

Contents

Special Issue:

New Trends for Knowledge Organization,

Guest Editor: Renato Rocha Souza

Articles

- Isidoro Gil-Leiva.
SISA—Automatic Indexing System for
Scientific Articles: Experiments with Location
Heuristics Rules Versus TF-IDF Rules 139
- Sandra Collovini de Abreu and Renata Vieira.
RelP: Portuguese Open Relation Extraction 163
- Maria Luiza de Almeida Campos and
Hagar Espanha Gomes.
Ontology: Several Theories on the
Representation of Knowledge Domains 178
- Fidelía Ibekwe-SanJuan and Geoffrey C. Bowker.
Implications of Big Data for Knowledge Organization..... 187

- Lígia Maria Arruda Café and Renato Rocha Souza.
Sentiment Analysis and Knowledge Organization:
An Overview of the International Literature 199

- Richard P. Smiraglia and Xin Cai.
Tracking the Evolution of Clustering,
Machine Learning, Automatic Indexing
and Automatic Classification in Knowledge
Organization 215

Reviews of Concepts in Knowledge Organization

- Daniel Martínez-Ávila.
Reader-interest Classifications:
An Alternative Arrangement for Libraries 234

- Books Recently Published** 247

Contents pages

Gil-Leiva, Isidoro. 2017. “SISA—Automatic Indexing System for Scientific Articles: Experiments with Location Heuristics Rules Versus TF-IDF Rules.” *Knowledge Organization* 44(3): 139-162. 123 references.

Abstract: Indexing is contextualized and a brief description is provided of some of the most used automatic indexing systems. We describe SISA, a system which uses location heuristics rules, statistical rules like term frequency (TF) or TF-IDF to obtain automatic or semi-automatic indexing, depending on the user's preference. The aim of this research is to ascertain which rules (location heuristics rules or TF-IDF rules) provide the best indexing terms. SISA is used to obtain the automatic indexing of 200 scientific articles on fruit growing written in Portuguese. It uses, on the one hand, location heuristics rules founded on the value of certain parts of the articles for indexing such as titles, abstracts, keywords, headings, first paragraph, conclusions and references and, on the other, TF-IDF rules. The indexing is then evaluated to ascertain retrieval performance through recall, precision and f-measure. Automatic indexing of the articles with location heuristics rules provided the best results with the evaluation measures.

Collovini de Abreu, Sandra and Renata Vieira. 2017. “RelP: Portuguese Open Relation Extraction.” *Knowledge Organization* 44(3): 163-177. 47 references.

Abstract: Natural language texts are valuable data sources in many human activities. NLP techniques are being widely used in order to help find the right information to specific needs. In this paper, we present one such technique: relation extraction from texts. This task aims at identifying and classifying semantic relations that occur between entities in a text. For example, the sentence “Roberto Marinho is the founder of Rede Globo” expresses a relation occurring between “Roberto Marinho” and “Rede Globo.” This work presents a system for Portuguese Open Relation Extraction, named RelP, which extracts any relation descriptor that describes an explicit relation between named entities in the organisation domain by applying the Conditional Random Fields. For implementing RelP, we define the representation scheme, features based on previous work, and a reference corpus. RelP achieved state of the art results for open relation extraction; the F-measure rate was around 60% between the named entities person, organisation and place. For better understanding of the output, we present a way for organizing the output from the mining of the extracted relation descriptors. This organization can be useful to classify relation types, to cluster the entities involved in a common relation and to populate datasets.

Campos, Maria Luiza de Almeida and Hagar Espanha Gomes. 2017. “Ontology: Several Theories on the Representation of Knowledge Domains.” *Knowledge Organization* 44(3): 178-186. 28 references

Abstract: Ontologies may be considered knowledge organization systems since the elements interact in a consistent conceptual structure. Theories of the representation of knowledge domains produce models that include definition, representation units, and semantic relationships that are essential for structuring such domain models. A realist viewpoint is proposed to enhance domain ontologies, as definitions provide structure that reveals not only ontological commitment but also relationships between unit representations.

Ibekwe-SanJuan, Fidelia and Geoffrey C. Bowker. 2017. “Implications of Big Data for Knowledge Organization.” *Knowledge Organization* 44(3): 187-198. 33 references.

