Domain Analysis of SIG/CR 2012

Richard P. Smiraglia
Information Organization Research Group
School of Information Studies
University of Wisconsin, Milwaukee

smiragli@uwm.edu

ABSTRACT
Domain analytic techniques are applied to the abstracts approved for the 2012 SIG/CR Classification Workshop. Results of preliminary metric analyses are described.

Keywords
SIG/CR, Domain-analysis, Classification, Knowledge-organization.

1. INTRODUCTION

The 2012 SIG/CR classification workshop was devoted to an evaluation of the "past and future prospects of SIG/CR," which was one of the group of special interest groups formed early in the history of ASIST. The structure of the workshop was diverse, incorporating papers, lightning talks, and doctoral submissions, as well as keynote presentations, in order to unearth the broadest possible range of research topics and methodologies currently being applied in classification research. The purpose of this presentation is to bring domain analytic tools to bear on the content of the 2012 workshop. Some simple metric techniques including keyword analysis and citation analysis were applied in advance to abstracts accepted for the workshop, then repeated on the final submissions.

2. TITLE KEYWORDS

Not including the keynote presentations, twenty-two abstracts were approved for inclusion in the workshop. Keywords in the titles of those papers were sorted by frequency of occurrence. Figure 1 shows the terms that were used more than once.
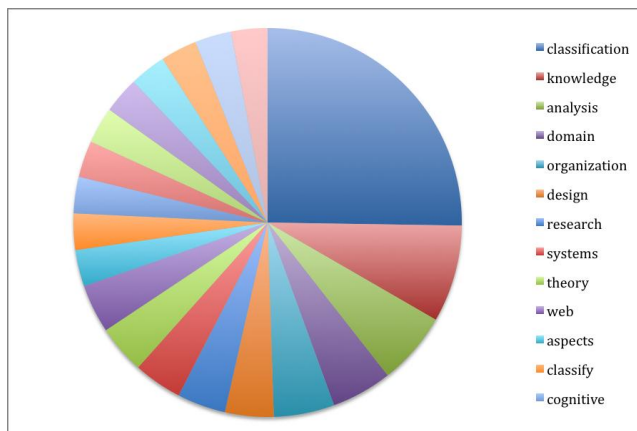
Figure 1. Title keywords

The word "classification" was used 25 times, "domain analysis" 8, "knowledge organization" 5, and the rest occurred 3 or 4 times. An MDS plot of these keywords was produced using WordStat™ software (stress = 0.2536; R-squared = 0.7915); this is shown in Figure 2.
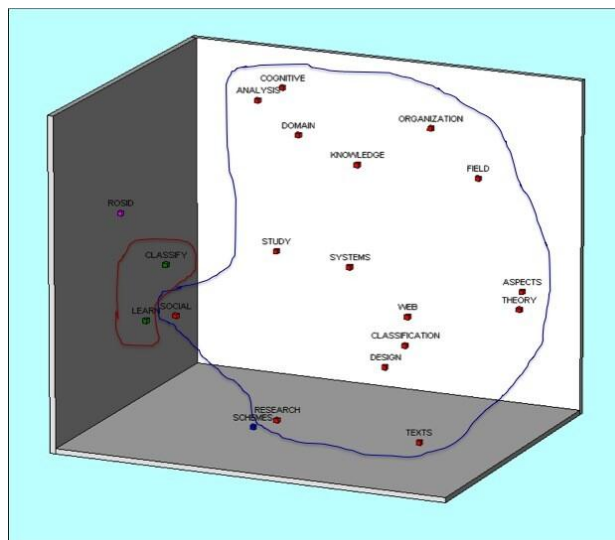

Figure 2. 3-dimensional plot of keywords

Low stress and relatively high R-squared suggest goodness of fit of the plot. There are essentially two regions, a small

cluster "learn" and "classify" in one, and nearly everything else in the other, which is a fair image of the granularity or breadth of intension in the domain. This larger cluster is spacious, suggesting clear delineation, or if one prefers, specificity. At the upper left is a sort of cluster formed from domain analysis and cognitive work analysis, for example. Knowledge organization seems to be hovering in the center at the top, anchoring the whole cluster. Similarly, classification is perhaps anchoring the cluster from the center but farther down. There is some suggestion of motion of the research front from classical "classification" toward "domain analysis."

The high frequency of usage of "classification" is not a surprise, and suggests the common unity of the workshop's participants, but the granularity in occurrence of other terms illustrates the thematic breadth of the workshop. A preliminary conclusion is that the participants see classification as central to their mission, but that a wide variety of approaches are considered acceptable, and perhaps even compatible.

