

David Jank
Long Island University, Brookville, NY

Toward a Unifying Ontology for Human-Information Interaction

Abstract: Research agendas in human-information interaction (HII) are varied and divergent. The interdisciplinary lens of information studies offers a convergent view of HII scholarship. The purpose of this paper is to taxonomically document the divergent scholarship in human-information interaction, and construct a unifying ontology of HII discourse, using bibliometric techniques.

Résumé : Les programmes de recherche en matière d'interaction homme-information sont variés et divergents. L'interdisciplinarité des sciences de l'information offre toutefois un portrait convergent de la mission professorale en matière d'interaction homme-information. L'objectif de la recherche est de documenter du point de vue taxinomique les divergences de la mission d'enseignement de l'interaction homme-information et de construire une ontologie unifiée du discours au moyen de techniques bibliométriques.

Epistemological views of human-information interaction (HII) can be seen as highly interdisciplinary. While scholarship in information science offers empirical documentation of HII in action, the cognate disciplines that inform this study are many, and not as highly documented. Among the primary cognate fields of scholarship documented in the current study as contributive to scholarship on the interaction between humans and information are four subfields of psychology (neurology, cognition, behavioral analysis, and social interaction theory), information science, librarianship, computer science, medicine, human-computer interaction, and interactive learning. These interdisciplinary components of HII seem to require consideration of this field from a broad ontological spectrum. The purpose of this paper is to highlight many of the findings of my dissertation research. These findings constitute the current study, where the interaction between humans and information from each of these epistemological viewpoints is examined. These results include the construction of a unifying ontology of HII based upon both qualitative and quantitative examination of bibliometric analyses performed on multidisciplinary scholarship published from 1996 to 2008.

An exploratory study was first conducted across all databases in both the Dialog Information Services system, and the federated platform of the ISI Web of Knowledge. The pilot study included axially coded search statements based upon validated terminology relating to HII and selected from both controlled vocabulary listings and online data dictionaries. The pilot study identified more than 24,000 pieces of scholarship, published during the time period specified, from the social sciences, medicine, computer science, business, education, and the humanities disciplines that utilized this terminology in descriptor, abstract, and title fields. These data were downloaded into Excel spreadsheet files and subsequently processed via QDA Data Miner software in order to ascertain prominent themes in scholarly discourse. Data mining was also utilized to identify leading scholars, co-citation behavior, publishing patterns, prominent research agendas, and a variety of descriptive bibliographic characteristics. The purpose of this paper is to document the construction and validation testing of the HII ontology, beginning from the collection of pilot study results.

The purpose of the bibliometric analyses conducted in the pilot study was to identify those academic disciplines with substantive scholarly discourse in the areas of research cognate to HII, as well as to document the ontological characteristics of that discourse. Data were sorted and ranked by usage and discipline. Frequency distribution data were utilized to determine prominent HII discursive terminology, research agendas, and domain applications of this scholarship. These data were also used to identify the most appropriate databases, across all Dialog and ISI offerings, in which to conduct more detailed analyses for the current dissertation research. Both a taxonomy and concept map of HII were constructed based upon the pilot study findings. The HII taxonomy was utilized in order to design more precise search strategies for the current study.

Search strategy design for the current study was based on facet analytic techniques, by which the taxonomic components of HII were combined with terminology from published thesauri for the most prominent scholarly disciplines contributive to HII discourse. The search strategy for the current study utilized field-delimited searching, based upon facet analysis of the most prominently ranked data in the pilot study. In order to reflect the use of emergent terminology that may not yet be reflected in specific thesauri, free text searching was also utilized based upon term ranking from the bibliographic analyses in the pilot study. Online searching was conducted on 35 Dialog files that reflected prominent indexing of taxonomic terms, as well as on the federated ISI platform “Web of Knowledge,” which encompasses the basic sciences, the social sciences, and the arts and humanities.

The current study identified 15,292 potential pieces of published HII scholarship over the 12-year period, based upon prominent usage of these terms. Random sample size calculations were utilized, resulting in a sample set size $n = 669$. Domain filtering techniques (McCain 1995) were utilized in conjunction with random number generator lists in order to build the sample set. All citations in the sample set were reviewed manually following validated methods of qualitative analysis (Sugimoto et al. 2008; White and McCain 1998). These records were also downloaded into spreadsheet software, and bibliometric analysis was conducted on all fields in each record, utilizing QDA data mining and WordStat content analysis software. Multidimensional scaling techniques were employed to determine word rankings and co-occurrence coefficients. Additionally, text data for all fields in the bibliographic records were filtered through automated axial coding parameters, utilizing online data dictionaries, to determine prominent thematic discourse of leading scholars and research agendas of the discourse community.

The study identified five “ontological families” comprising the scholarly discourse of human-information interaction. These were: neurological health, cognitive function, human information processing and use, social interaction, and interactive learning. In order to determine levels of significance for contributive fields of scholarship, analysis of variance was conducted on research findings for all scholarly disciplines contributive to this scholarship, utilizing Chi Square analysis of bibliometric data. Goodness of fit testing was utilized with qualitative coding to determine the representativeness of study findings to the entire body of HII scholarship. Domain analytic axes (Tennis 2003) and Jaccard coefficient measures (at $J \geq .50$) were also employed to determine shared ontological characteristics of these data across disciplines.

A unifying ontology and corresponding quadrant displays were constructed. ANOVA calculations revealed that 12 scholarly fields of research were uniquely contributive to interdisciplinary HII discourse. Goodness of fit testing utilizing stress values and *Pearson R* calculations for all multidimensional coding ranged from $S=.09$ to $S=.21$, while correlation values ranged from $R^2=.78$ to $R^2=.93$. At 95% confidence levels, and confidence intervals ranging from $CI=.04$ through $CI=.06$, the final ontology illustrates that human-information interaction scholarship is founded on cross-disciplinary collaborative discourse, is highly thematic in terms of terminology use and research agendas across disciplines, is largely international in scope, and is published primarily in interdisciplinary scholarly journals. This paper includes HII ontological breakdowns for thematic terminology use, Euclidean distance modeling for graphing the ontology, quadrant displays for domain analytical axes, and various conceptual representations of the published scholarship on human-information interaction.

References

- McCain, Katherine W. 1995. Biotechnology in context: A database-filtering approach to identifying core and productive non-core journals supporting multidisciplinary R&D. *Journal of the American Society for Information Science*, 46: 306-317.
- Sugimoto, Cassidy R., Jean A. Pratt, and Karma Hauser. 2008. Using field cocitation analysis to assess reciprocal and shared impact of LIS/MIS fields. *Journal of the American Society for Information Science and Technology*, 59: 1441-1453.
- Tennis, Joseph T. 2003. Two axes of domains for domain analysis. *Knowledge Organization*, 30: 191-195.
- White, Howard D., and Katherine W. McCain. 1998. Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49: 327-355.