

Producing Navigable Knowledge Organization with Knowledge Interaction

Richard P. Smiraglia and Jihee Beak, University of Wisconsin, Milwaukee

Abstract: Knowledge interaction combines concept theory, instantiation theory, and the taxonomy of knowledge elements to suggest that knowledge organization systems might be used effectively to integrate different dimensional representations of information objects. Understanding knowledge structurally as well as semantically can lead to a variety of implementations that might provide temporal interfaces for understanding relationships among information objects that are not obviously semantically related. An experimental test-bed would rely on the actual experience of working scholars. Preliminary results come from a case study of the works of one prolific New Testament scholar whose works are available in digital form. We see clearly the distance between the theological positions, sociological interpretive positions, and methodological positions, indicating three interacting intellectual poles in this scholar's writing.

1.0 Knowledge interaction for interfacing systems

Knowledge interaction is the theory that combines concept theory, instantiation theory, and the taxonomy of knowledge elements to suggest that faceted knowledge organization systems might be used effectively to integrate different dimensional representations of information objects (Smiraglia and van den Heuvel 2013). This approach to understanding knowledge structurally as well as semantically can lead to a variety of implementations that might provide temporal interfaces for understanding relationships among information objects that are not obviously semantically related. Using the interplay between syntax and semantics, it is possible to expose points of interfaces among such objects that might usefully be mapped (Smiraglia, van den Heuvel and Dousa. 2011). An interesting illustration of the potential power of applications built from this perspective is the chain of associated multimedia objects associated with Barnett Newman's *Who is afraid of Red, Yellow and Blue III* reported by van den Heuvel and Smiraglia (2013). The capability of faceted classification to be used to create navigable routes through structural bibliographical elements has been demonstrated in three recent papers (Smiraglia 2014a, b and c).

Producing workable knowledge organization interfaces based on new theoretical knowledge such as that summarized above requires a fair amount of experimental development. Little is known about how to search for, or identify with high precision, critically important non-semantic similarities. Computer science and its allied domains have developed experimental techniques for concept extraction and topic modeling such as latent semantic analysis, latent dirichlet analysis, syntactic pattern recognition—which can be characterized as attempts to use patterns among terms occurring in documents (see for example Flasiński and Jurek 2014). Likewise, computational linguists have developed domain ontology engineering tools (e.g., Navigli and Velardi 2004). All of these approaches require parallel or subsequent semantic interpretation. The case described here is an attempt to use the domain analytical tools from

the knowledge organization domain to provide a triangulated data set for the purpose of providing scholar-based semantic interpretation.

What is needed for knowledge discovery is a system that can prompt a searcher for different ways to combine non-obvious modalities to retrieve efficiently unexpected and meaningful results. One potential direction would be the development of a device for generating unknown and new connections that cannot be found by existing search engines. Such a system would be based on the ability to switch between use of keywords and pattern or structure recognition. To some extent, an experimental test-bed would rely on the actual experience of working scholars. Our objective is to build a test-bed by generating structural ontologies based on concept, idea, and topic extraction, and on mapping language onto structural relations that emerge from the ontologies. This research presents preliminary results from such a study.

2.0 New Testament scholarship as a test case

Domain analytical tools are used in knowledge organization to discover the knowledge base common to groups of people—everything ranging from a corporate office to an academic department to an entire field of study. The same tools have been used only sparingly to analyze the knowledge base of a single individual; Hartel (2003 and 2010) used qualitative methods to study hobbyists and gourmet home cooks. In this case study, we will look at a selected portion of the published writing of one scholar to see what we can learn about the operative knowledge base of a single individual. We are limited by chronology—we will see in this case only work published in the past and thus not work in progress.

For the present analysis we have chosen one prolific New Testament scholar whose works are available in digital form. We want to know what sort of structural ontology might represent this one scholar—a potential user of our future searching platform. We will use content analysis software to extract keywords, terms and phrases as exemplars of the concepts, topics and potential research questions that emerge from a set of writing.

Deirde Good is professor of New Testament at The General Theological Seminary of the Episcopal Church, New York, New York. Professor Good's research has foci in synoptic gospels, Christian origins, noncanonical writings and biblical languages, according to her official biography on the GTS website (http://www.gts.edu/index.php?option=com_workforce&view=employee&id=16&Itemid=108). In fact, she is a scholar engaged in analysis of texts from the collection of gnostic scrolls found at Nag Hammadi, as well as studying the roles of gender and sexuality in what are known as hospitality texts from the New Testament. For this case study we gathered texts of sixteen of the fifty-nine works listed in her abbreviated c.v., digitized them, and submitted them to co-word analysis using Provalis' QDA Miner and WordStat software to see whether we could discover thematic patterns in the terms used in her texts. The texts analyzed are listed in Table 1.