3. CITATIONS

Citation practices were quite mixed in the early submissions, ranging from formal papers with full reference lists to short abstracts with no references. The "final complete draft" submissions were more formal in nature, although there was no standard approach to citation formatting, which made manual indexing problematic. 234 citations were harvested from the 22 submissions in the proceedings. The number of references present ranged from 0 to 23 with a mean of 7.8. Usually references are in one style, and thus can be parsed to extract the dates of publication, which is a useful indicator of obsolescence or absorption in a domain. Dates of citation ranged from 1771 to 2012. The mean age of citation was 15.4 years, ranging from 0 (for works published in 2012) to 240, although the majority of works cited were in the range from 0 to 60 years. Most citations (72%) were to works within the current 15 years. This age of citation is consistent with a social scientific domain.

4. HIGHLY CITED AUTHORS

The citations were arrayed by first author and sorted by frequency of occurrence. 31 authors were cited more than once, and these names appear in Figure 3.
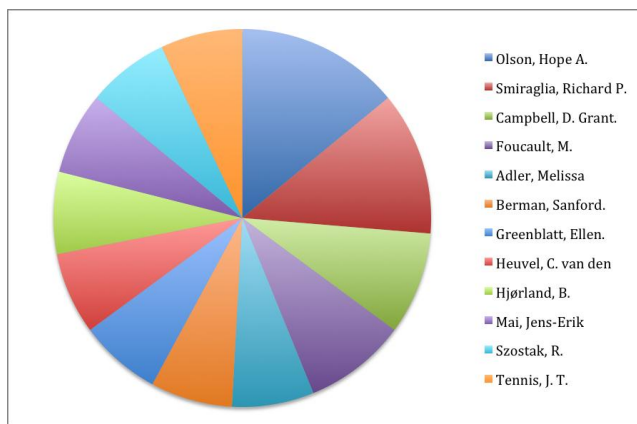

Figure 3. Authors cited more than once

There is, again, a fair amount of breadth. There is also much self-citation by the workshop participants. Self-citation is often eliminated from bibliometric analyses for a variety of reasons. But in an emerging scientific domain it is to be expected that researchers on the cutting edge might have only their own prior papers to cite, especially in a research forum such as SIG/CR. Be that as it may, likely, the core of the workshop as a domain is represented by the names of authors cited 3 or more times, which have been highlighted here. Mostly, the names in the highlighted region are consistent with the names usually cited in domain analytic studies of knowledge-organization (see Smiraglia 2012).

However, it was interesting to observe the ways in which this list differs from the lists that usually emerge from domain analyses of Knowledge Organization. Typically those lists have Hjørland at the top near names such as Dahlberg and Beghtol. Here we have rather a clearer reflection of the actual work of the SIG/CR Workshop participants. Attempted author co-citation within the 22 presentations was not successful because there were too few co-citations to run the SPSS procedure, which in itself says something about the granularity of the research presented here. That is, in these 22 papers there is not much of a common theoretical foundation. But there is a newly distinct cluster, "Olson Adler Berman Greenblatt." So we can say that at least for this SIG/CR Workshop, a relatively new body of work is apparent. Campbell, Hjørland and Mai also appear, but do not form a cluster and are not much cited by the Olson et al. cluster.

5. CONCLUSION

The SIG/CR workshop shows unity around the concept of classification, but great granularity (or breadth, depending on point of view) otherwise. That is a sign of a healthy, evolving domain. There is a wide range of obsolescence in the materials cited by these authors, although the majority

align with a social-scientific domain and seem to be citing materials produced within the current decade or so. The workshop participants are bringing current research to the workshop, which is one cause for the high level of self-citation among the participants. Otherwise, a fairly narrow band of core cited-authors, consistent with parameters recognized in knowledge-organization, are seen as representative. "Classification," and "domain analysis, cognitive, theory, design, and information systems" are themes represented in the titles of workshop presentations.

This list of terms is consistent with work by the core cited-authors.

REFERENCE

Smiraglia, Richard P. 2012. Universes, dimensions, domains, intensions and extensions: knowledge organization for the 21st century. In Neelameghan, A. and Raghavan, K.S. Categories, relations and contexts in knowledge organization: Proceedings of the Twelfth International ISKO Conference, Mysore, India, August 6-9, 2012. Würzburg: Ergon Verlag, pp. 1-7.

The columns on the last page should be of approximately equal length.