

headings, because of what we know about cataloging practices. It also seems logical that names as subjects are not associated with other kinds of subject headings; typically biographical works might not have many other headings assigned to them. Similarly, names as subject headings are not associated with the other detailed bibliographic characteristics. What we know about the associations of UDC classes with auxiliary operators or other classes also makes sense, given the age of the population of classified works and the population of the classification in general.

When we looked for associations between the UDC components and the bibliographic characteristics we also found logical associations occurring as statistically significant. Works in class 8 “Languages and literatures” seem to be associated with richer subject vocabulary, represented here by a more complex mix of subject headings as well as the simple relation operator (to express phase relations) and the consecutive extension operator (to express broader subject areas). Class 0 “Generalities” contains many reference works, thus it is associated with uncontrolled index terms that likely are more specific, genre terms, the simple relation operator and linked classes. Classes 6 and 7 are the sciences and the arts, and thus are associated with linked classes, as well as names as subject headings and the consecutive extension operator. Class 1 “Philosophy and psychology” is associated with names as subjects.

All of the associations enumerated here could be considered footprints of sorts, which is to say they represent ways in which associations among classified bibliographic entities are reliably present and predictable. The associations themselves could be classified. In any case, these associations demonstrate the power of classification for indicating nonlinear pathways through a collection of bibliographic entities.

An artifact of this study (or it could be termed a limitation) is the confidence level of 95%, which therefore rendered other likely associations not statistically significant. If we were to reprise table 10 here and this time shade the cells that would have been statistically significant at 90% confidence we get the result shown in table 11.

Now we have a slightly richer picture, with a few more sensible associations. This suggests that a slightly larger sample might have yielded this many more statistically-significant correlations, and therefore, predictable, navigable footprints. Also, it is important to consider what we learn from all of the blank cells where no associations were apparent. This tells us that those characteristics are ubiquitous and therefore predictable because they are always present across all classes to some degree.

5.0 CONCLUSIONS: EMPIRICAL POWER OF CLASSIFICATION INTERACTION

With this simple preliminary investigation we have seen a first glimpse of the empirical power of classification interaction. The shape of the correlations we observed aligns with the population of the UDC across the bibliographic domain represented. Predictable footprints demonstrated here could be used as pathways for navigation across the structural relationships shared by works in differing domains, in addition to traditional author and title navigation of known item searching or gatherings of like subjects in traditional classification. In this sense the UDC is shown to represent Beghtol’s notion of a classification as cultural artifact, because it both gathers works together and facilitates navigation across their bibliographic characteristics. In other words, the classification here is a footprint, of sorts, of the works it classifies.

At another level we can consider all of the results of this study to be hypotheses for further research. A limitation of this study is the post-1979 chronology of the UDC numbers assigned to works in the WorldCat. The majority of the nine million UDC numbers in the sample under analysis represent works cataloged in the last quarter century. This means that the UDC practices and the population of the classification revealed here does not represent the whole history of the UDC, but just a recent snapshot. The Knowledge Space Lab team also received output of all UDC numbers from the online catalog of Catholic University Leuven (KU Leuven) in order to compare the population of the UDC in a specific library to that found in the WorldCat. An obvious next step for this research is to repeat the analyses using the data from KU Leuven for comparison.

In this study only content-designated elements were analyzed for correlation, and not any of the values of any of those elements. In other words, another obvious next step is to look at individual names, works, subject terms, and so forth, to see what more detailed network of associations might be present in the classified bibliographic universe, and whether those associations are predictable. The technique of identifying statistically-significant correlations among nominal level variables also has the limitation that no direction or causality can be demonstrated. All we can say at this point is that we see certain elements occurring in conjunction with each other in this data set.

However, it is already clear from this simple study that in a classified bibliographic dataset predictable pathways of association exist through the data. The raw data associated with the elements in table 11 were entered into a matrix to generate a network diagram of the associations uncovered in this research. Figure 3 shows the network.

