Abstract: The new electronic environments pose a threat and challenge to the theory and practice of knowledge organization. Documents can be approached in electronic retrieval activities in ways not dependent on ‘classical’ knowledge organization activities such as indexing or classification. Accordingly, an argument stating the qualitative difference of knowledge organization in the new electronic environments must show that knowledge organization is worth pursuing and that it is a valuable support to users of information retrieval (IR) systems. In this paper the qualitative difference of knowledge organization and its role in scholarly communication is framed as a question of ascribing cognitive authority to documents. The concern is to examine and discuss how and to what extent knowledge organization as an epistemic instrument in scholarly communication can contribute to ascribe cognitive authority to scholarly documents. The paper is structured in the following way. Initially, a brief examination of the appearance of cognitive authority in knowledge organization, and how that affects an argument stating the qualitative difference of knowledge organization shall be presented. Secondly, the theoretical approach will be outlined and discussed. Then the empirical analysis applying the theory will be presented. The last part will point to the benefits, limitations, and possibilities of the proposed theoretical approach in relation to the conception of knowledge organization as an epistemic activity in scholarly communication.

1. The Argument: Cognitive authority in knowledge organization

When organizing and representing documents in information systems for the purpose of retrieval, these documents are the product of an already established social organization of knowledge or division of labor. They are so because they have been produced with aim of fulfilling or achieving some human activities organized socially. Insofar the social organization of knowledge expresses a kind of cognitive authority this also means that there are several levels of cognitive authority present in a document representation. A traditional document representation consists of various textual elements such as author, title, publisher, institution (or corporate source), journal name, abstract, index terms, and, in the case of the citation indexes, the reference list. It can be argued that these textual elements both in themselves and together contribute to the cognitive authority of a document. These levels can all be seen as communicative interactions mediating between document, author, publisher, library, and reader. Moreover, the scholarly journals are but one expression of the social organization of knowledge within a scholarly domain. They are (usually) organized around some particular subject matter. Author and reader are also organized within the social activities of writing and reading that tie them to the social organization of the domain. The publishers are also part of the...
social organization of knowledge. They are responsible for the distribution of the material output (i.e. the documents) of the disciplines both to the esoteric circles of the disciplines and a wider public. Thus, the knowledge materialized in scholarly documents, and the object for knowledge organization, is the product of already socially organized human communicative activities structured around the social interactions between author, document, reader, and publisher. Bazerman (1988) illuminated this in his analysis of the experimental article in science. He paid particular attention to how this genre emerged historically and how it was, and is, shaped by the typified communicative activities of writers, editors, and readers, and how the genre shaped the knowledge producing activity into a typified activity exactly as a product of history. In addition the actual scholarly document itself may also have an inherent cognitive authority in terms of the credibility, validity, and reliability of what it reports, its theoretical claims, methodology, arguments, or the validity of the interpretation given (Hjørland, 1992a). Thus, given there are already several levels of cognitive authorities present in and ascribed to a document in advance, what is the role left for knowledge organization? This is a theoretical question facing knowledge organization and it must try to offer an answer. Knowledge organization must seek to spell out how it can contribute to the overall characteristic of documents in the communication of knowledge in scholarly communication. This can, among other things, be framed as a question of cognitive authority (Wilson, 1983). For that reason, knowledge organization, as an epistemic instrument in the control of writings (and reading), should help scholarly communication work optimally, and it must try to explain how and to what extent it can contribute to ascribe to documents cognitive authority. By epistemic instrument is meant knowledge organization as a means to maximize over time the intellectual utilization of recorded scholarly knowledge and minimize falsity, error, or ignorance in this process. This is interpreted as being in accordance with Wilson’s (1968) notion of “the best textual means to an end” in connection with bibliographical control, and Hjørland’s (1992b, 1997) “epistemological potentials of documents” in connection with subject analysis.