Abstract: In this paper, we propose a high-level analysis of the implications of big data for knowledge organisation (KO) and knowledge organisation systems (KOSs). We confront the current debates within the KO community about the relevance of universal bibliographic classifications and the thesaurus in the web with the ongoing discussions about the epistemological and methodological assumptions underlying data-driven inquiry. In essence, big data will not remove the need for humanly-constructed KOSs. However, ongoing transformations in knowledge production processes entailed by big data and Web 2.0 put pressure on the KO community to rethink the standpoint from which KOSs are designed. Essentially, the field of KO needs to move from laying down the apodictic (that which we know for all time) to adapting to the new world of social and natural scientific knowledge by creating maximally flexible schemas—faceted rather than Aristotelean classifications. KO also needs to adapt to the changing nature of output in the social and natural sciences, to the extent that these in turn are being affected by the advent of big data. Theoretically, this entails a shift from purely universalist and normative top-down approaches to more descriptive bottom-up approaches that can be inclusive of diverse viewpoints. Methodologically, this means striking the right balance between two seemingly opposing modalities in designing KOSs: the necessity on the one hand to incorporate automated techniques and on the other, to solicit contributions from amateurs (crowdsourcing) via Web 2.0 platforms.

Café, Lígia Maria Arruda and Renato Rocha Souza. 2017. “Sentiment Analysis and Knowledge Organization: An Overview of the International Literature.” *Knowledge Organization* 44(3): 199-214. 32 references.

Abstract: Knowledge organization (KO) as an activity is, among other meanings, a process for conceptual modeling of knowledge domains that produces a consensual abstraction model of this domain with a particular purpose. It adopts a myriad of techniques to analyze and build efficient knowledge organization systems, and one of these techniques is called sentiment analysis (SA) or opinion mining, which is emerging as promising and useful in a variety of ways. It is based in NLP and AI algorithms, and aims at identifying opinions and emotions toward any person, organization or subject; evaluating them as positive or negative, in both binary and graded fashions. This study sought to show various aspects of the implementation of SA for knowledge organization tasks as registered in the scientific literature. We began with exploratory bibliographic research and built a corpus of 91 scientific papers, written in English, selected in the LISA Database, between 2000 to 2016. We analyzed these papers and extracted title, year of publication, author(s) and institution(s), title of the journal where they were published, keywords, the LISA classification code, methods/techniques adopted and its application areas. Our main findings are that theoretical papers still prevail, which may indicate a field in the early stages. We found many institutions and authors from Asia, which points to a new shift in world expertise. We concluded that SA is still a novelty in the KO field, being slowly adopted as an aid to the main tasks, as document classification.

Smiraglia, Richard P. and Xin Cai. 2017. “Tracking the Evolution of Clustering, Machine Learning, Automatic Indexing and Automatic Classification in Knowledge Organization.” *Knowledge Organization* 44(3): 215-233. 12 references.

Abstract: A very important extension of the traditional domain of knowledge organization (KO) arises from attempts to incorporate techniques devised in the computer science domain for automatic concept extraction and for grouping, categorizing, clustering and otherwise organizing knowledge using mechani-

cal means. Four specific terms have emerged to identify the most prevalent techniques: machine learning, clustering, automatic indexing, and automatic classification. Our study presents three domain analytical case analyses in search of answers. The first case relies on citations located using the ISKO-supported “Knowledge Organization Bibliography.” The second case relies on works in both Web of Science and SCOPUS. Case three applies co-word analysis and citation analysis to the contents of the papers in the present special issue. We observe scholars involved in “clustering” and “automatic classification” who share common thematic emphases. But we have found no coherence, no common activity and no social semantics. We have not found a research front, or a common teleology within the KO domain. We also have found a lively group of authors who have succeeded in submitting papers to this special issue, and their work quite interestingly aligns with the case studies we report. There is an emphasis on KO for information retrieval; there is much work on clustering (which involves conceptual points within texts) and automatic classification (which involves semantic groupings at the meta-document level).

Martínez-Ávila, Daniel. 2017. “Reader Interest Classifications: An Alternative Arrangement for Libraries.” *Knowledge Organization* 44(3): 234-246. 72 references.

Abstract: The concept of reader-interest classifications and its related terminology have shown a well-established presence and common characteristics in the knowledge organization literature for more than half a century. During the period 1952-1995, it was not unusual to find works, projects and discourses using a common core of characteristics and terms to refer to a recognizable type of projects involving alternative classifications to the *DDC* and other traditional practices in libraries. The use of reader-interest classification related terms and references drastically declined since 1995, although similar projects and characteristics are being used until the present day such as those of implementation of *BISAC* in American public libraries. The present paper attempts to overview the concept and terminology of reader-interest classifications in a historical perspective emphasizing the transformation of the concept and its remaining characteristics in time.