Derrida, Logocentrism, and the Concept of Warrant on the Semantic Web

Abstract
The highly-structured data standards of the Semantic Web contain a promising venue for the migration of library subject access standards onto the World Wide Web. The new functionalities of the Web, however, along with the anticipated capabilities of intelligent Web agents, suggest that information on the Semantic Web will have much more flexibility, diversity and mutability. We need, therefore, a method for recognizing and assessing the principles whereby Semantic Web information can combine together in productive and useful ways. This paper will argue that the concept of warrant in traditional library science, can provide a useful means of translating library knowledge structures into Web-based knowledge structures. Using Derrida’s concept of logocentrism, this paper suggests that what while “warrant” in library science traditionally alludes to the principles by which concepts are admitted into the design of a classification or access system, “warrant” on the Semantic Web alludes to the principles by which Web resources can be admitted into a network of information uses. Furthermore, library information practice suggests a far more complex network of warrant concepts that provide a subtlety and richness to knowledge organization that the Semantic Web has not yet attained.

Introduction
The concept of warrant in traditional library science has been a cornerstone of our understanding of, and our training in, the major tools of subject access, including the Library of Congress Subject Headings, Library of Congress Classification, and the Dewey Decimal Classification. As the articulation of the intellectual basis for the inclusion of entities in a formal ontology, it offers a suggestive means of exploring how LIS traditions can play a role in emerging Web technologies. When we compare the principles of warrant in the two domains, we find that library articulations of warrant principles center on the decisions of a central repository, or of those responsible for large, ubiquitous tools for subject access, tools that give a clear and distinct shape to library collections. On the Semantic Web, questions of warrant surface at the individual level who would set individual preferences: the warrant involved is the decision of which namespaces to admit as trusted and useful domains of information use.

Furthermore, library articulations of warrant are a curious combination of the formal and the informal, the written and the spoken, subjects for ongoing debate, deliberation and slow consensus-building. Semantic Web standards, on the other hand, take these urgent questions and turn them into formal, machine-readable standards. This paper uses Derrida’s theories of logocentrism to explore these two tensions: between centralized and decentralized questions of warrant, and between warrant as a set of written and oral traditions and warrant as programmable code.

The Concept of Warrant in Library and Information Studies
Svenonius (2000) describes warrant as “the specifying criteria for term selection” in a subject language (135). While this definition is certainly true, Beghtol’s definition (1986) provides a more explicit reference to the status of warrant as a two-stage “authority”: first to justify and subsequently to verify decisions about what classes/concepts to include in the system, in what order classes/concepts should appear in the schedules, what units classes/concepts are divided into, how far subdivision should proceed, how much and where synthesis is available, whether citation orders are static or variable and similar questions. (110)
Warrant, according to Beghtol, is an authority that underlies an entire network of practices and decisions that occur not just during the creation of the subject access scheme, but also during its subsequent use and subsequent revision. Beghtol isolates four distinct concepts of warrant:

1. **Literary** warrant, in which the subject scheme is shaped by a pre-existing literature, either in a specific library or a specific bibliographic domain (112);
2. **Scientific** or philosophical warrant, in which the scheme reflects a prevailing scientific or philosophical consensus about the nature of knowledge in its various fields (115);
3. **Educational** warrant, in which the scheme reflects the practical needs of educational institutions (117);
4. **Cultural** warrant, in which the scheme reflects prevailing cultural values and assumptions (119).

Svenonius, by contrast, isolates three conceptions of warrant: literary, use and structural. Use warrant reflects the perceived needs of information users (135), while structural warrant allows for the inclusion of terms that may not be in common practice, but which provide a means of collocating concepts in a useful way (136).

Despite their different formulations, both Beghtol and Svenonius agree that subject languages rely on the useful exploitation of semantic and syntactic dimensions of subject access, and that warrant provides a means of understanding how the semantic dimension—the selection, arrangement and normalization of terms within a subject domain—manifests itself over time. Warrant can be viewed as a library’s ongoing commitment to analyze, justify and revise the semantic elements of its access systems.

Warrant, then, introduces a new dimension to the relationship between a subject access scheme and the documents that it organizes. Classification schedules typically relate to library collections in a way similar to Ferdinand de Saussure’s distinction between *langue* and *parole*: between the sum of all that could potentially be expressed by a language, and the sum total of language “acts” that comprise what is actually expressed. (Saussure 1986, 29) But if we include warrant in this relationship, we realize that warrant criteria monitor and shape the relationship between the system and its documents, between the class numbers we could assign and the class numbers that we actually assign.