2. Theoretical framework: A Genre and Writing Approach

In an article analyzing the trajectory of scientific facts from primary literature to popular literature, Fahnestock (1986) presented this as both a shift in rhetorical situation and in genre. With this, Fahnestock points to that because of this, scientific facts cannot just be distilled without a loss of certainty and credibility. In a similar way how can the shift in rhetorical situation and genre be accounted for when it comes to the trajectory of knowledge claims from primary literature to bibliographic literature without a complete loss of certainty and credibility? Being two different genres, primary literature and bibliographic literature have different rhetorical problems because they constitute two different, though interdependent, activities in scholarly communication. The former has to produce, or make claims for, new knowledge. The latter registers and describes primary literature for the purpose of retrieval and bibliographic control. It seems to be essential for the functioning of both primary and bibliographic literatures in scholarly communication that this change is accounted for and understood in order for the bibliographic literature to be an adequate epistemic response to the knowledge producing activity. Studies of scholarly writing (e.g. Bizzell, 1982; Bazerman, 1988; Myers, 1990; Swales, 1990;
Selzer, 1993; Prior, 1998), emphasizing it as a situated and mediated epistemic activity embedded in various discourse communities, have proven to be a fruitful way of understanding the knowledge producing activities, scholarly literature both shapes and is being shaped by. By treating scholarly writing as a product of situated socio-historical and literate activity, and by relying on Miller's concept of genre as 'typified rhetorical actions based in recurrent situations' (Miller, 1994), these studies have pointed to how any claim for knowledge is embedded in typified rhetorical activities. Hence, by uncovering the rhetoric, an insight into what constitutes scholarly knowledge and how it helps organize the recurrent situated knowledge producing activities can be gained. Given that the role of knowledge organization in written scholarly communication can be seen as that of providing and optimizing intellectual access to scholarly literature, it seems reasonable to assume that this cannot be achieved without recognizing the activity of scholarly writing. Thus, it is argued in this paper that an examination of the way scholarly documents are written can contribute to an understanding of its knowledge producing activity and to a more convenient or appropriate organization and representation of scholarly documents in information systems. Following Bazerman (1988), I will analyze a primary article in relation to four contexts as to how they are referred to, invoked, or acted on in the article. The four contexts are 1) the object under study; 2) the literature of the field; 3) the anticipated audience and 4) the author's own self (Bazerman, 1988, p. 24). Extending this analysis to also cover the corresponding organization of knowledge claims in knowledge organization systems (e.g. bibliographies, catalogs, or classification systems), the objective is to examine the relationship between the knowledge the documents claim to produce and contribute with, and how they are in fact indexed. This can account for the shift in genre that takes place when indexing scholarly documents in information systems.

3. Empirical study: The Case of a LIS-article

The article chosen for examination is Harter's (1992) article on psychological relevance. In here Harter sets out to introduce Sperber & Wilson's psychological concept of relevance presented in their book 'Relevance: Communication and Cognition' (1986). Psychological relevance is what affects or causes a cognitive change in the mind of the user. Harter's agenda implies two things. First, Harter has to believe that this kind of relevance is useful for LIS. Second, he has to show this usability, and thereby argue in favor of it. Consequently, Harter finds himself in an epistemic-rhetorical situation. The significance of this epistemic-rhetorical situation is further underscored by the fact that relevance research since the Cranfield experiments has been, and still is, a classic area of research in information retrieval (IR), challenging Harter with a heritage that he must take into account. Moreover, Harter's article is a theoretical article about a concept that is born out of a highly experimental tradition. Thus, Harter speaks through a particular research genre. What, then, does Harter contribute with? The object under study in Harter's article is the concept of relevance. Harter begins his article with contextualizing the historical rootedness of relevance research in LIS. In the opening paragraph of the Introduction relevance is referred to as 'central to information retrieval (IR)' and 'as the basis of nearly all experimental evaluation and testing.' (p. 602). As a research area relevance is further invoked, when Harter identifies two classic relevance 'camps' in IR research: The objective sense of relevance, based on the Cranfield