Title	Date
Sophia in Valentinianism	1984
Sophia in Egnostos The Blessed (NHC III, 4 and V,I) and the Sophia of Jesus Christ (NHC III, 3 and BG 8502 ,3)	1990
Divine Noetic Faculties in Egnostos the Blessed and Related Documents	1986
The verb ANAXOREO (to withdraw) in Matthew's Gospel	1990
Rhetoric and Wisdom in Theophilus of Antioch	1991
The Canaanite Woman: Patristic Exegesis of Matthew 15:21-28	1991
The Dead Sea Scrolls in Current Scholarship	1994
Moral Dualism and Virtues in Matthew's Gospel	1997
Sacramental Language and Verbs of Creating, Generating and Begetting in the Gospel of Philip	1997
Prayer of the Apostle Paul	1997
Reading Strategies in Biblical Passages on Same-Sex Relations	1997
The New Testament and Homosexuality: Are We Getting Anywhere?	2000
Why it is inaccurate and unprofitable to argue that the Bible condemns homosexuality	2001
Bible, New Testament	2008
A theology of marriage including same-sex couples: a view from the liberals	2011
The liberal response	2011

Table 1. Texts analyzed

An initial term frequency analysis reveals 8267 individual keywords used in the core texts. 745 words each occurred more than 10 times. Curiously, “gnostic” occurred only 49 times and thus does not fall into the present data analysis. Table 2 shows are the forty-two words that occur most frequently in the corpus.

<i>Jesus</i>	236	1.30%	<i>World</i>	82	0.40%
<i>Matthew</i>	187	1.00%	<i>Meekness</i>	80	0.40%
<i>God</i>	182	1.00%	<i>Scrolls</i>	79	0.40%
<i>Text</i>	163	0.90%	<i>Virtues</i>	79	0.40%
<i>Sophia</i>	156	0.80%	<i>Verb</i>	78	0.40%
<i>Gospel</i>	151	0.80%	<i>Phil</i>	77	0.40%
<i>Christian</i>	145	0.80%	<i>Body</i>	73	0.40%
<i>Marriage</i>	134	0.70%	<i>Gos</i>	73	0.40%
<i>Paul</i>	131	0.70%	<i>Context</i>	72	0.40%
<i>Wisdom</i>	130	0.70%	<i>Homosexuality</i>	70	0.40%
<i>Church</i>	121	0.60%	<i>Passage</i>	70	0.40%

<i>Community</i>	118	0.60%	<i>Life</i>	69	0.40%
<i>Sex</i>	114	0.60%	<i>Testament</i>	69	0.40%
<i>Texts</i>	112	0.60%	<i>Bible</i>	68	0.40%
<i>Greek</i>	111	0.60%	<i>Women</i>	68	0.40%
<i>Sexual</i>	103	0.50%	<i>Word</i>	68	0.40%
<i>Christ</i>	97	0.50%	<i>Account</i>	66	0.40%
<i>Male</i>	93	0.50%	<i>Dead</i>	66	0.40%
<i>Female</i>	92	0.50%	<i>Father</i>	66	0.40%
<i>Biblical</i>	91	0.50%			
<i>Spirit</i>	90	0.50%			

Table 2. Words occurring across the entire corpus

The software's keyword-in-context feature was used to detect terms in which each of these keywords appears; a chart of these terms was then created. For example, "God's redemptive work," "Christian tradition," and "Greek verb" are all frequently occurring terms in the corpus. The software also has a phrase detection capability, which gathers phrases according to parameters set by the investigators; we extracted 2-5 word phrases that appear in the corpus more than twice, which yielded 512 phrases. These numbers are an indicator of the complexity of the task, of course. Just as an illustration, we reproduce in Table 3 the phrases appearing 20 or more times in the corpus.

