sweet & Maxwell, a British legal publisher, created a "Legal Taxonomy" for structuring its indexers' thesaurus in accordance with the principles of facet analysis (Scott and Smith 2010, at 217). In our view, these attempts do not fulfil the promise of faceted schemes applied to law, for reasons that we have explained elsewhere (see Reiner *et al.* 2019, at 350, 351). One reason is that they continue to rely on lists of topics that do not reflect the factual dimensions of legal information.

3 Our Project and Prototype

Inspired by Ranganathan's theory of faceted classification, we have developed an indexing model that follows a content grid of six predefined categories (or facets in a broad sense). This structure represents the "grammar" of legal information contained within documents of the database. It facilitates the search for cases with similar facts and supports legal thinking by revealing relationships between facts and legal responses, as they co-occur within documents of the database.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ICAIL '19, June 17–21, 2019, Montreal, QC, Canada

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6754-7/19/06.

<https://doi.org/10.1145/3322640.3326735>

We found Ranganathan's scheme to be simpler and more intuitive than the CRG's. We noted its similarity with Gaius' famous tripartite division, well known to all Western legal systems. Gaius, a Roman jurist, declared: "all our law is about persons, things and actions" (*Institutiones*, at 1.8). For Ranganathan, persons and things belong to the facet Personality, but in law, the distinction is so fundamental that it is necessary to provide distinct facets. After experimenting with various schemes, we arrived at the following six facets: 1) Person, 2) Action, 3) Thing, 4) Context, 5) Legal Category and 6) Sanction. The "Context" facet comprises Ranganathan's "Matter," "Time" and "Space." "Legal category" and "Sanction" are not true facets; they resemble traditional classifications of legal topics.

Traditional classifications and indexes favor the use of abstract legal concepts such as "civil liability" and "damages," which belong to our facets "Legal Category" and "Sanction." However, such concepts are not very helpful in describing the facts of a case. The first four facets of our scheme represent the facts within a structure that is both intuitive and legally relevant. To give them content, we had to develop our own controlled vocabulary, since we could not rely on existing thesauri. Our controlled vocabulary (in French) was developed incrementally, during the indexation process, under the supervision of an expert librarian.

We have built a prototype database in accordance with our indexing model. The prototype contains 2,500 cases, mostly in French, in the areas of administrative law, labor law and the law of obligations. We have manually indexed each case using our controlled vocabulary. Our indexing policy indicates how to assign indexed terms to the appropriate facet. Figure 1 provides an example of our indexing scheme applied to a case of the database (*L'Italian c. Beauguard*, 2012 QCCQ 5822).

Indexation ▾	
Personne	voisin
Action	attaque empiètement
Chose	animal > animal de compagnie
Contexte	blessure > morsure
Régime juridique	responsabilité civile > fait d'un animal
Sanction	dommages-intérêts > dommages-intérêts compensatoires dommages-intérêts > dommages-intérêts moraux

Figure 1: Indexation of a case from our prototype

Our prototype's search interface offers three search modes that a user may combine at will: a full-text search, a search using indexed terms and the selection of indexed terms from within each facet. Facets appear alongside the results list and operate as filters.

We expect our model to enhance the use of indexed terms in searching for comparable cases, to help users design targeted searches with greater precision and to present more clearly the information that the search itself generates. The collocation of indexed terms within documents of the database represents

meaningful relationships between concepts, which the interface renders more visible; it also suggests possible analogies.

4 Steps Ahead

The next step is to test the performance of our model, for purposes of information retrieval, by conducting a user study. If results are positive, a further step will be to consider whether our model may serve as a tool for the (semi-)automatic indexing of legal documents. It is conceivable that our facets and controlled vocabulary may contribute to a process of machine learning. The current size of our prototype database is probably too small for supervised machine learning algorithms to become effective at the task of automatic indexing. However, it is possible to expand it manually thanks to our detailed indexing policy. With an expanded database, we could explore supervised learning approaches and adapt them to the characteristics of our indexing scheme.

Finally, the 2,500 sets of manually attributed index terms are valuable in themselves. They formalize the legal essence of the database content in a matrix (or template) that resembles a light ontology, revealing statistically analyzable information about the relationships between concepts (see Schweighofer 2009). Such information may support the development of other AI applications in the legal domain.

ACKNOWLEDGMENTS

This research program is funded by the Social Sciences and Humanities Research Council of Canada (2015-2020). We are grateful to Michèle Hudon, retired professor from the Université de Montréal, for her invaluable contribution to this project. We are also grateful for the research assistance of Christophe Achdjian, Dominique Boulanger, Cécile Gaiffe, Marjorie Gauchier, Joanie Gosselin-Lepage, Ariane Joncas, Béatrice Lecomte, Amélie Roy, Erin Sandberg, Forrest Wakarchuk and Iryna Zazulya. Thanks to Alexandre Fortier, our expert librarian and Charles Tremblay-Potvin, our project coordinator. Finally, we would like to thank the Société québécoise d'information juridique (SOQUIJ) for their cooperation and interest.

REFERENCES

- [1] Broughton, Vanda (2001). Faceted classification as a basis for knowledge organization in a digital environment; the Bliss Bibliographic Classification as a model for vocabulary management and the creation of multi-dimensional knowledge structures. *The New Review of Hypermedia and Multimedia* 7(1): 67-102.
- [2] Broughton, Vanda (2010). The use and construction of thesauri for legal documentation. *Legal Information Management* 10(1): 35-42.
- [3] Broughton, Vanda (2011). Brian Vickery and the Classification Research Group: the legacy of faceted classification. <https://www.researchgate.net/publication/266184326>.
- [4] Francesconi, Enrico, Montemagni, Simonetta, Peters, Wim, Tiscornia, Daniela (eds.) (2010). *Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language*. Heidelberg.
- [5] Nevelow Mart, Susan (2013). The Case for Curation: The Relevance of Digest and Citor Results in Westlaw and Lexis. *Legal Reference Services Quarterly*, 32:1-2: 13-53.
- [6] Ranganathan, Shiyali Ramamrita (1959). *Elements of Library Classification*. 2nd edition. Edited by B.I. Palmer. London: Asia Publishing House.
- [7] Reiner, Günter, Cumyn, Michelle, Hudon, Michèle and Mas, Sabine (2019). Designing a database to assist legal thinking: a new approach to indexing using facets. In: Erich Schweighofer, Franz Kummer, Saarenpää (eds.), *Internet of Things, Proceedings of the 22nd International Legal Informatics Symposium IRIS 2019*, p. 345-356.
- [8] Satija M. P. (2017). Colon Classification (CC). *Knowl. Org.* 44 (2017) No.4: 291- 307.
- [9] Schweighofer, Erich (2009). Learning and Verification of Legal Ontologies by Means of Conceptual Analysis. <https://www.researchgate.net/publication/242641747>
- [10] Scott, Mark and Smith, Nigel (2010). Legal Taxonomy from Sweet & Maxwell. *Legal Information Management* 10(3): 217-222.