model, and the subjective sense of relevance. The former contemplated in terms of the topicality of the search request and document(s) retrieved and the degree of match between these two. The latter as the view that takes the relevance judgments of the user into account. That way Harter establishes a context for his upcoming argument and analysis. Harter invokes the literature of the field in two ways. When talking about the objective view of relevance the literature invoked is the critics of this view, no proponents are cited. In continuation of this, Harter comments that despite the critique of the Cranfield model '...the model has been accepted, generally uncritically, as the basis for most experimental research done on IR systems to the present day.' (p. 602). Thus, the author's own self become present when commenting on the acceptability of the Cranfield model and on the lack of critical distance to it. Harter's critical attitude toward objective relevance is further underlined when stating that: 'The idea that users want, or will be most happy with, documents on the topic of a search statement is an assertion that has absolutely no empirical support.' (p. 603). Thus, relevance is also referred to as an empirical matter. As a consequence of this, Harter makes himself present by claiming that 'I believe that a view of relevance that equates relevant documents with documents "on the subject" has been pernicious, both to the ultimate goals of the user and to the development of theory in information science.' (p. 603). Thus, Harter is very clearly distancing himself from objective relevance. He does by making certain rhetorical moves: He dos not cite proponents of the view. This can by explained by the fact that it may be assumed to be known by the audience. He emphasizes the critics of the Cranfield model. He further invokes his audience by claiming that objective relevance does not have empirical support, implying that if it had, it could have been more acceptable. Further, objective relevance is accused of having been a major hindrance for theory development in LIS. This is a strong attack on objective relevance implying that Harter's contribution must be considered as something that can contribute to and re-create theory development in LIS. In contrast to objective relevance, proponents of subjective relevance are cited by providing what Swales (1990, p. 148) calls non-integral citations. That is, citations occurring in parentheses rather than part of the citing sentence as a sentence element. Different notions of subjective relevance such as pertinence, situational relevance, perceived utility, informativeness, beneficially are referred to. Furthermore, the review of relevance given by Schamber, Eisenberg & Nilan (1990), arguing for a dynamic and user-oriented view of relevance, is characterized as excellent. Without having introduced readers yet to the ideas of Sperber and Wilson (1986), Harter nonetheless connects the psychological concept of relevance to the subjective view of relevance in LIS: 'It roughly incorporates or is consistent with the notion of pertinence, situational relevance, perceived utility, informativeness, beneficiality, and other types of subjective relevance... However, without a supporting theory, subjective relevance is too vague to offer much insight.' (p. 603). Hence, Harter claims that subjective relevance needs a theory and offers Sperber & Wilson's theory of relevance as a solution: 'Not only is Sperber and Wilson's theory of psychological relevance directly applicable to the notion of relevance in information retrieval, it offers valuable insights into its essential nature.' (p. 604). Thus, already in the Introduction the theory is decided to be applicable and valuable for IR. Accordingly, in the Introduction Harter has set the stage for his further knowledge claims. He has dissociated himself with the notion of objective relevance and sets out to argue that Sperber & Wilson's psychological relevance can be a supporting
theory for subjective relevance. The latter needs that and the former has been a hindrance for the development of theory. That way, the concept of relevance (the object of study) is given an epistemic-rhetorical context from where to argue, discuss, and analyze it. Sections three and four of the article, ‘Psychological Relevance and Information Retrieval’ and ‘An Extended Example’ are decisive for Harter’s agenda in that he has to show the applicability of Sperber & Wilson’s relevance theory to IR; i.e. his contribution to knowledge is what is going to be presented here. This is done through theoretical analysis and exemplification. Because Harter is about to make a knowledge claim, it is important for him that he invokes his audience since he is not allowed to state whatever he likes. In view of that, the audience is invoked as being part of Harter’s activity of interpreting the theory: ‘Let us see how psychological relevance can be interpreted in terms of a document retrieval system.’ (p. 606). Thus, the interpretation of the theory is a collective activity meaning all participants take responsibility. This way of invoking the audience is repeated when revising the definition of information need on the basis of the theory of psychological relevance: ‘Let us revise our definition of information need, then,...’ (p. 606). Having analyzed psychological relevance in relation to IR, Harter once again invokes his audience, signaling the collective nature of the activity: ‘Psychological relevance allows us to talk about an “information need” as the current context – the cognitive state at a given time – of an individual who consults an information system.’ (p. 607). Accordingly, the object of study (i.e. relevance) is turned into a psychological issue. However, when finishing this section, Harter makes clear that he has done the interpretation but with permission from his audience: ‘My interpretation of information need suggests that we will...’