Gospel of Philip (Gos Phil)	101
Dead sea	56
New testament	54
Dead sea scrolls	45
Bridal chamber	35
Moral dualism	32
Early christian	29
Male and female	27
Nag hammadi	21
Sex relations	21
Sexual relations	21
Moral dualism and virtues	20

Table 3. Phrases occurring more than 20 times

Keywords, terms and phrases—as defined here—all provide indicators of concepts and themes in the researcher's corpus. For instance, there are many theological terms (Jesus himself, God's redemptive work) but there also are many methodological cues (Greek verbs, Christian texts, Co-occurrence

processes can be performed on each of these. For example, figures 1-3 show MDS plots of the most-frequently occurring keywords, terms and phrases.

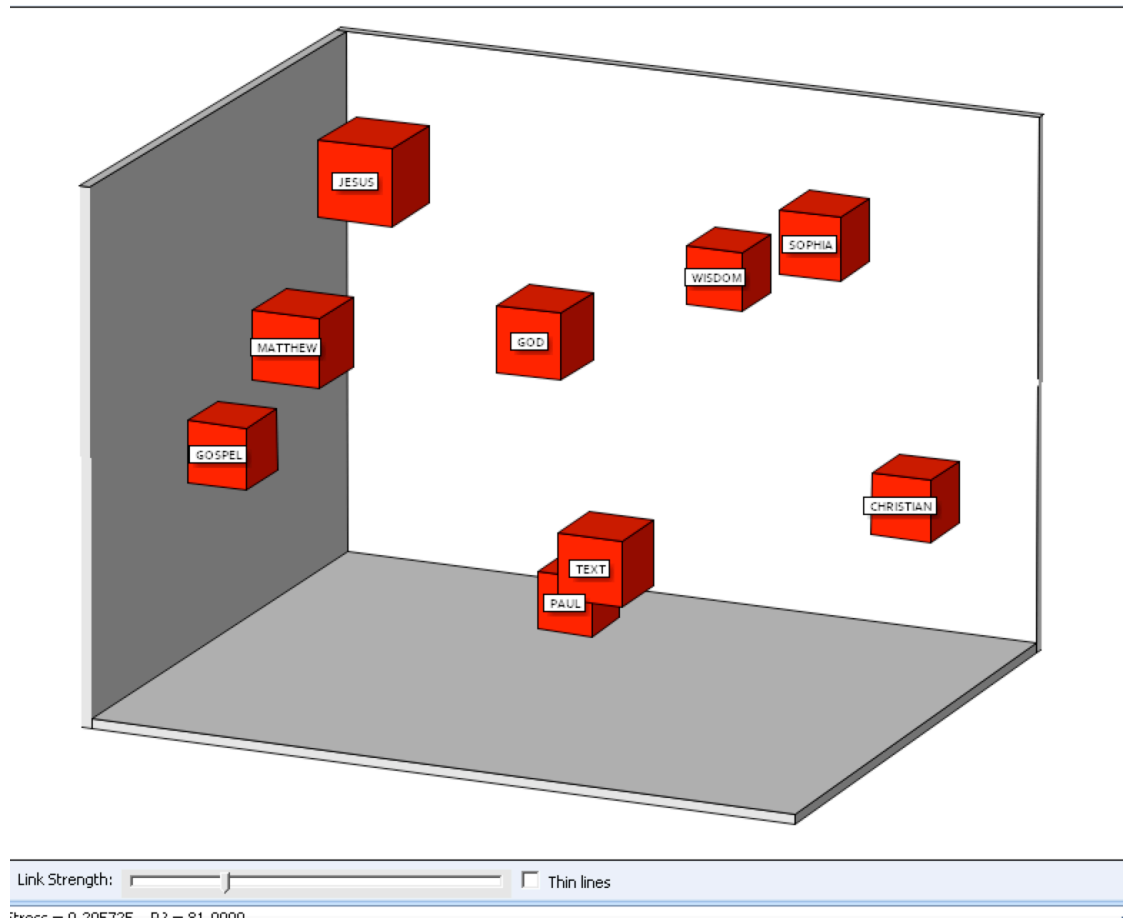


Figure 1. Keyword MDS Plot

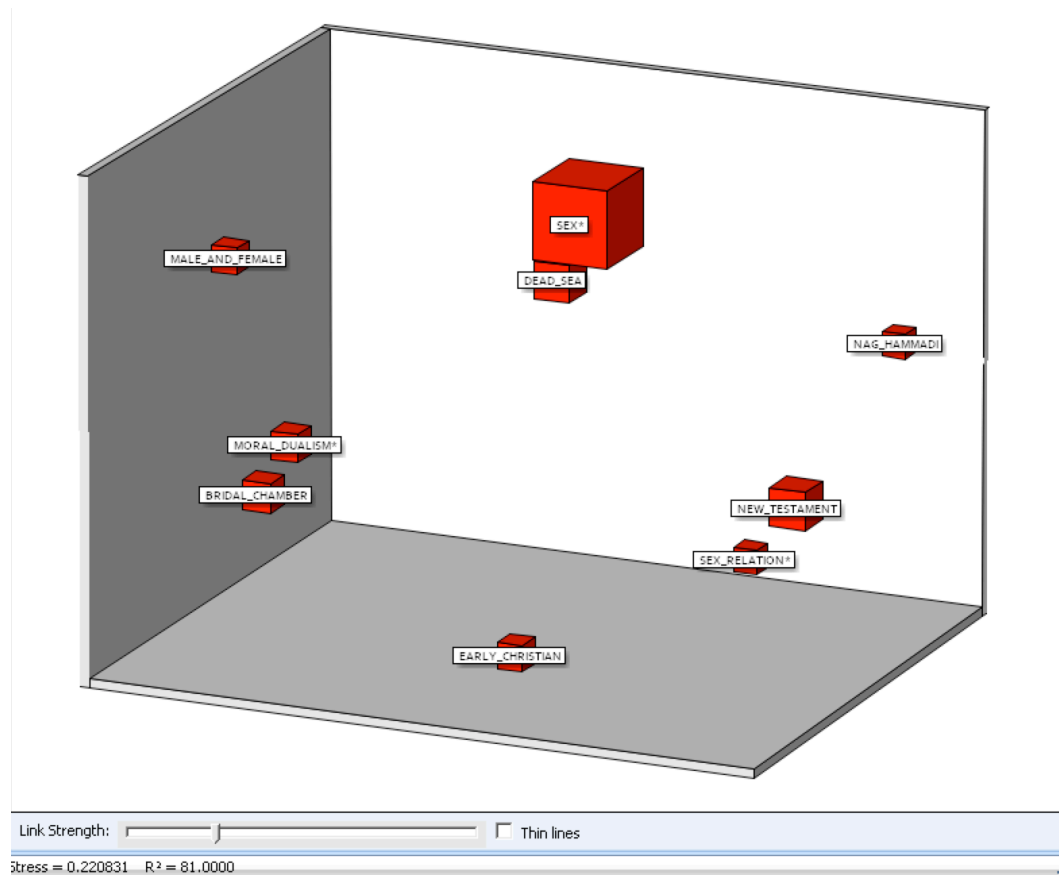


Figure 2. Term MDS plot.

