

distinctions. Second, the categories are somewhat vague. For instance, the distinction between cognitive-based exclusive and nonexclusive is not really clear: After all, one could declare all processes that have to do with Data (or Information or Knowledge) processing as *cognitive*. This is supposed to distinguish between the approaches of “Højrlund, which defines information in terms of biological mechanisms and signals, while Poli defines information in terms of signs and meanings” (Zins, 2007, p. 488). But one could argue that signals, signs, and meanings are all cognitive concepts, and that biological mechanisms can implement cognitive processes. Third, the categories do not clearly distinguish among approaches. For instance, the propositional versus nonpropositional distinction has to do with how the approaches consider Knowledge, but it tells us nothing about how they see Data and Information.

One could call Zins’s (2007) analysis *qualitative* and mine *quantitative*. In my analysis, I take the opposite route from Zins: Although he asserts that “words can be misleading”, (p. 487) I assume that most of the time they are not; after all, the messages we use to communicate are composed with words, and in the case of a questionnaire, words are all we have. Thus, my first assumption is that *which* words are used is significant. I take the answers at face value: I simply count word appearances, and even my semantic analysis is based on the appearance of certain syntactic structures (negation, and so on). My results are of a very different nature, as I do not classify individual answers in categories. Rather, I try to find salient features at the aggregate level. Thus, a second assumption is that aggregating the answers, while making important individual differences disappear, is the best way to obtain the voice of community. And although I agree with Zins (2007, p. 487) that “definitions are theory-laden,” the question is exactly what (kind of) theory they call for (or rely upon). In this sense, the IR analysis may be criticized by ignoring too much. The context of a word is indeed important, as well as the ways that words relate to each other.<sup>4</sup> I have attempted to remedy the grossest distortion with my analysis of negative context (i.e., when a word *x* is used to state that “data|information|knowledge is not *x*”). Arguably, such analysis is necessary but not sufficient to provide context and depth of interpretation. However, by letting raw data (counts of occurrences) be my guide, I am minimizing the use of conceptual categories, which are chosen depending on one’s conceptual framework.

It is interesting that there are points of agreement with Zins’s (2007) evaluation (e.g., his assessment of the human-centeredness of most definitions), but I simply have nothing to say about his other categories, as I have found no evidence for or against them. Whether this means that such categories are an artifact of Zins’s (2007) framework, not supported by the answers themselves, or that my analysis is too literal and lacks the depth to get at the meaning behind the statements, depends on one’s viewpoint. I certainly do

*not* argue that the analysis presented here, is more objective or “better” in any sense. The value one gives to Zins’s analysis depends considerably on how much one agrees with his categories, and the use he makes of them. An obvious question is whether establishing a conceptual framework must commit us so much as to accept such choices, or whether it is possible to carry out another analysis that leaves the given definitions to “speak for themselves” as much as possible. This is what I have attempted in this article.

## Conclusions and Further Research

I have carried out a text analysis of the article by Zins (2007). I analyzed the terms used in experts’ definitions of the concepts of Data, Information, and Knowledge. My analysis uses only standard IR and IE tools; therefore, it is guided by the data (words in the text), as opposed to by any *a priori* theory.

Because my approach is so different from Zins’s (2007), my results can be seen as complementing and enriching Zins’s presentation. Also, by proposing an alternate analysis, I hope to encourage readers to consider different possible approaches (their assumptions, strengths, and drawbacks) to defining these and other concepts in the field.

I must make clear that I do not consider this to be the definitive analysis of the data gathered by Zins (2007). There are clear limitations to this type of analysis. I consider it a starting point from which a more nuanced view can be developed by careful consideration of the text’s meaning. Unfortunately, such process is difficult and can easily lead to subjective judgments. On the other hand, the analysis presented in this article is guided entirely by the data and makes minimal assumptions; therefore, it can serve as an ancillary point for further, deeper analysis. In this sense, this work constitutes simply a starting point. It would have been tempting to use more sophisticated tools for data analysis. For instance, I could have used one of several clustering algorithms to determine whether the concepts defined are indeed disjointed (i.e., to check whether I obtain three distinct clusters when analyzing the whole text). But clustering algorithms have built-in assumptions, in that they all require some definition of *similarity* (or, inversely, distance) among data points (Han, Kamber, & Pei, 2011). Thus, I have left such analysis for further work, perhaps when more data (see later) are available.

Because the original data come from a questionnaire that was deployed in 2003–2005 and reached what would be considered today as a small number of participants, two immediate questions come to mind: If repeated today with the same participants, would the results change? In other words, would the experts who participated in the original survey still hold the same views? And if a new survey similar to the original one were given out today, but tried to reach a much wider audience (involving, perhaps, experts from not only information science, but IR, computer science, economics, philosophy, and others), would these results hold?

<sup>4</sup>The old joke is that, by ignoring word order, for IR, a *Venetian blind* and a *blind Venetian* are one and the same thing.