(p. 607). Thus, when Harter has to argue for his contribution to knowledge and show its innovativeness, he invokes his audience in this activity and re-interprets old conceptions of information need in light of psychological relevance. The explanation for this may be due to that knowledge claims, following Fleck (1979), in journal articles are at the time of publication not yet to be considered a part of the certified body of knowledge of a research community (the thought collective in Fleck’s terms). The topics or problems treated in a particular journal article must therefore in some way or another address topics or problems already known to the thought collective in order to acquire a possible acceptance. This is in line with Fleck (1979, p. 98) when he defines a fact as “...a stylized signal of resistance in thinking.” and because the thought style is determined by the thought collective, a fact is then designated “...as the signal of resistance by the thought collective.” (Fleck, 1979, p. 98). Harter exactly addresses psychological relevance in relation to topics already known by the LIS research community in order to overcome, or deal with, the signal of resistance, and to turn his interpretation of psychological relevance into a signal of resistance, into a fact. Otherwise he cannot claim to have made a contribution to knowledge. After this theoretical analysis and the establishment of his contribution to knowledge, Harter goes on with exemplifying his new theory, as a way of demonstrating to his audience the concept of psychological relevance and how it can be interpreted in relation to LIS-issues. That way he invokes a familiar ground for both author and audience. Turning to the implications of the concept of psychological relevance for IR, Harter has to demonstrate how his contribution to knowledge makes a difference. He does this by discussing the concept in relation to a range of IR research issues. Throughout this section the literature of the field is invoked when
something similar to Harter’s suggestion has been proposed or when the critics of objective relevance are referred to. Like in the Introduction, potential proponents of the objective view of relevance are not cited. It is assumed that this is part of the common knowledge of the audience. This is clearly illustrated when discussing the notion of information need and how it has been employed in IR: ‘...it [information need] has usually been taken as static – representing something that is fixed an unchanging – even in very recent writings; the work of Schamber, Eisenberg & Nilon (1990) and Katzer and Snyder (1990) are rare exceptions.’ (p. 610). The ‘very recent writings’ are not given examples of, while those representing an exception are cited. Harter thereby invokes his audience through his passive and active use of the literature. Showing the implications of psychological relevance to selected information seeking studies, Harter makes his presence and refers to psychological relevance as has been ever present in these, but not explicated: ‘I suggest that psychological relevance may be the concept that underlies these models of the information-seeking process, and that unites them. Psychological relevance helps explain why these models work, as behavioral descriptions of information-seeking behavior.’ (p. 611). By arguing that psychological relevance underlies and can help explain the successes of these models, Harter seeks to justify the concept and thereby shows its role as a supporting theory. Those who do not agree with this then, now have the burden of proof for showing this is not so. Because psychological relevance implies, in Harter’s interpretation of it, changing and dynamic relevance judgments, retrieval tests and relevance judgments based on the measures of recall and precision are referred to as wrong (p. 612). Harter further underscores this disapproval of retrieval tests by making one of the most explicit claims and evidence of his persona in the article: ‘I no longer believe that there is a valid interpretation of the meaning of the results of such tests.’ (p. 612). By referring to the results of traditional retrieval testing as invalid and wrong, Harter has just refuted four decades of traditional, experimental IR research. This is a strong claim, which presumably cannot go unnoticed and unconunented in the IR research community. It could be speculated that this claim has been the motivation for the writing of the article. Throughout the entire article the concept of psychological relevance is constantly held in opposition to this established research area. Further, the sentence ends the paragraph and no other evidence is invoked or referred to in order to strengthen the claim. It is a statement that needs no further explanation, insofar the theory of psychological relevance is accepted. When analyzing the implications of psychological relevance in connection with IR and bibliometrics, relevance is invoked as the notion that brings these two research areas together (p. 612). Psychological relevance and its relation with citation practices is invoked as a measurable and empirical concept that needs to be tested in order to ‘prove’ its applicability. Thus, besides the theoretical approach taken in the article, Harter also appeals to and recognizes the need for empirical research on this matter. An article co-authored by Harter is here referred to as an example of an empirical test that has been conducted. In the last section of the article suggestions for further research are proposed. Once again psychological relevance is invoked as something that in future research needs to be empirically tested and measured (p. 614). Harter poses several questions as to how this can be done and what it will imply. This is also a way of accommodating his audience. Harter has not presented empirical evidence in his article, but acknowledges that psychological relevance needs an empirical basis if it is going to have a future role to play in LIS research.