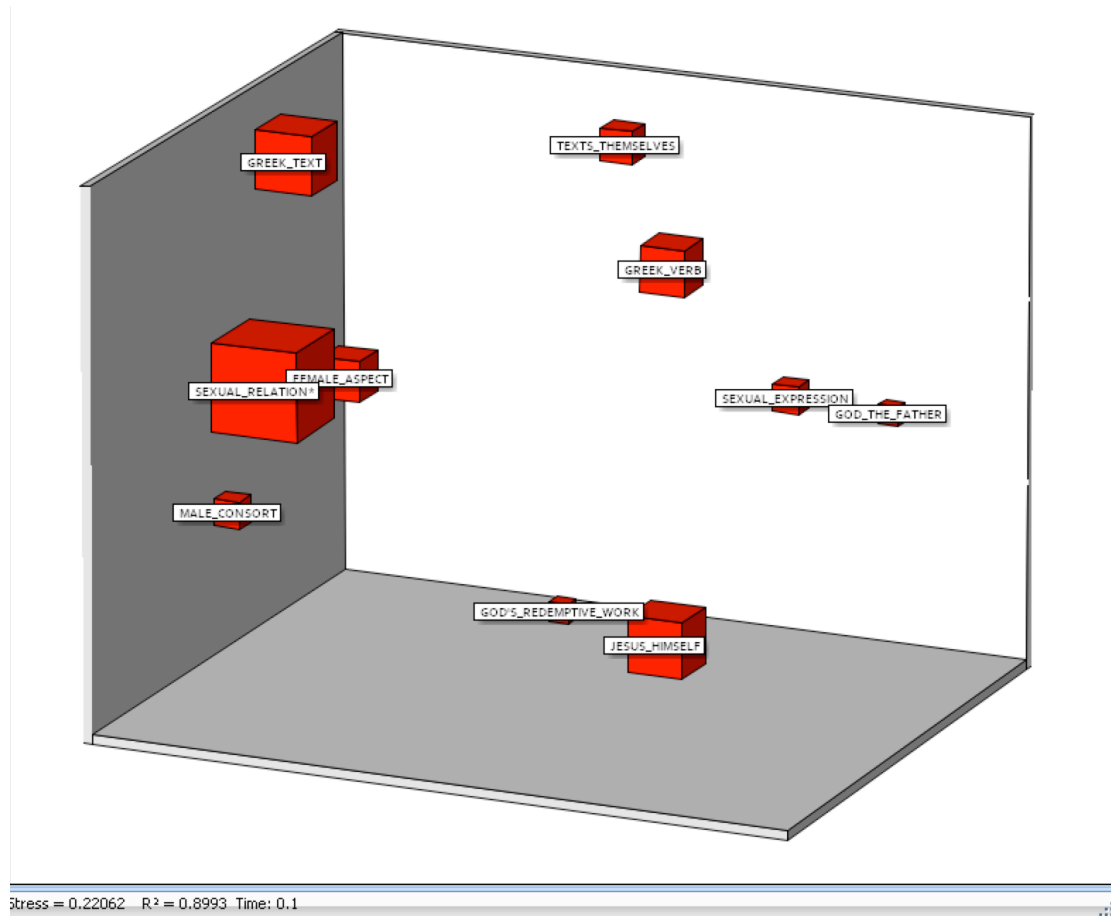


Figure 3. Phrase MDS plot.

What we see in all three figures but most clearly in Figure 3 is the distance between the theological positions, sociological interpretive positions, and methodological positions. These very limited results from a large case study of a single scholar's corpus have easily demonstrated three interacting intellectual poles in this scholar's writing. Of course, even the limited data in this case study can reveal much richer details. Producing a knowledge organization interface, perhaps in the form of a visual map could allow the system to learn a scholar's domain in order to present not only summaries but also potential next steps.

References

- Flasiński, Mariusz and Janusz Jurek. 2014. Fundamental methodological issues of syntactic pattern recognition. *Pattern analysis and applications* 17: 465-80. <http://dx.doi.org/10.1007/s10044-013-0322-1>.
- Hartel, Jenna. 2003. The serious leisure frontier in library and information science: hobby domains. *Knowledge organization* 30: 228-38.

- Hartel, Jenna. 2010. Managing documents at home for serious leisure: a case study of the hobby of gourmet cooking. *Journal of documentation* 66: 847-74.
- Navigli, Roberto, and Paola Velardi. 2004. Learning domain ontologies from document warehouses and dedicated web sites. *Computational Linguistics* 30: 151-179.
- Smiraglia, Richard P. 2014a. Big Classification: using the empirical power of classification Interaction. In *Advances in classification research online* 24(1),doi:10.7152/acro.v24i1.14673
- Smiraglia, Richard P. 2014b. Classification interaction demonstrated empirically. In Wiesław Babik, ed. *Knowledge organization in the 21st century: between historical patterns and future prospects: Proceedings of the Thirteenth International ISKO Conference 19-22 May 2014, Kraków, Poland*. Advances in knowledge organization 14. Würzburg: Ergon Verlag, pp. 176-83.
- Smiraglia, Richard P. 2014c. Extending the visualization of classification interaction with semantic associations. SIG/CR proceedings. Forthcoming.
- Smiraglia, Richard P. and Charles van den Heuvel. 2013. Classifications and concepts: towards an elementary theory of knowledge interaction. *Journal of documentation* 69: 360-83.
- Smiraglia, Richard P., Charles van den Heuvel Thomas Dousa. 2011. Interactions between elementary structures in universes of knowledge. In Slavic, Aida and Civallero, Edgardo eds., *Classification & Ontology: Formal Approaches and Access to Knowledge: Proceedings of the International UDC Seminar 19-20 September 2011, The Hague, Netherlands*. Würzburg: Ergon Verlag, 2011, pp. 25-40.
- van den Heuvel, Charles and Richard Smiraglia. 2013. Visualizing knowledge interaction in the multiverse of knowledge. In Aida Slavic, Almila Akdag Salah and Sylvie Davies eds., *Classification & visualization, interfaces to knowledge: Proceedings of the International UDC Seminary 24-25 October 2013, The Hague, Netherlands*. Würzburg: Ergon Verlag, pp. 59-72.