Thereby, Harter accommodates the kind of audience that wants empirical (experimental) evidence in favor of a theory. The audience is invoked several times by the use of ‘we’, ‘us’ and ‘our’ when Harter is pointing to the need for investigating, or re-investigating, LIS research topics in light of psychological relevance. Once again is it being stressed that psychological relevance needs to be tested empirically (p. 614). Consequently, the structure of the article can be regarded as a rhetorical structure because it guides Harter’s overall argument and his knowledge claims. The relationships between author, audience, literature and the object of study are in constant interaction, demonstrating how knowledge production can be seen as a situated literate activity. How then is Harter’s article indexed in LIS-bibliographies? How is this activity accounted for? LISA has assigned to the article the descriptor-string “Information science; Implications for; Psychological relevance”. Where does such kind of indexing lead to? What kind of activity does it serve? The implications of psychological relevance for information science is also part of Harter’s knowledge claim. If it is accepted as a suitable relevance concept for LIS, then it has some implications for existing LIS research and then the indexing to some extent reflects the knowledge producing activity of the article and contributes to its cognitive authority. However, the title of Harter’s article, ‘Psychological Relevance and Information Science’ indicates that this article is about the implications of psychological relevance for LIS. Further, if title access does not fully suggest this, abstract access certainly does. Here it is stated that the implications of psychological relevance for LIS is treated in the article. On that point it could be argued that the indexing undertaken is redundant because it does not make a qualitative difference, though pointing to Harter’s knowledge claim. In Information Science Abstracts the article is indexed with these descriptors: “Bibliometrics”, “Information retrieval”, “Information science”, and “Psychological aspects”. These descriptors are broad and do not as such refer to relevance. “Bibliometrics” is not a totally odd descriptor. Harter suggests that psychological relevance underlies the citation behavior of scholars (and his also suggests that his article is not about bibliometrics (Harter, 1992, p. 603)) and that IR and bibliometrics are connected through this notion of relevance. This is part of his general knowledge claim. The descriptors “Information Retrieval” and “Information science” must be considered inappropriate. As a descriptor “Information Retrieval” does not reflect, or relate to, Harter’s overall knowledge claim or argument, but is of course affected insofar his knowledge claim is accepted and as an access point it does not have enough discriminating value. This also applies to the descriptor “Information Science”. In a bibliography covering LIS, the descriptor is useless since all documents in terms of their knowledge claims must necessarily either be about or relate to LIS. Indeed it must be the criterion for inclusion in the bibliography. From a retrieval point of view, the title provides access through this term suggesting that also indexing the document with that descriptor is redundant. Despite its discriminating value, “Psychological aspects” is ambiguous as a descriptor. Harter clearly implies psychological aspects with his knowledge claims. However, psychological aspects in the present of Harter’s article are not specific enough since this does not indicate what kinds of psychological aspects (e.g. cognitive, behavioral, or cultural) there are at stake. The indexing of Harter’s article in two LIS bibliographies reveals that it does not fully reflect the knowledge producing activity of Harter’s article. The shift in genre implies that Harter’s knowledge claims are not clearly accounted for and the indexing
undertaken cannot be justified in terms of its contribution to the cognitive authority of the article. What, then, can the above close reading of Harter’s article contribute with in terms of appropriate descriptors that reflect, or mediate, his arguments and knowledge claims? Harter examines Sperber & Wilson’s (1986) concept of psychological relevance and proposes it as an appropriate relevance concept for LIS studies. It implies re-thinking the whole notion of relevance within LIS. To do that Harter carries out a concept study with the aim of arguing for and showing its relevance for LIS. In this sense a descriptor reflecting his mode of argument and his knowledge producing activity could be “Concept Study”. In opposing it to topical relevance, relevance is turned into a cognitive concept. A descriptor for this argument and knowledge claim could be “Cognitive Relevance Theory” revealing the implicit theoretical assumption of the theory and concept proposed. A methodological implication of this theory is its focus on the individual relevance judgments, which is why “Methodological Individualism” could be suggested as a descriptor too. That way the knowledge claim and its methodological consequences are indicated. Further, despite his critique of traditional retrieval testing, Harter emphasizes relevance as a measurable and empirical concept. This suggests an underlying epistemological assumption of Harter’s knowledge claim. An appropriate descriptor that would indicate this could be “Empiricist Epistemology”. As a consequence of psychological relevance, Harter also criticizes and rejects the validity of the results of recall and precision measures in applied in experimental evaluation and testing. Thus, a descriptor like “IR-Criticism” could be assigned since that would indicate the critique. This critique is part of Harter’s overall knowledge claim exactly because what the knowledge claim offered (psychological relevance) is the negation of what it criticizes. Since Harter aligns psychological relevance with various behavioral information seeking studies and argues that it underlies these, “Behavioral Studies” can also be suggested as a descriptor. This descriptor can be considered to be part of the argument because behavioral studies are a way of examining psychological relevance. Accordingly the descriptors suggested can be seen as a result of the epistemic-rhetorical situation Harter was faced with when writing his article. Exactly because they are part of this, we cannot expect that these be extracted on a mechanistic basis (e.g. by the use of automatic means). They form part of Harter’s knowledge producing activity and are as such embedded in a socio-rhetorical situation. The argument and the knowledge claims are something that must be contextualized in the social act of writing. Therefore, looking at the ways of invoking on, referring to, and evoking the contexts of the object under study, the literature of the field, the audience, and the author’s self can tell us what kind of argument is being made and in what the knowledge claims consist. These contexts reveal the very activity of the text of which it cannot do without. The activity of the text is situated in the activity of knowledge production. In terms of knowledge organization the above leads to the following claims: First, the reason why we can discuss the appropriateness of descriptors has to do with what arguments and knowledge claims a document put forward. This is again bounded to the rhetorical situation of the document and thereby its genre. Second, an understanding of a document demands more than its mere content in itself can explain; it demands an understanding of its activity as historically situated. The analysis presented above can provide a reasonable basis for discussing how the article is indexed in various scholarly bibliographies. A discussion of what the shift in genre implies. While the indexing of the article cannot be discussed in terms of
the truth or falsity of the indexing, the degree of appropriateness of the indexing undertaken can be discussed. Not all words and concepts are equally appropriate for an indexing of an article. In order to index scholarly documents and support the knowledge producing activity, knowledge organization must not fail the researcher (Weinberg, 1988). A scholar writing an article is faced with rhetorical problems in order to produce a knowledge claim. That way the analysis can be used to discuss strengths and weaknesses of knowledge organization and to what extent it contributes to cognitive authority, which may ultimately lead to specific descriptors stating a qualitative difference in relation to other access points. It may be objected that this kind of analysis is unrealistic because such a close reading cannot be done in ‘real-life situations’. There may be a sense of truth in this. But then we will also have to admit that we cannot account for the shift in genre that happens when indexing scholarly documents in bibliographies and why that shift may pose a problem for searchers. We cannot expect then that scholars will use the systems for knowledge organization, as most studies show they do not. We cannot conceive of our knowledge organizing activities as ‘user-friendly’. It is simply a chimera.

4. Practical Implications for Teaching Indexing

The problem of writing knowledge is discussed by Bazerman (1988). He poses the question of how students learn to write academic essays and what kind of problem writers are faced with: ‘Very soon into engaging this problem, I found that I could not understand what constituted an appropriate text in any discipline without considering the social and intellectual activity which the text was part of.’ (Bazerman, 1988, p. 4). As researchers and teachers of knowledge organization we face exactly the problem as Bazerman. The answer(s) to this question is exactly the same as those to the question ‘What do you need to know in order to index a document?’ It is of crucial importance for those having the task of indexing documents that they recognize the rhetorical situation they are in. This implies recognizing the social and intellectual activities documents are part of, or may serve, and that a choice must be taken as to what kind of descriptors are appropriate to assign in order to support the activity. This calls for a need of genre and activity knowledge. The indexer, like the informed writer (Bazerman, 1995), must know how to act within activities structured and shaped by the activities of the production, distribution and use of documents. These activities are historically constituted that have turned them into typified actions based in recurrent situations. This shows then a strong connection between how knowledge is socially organized through typified actions and how it is organized in information systems. The above analysis has, hopefully, revealed what we need to know about scholarly documents when organizing and representing them in information systems, and the limitations there exist when trying to ascribe cognitive authority to scholarly documents. Only then can the role of knowledge organization in scholarly communication be fully acknowledged and understood and the shift in genre from primary literature to bibliographic literature appropriately accounted for without a complete loss of cognitive authority. Only then does knowledge organization make a qualitative difference.
References