But how does warrant actually do this monitoring and shaping? This is a difficult question, largely because warrant appears in many different forms: in the instructions, manuals and formal procedures that accompany our subject access tools, in the formal written policies that underlie the tool’s use in a specific organization, in the inculcation of classification principles in library schools and training sessions, in the deliberations and procedures for revising subject tools, in the various ways in which the tool is discussed, debated, revised, queried, challenged, defended and rejected. But warrant monitors and shapes subject access on at least three distinct axes of implementation:

- the use of the tool to classify existing documents, in accordance with the rules laid out in the accompanying manuals;
- the decision to reclassify documents, in accordance with a change in cultural perspectives or user needs;
- the decision to change the classification structure itself, in accordance with changes in cultural perspectives or user needs.

Along all three axes, libraries and information organizations of all types use principles of warrant to scrutinize both the success with which they have implemented a specific
subject structure, and the degree to which that underlying structure—be it a reflection of existing literature, philosophical or educational consensus, cultural values or user patterns—remains valid.

**Analogies to Warrant on the Semantic Web**

The Semantic Web is a visionary project spearheaded by Tim Berners-Lee and carried out largely under the auspices of the World Wide Web Consortium. It rests on a principle of machine-understandable data: systems that would enable an intelligent agent to perform logical inferences upon data retrieved, to get specific answers to specific questions, to mine databases, and to discover relevant resources in diverse knowledge domains. These principles manifest themselves through a layered architecture, in which sophisticated ontologies such as OWL (Working Ontology Language) rest upon broader stores of metadata such as RDF (Resource Description Framework), all of which is encoded in XML.

Like many visions, this one is ambitious; unlike many visions, this one is focused on standards:

> Progress is measured on a scale that starts with the conception of an idea, which works through its acceptance as a common technology and becomes an interoperable standard, and ends with a global market based on it. … [The Semantic Web] absolutely needs the interoperable infrastructure that only global standard protocols can provide”. (Berners-Lee 2003, xi)

The W3C promotes these standards, in hopes that their widespread adoption will lead to a new transformation in Web culture, a transformation that will broaden the range of Web applications available to us, and in so doing dramatically reinvent the relationship between humans and computers.

Not everyone believes that the Semantic Web will achieve such widespread adoption: some believe, along with Clay Shirky, that the Semantic Web is good for creating syllogisms, but not much else (Shirky 2005). But the library community, with its well-developed access tools, has begun to look to the Semantic Web with considerable interest. Greenberg (2007, 203) notes the close analogies between library services and Semantic Web activities. Harper & Tillett (2007) suggest that the Semantic Web could serve as a means of moving these highly-structured subject access systems into Web environments, where they could greatly enhance knowledge accumulation and sophisticated retrieval (47), while Campbell & Fast suggest ways in which bibliographic description in academic libraries could be transformed by closer ties with Semantic Web data (2004).

If libraries do play a part in the Semantic Web, what role, if any, will concepts of warrant play? Initial indications suggest that warrant will be of pressing importance. While the Resource Description Framework provides a useful and rich means of creating metadata about Web resources, some allowance will have to be made for the fact that multiple communities will be creating and using this metadata, and this will cause problems when search agents attempt to aggregate Web data from multiple sources. Most Semantic Web applications will rely heavily on an ontology: “a formal, explicit specification of a shared conceptualization” that offers to both humans and web agents “a shared and common understanding of some domain that can be communicated among people and application systems” (Fensel et al. 2003, 11). Like the tools of libraries, ontologies will determine not just the terms to be admitted, but the relationships among them:

> An ontology establishes the things that a system can talk and reason about. This means the vocabulary, but … there’s more to it than just a collection of words and names—the terms have logical relationships to each other that need to be specified, and this in turn means that any ontology system must adopt some variety of logic, either formally or informally. (Passim 2004, 133)
Ontologies, like classification systems, enumerate the entities in a domain, assign them names, and establish relationships among them: all activities which principles of warrant both initially justify and subsequently verify (Beghtol 1986, 110). Ontologies, like classification systems, are presumably products of warrant. Unlike most traditional classification systems, ontologies are designed for machine readability and machine implementation. And unlike most classification systems, ontologies are designed to extract and aggregate data at a very granular level, from multiple sources.

What matters, then, is not the structuring principle of each individual ontology, but the principle by which data from each ontology is combined together. If the closest analogy to a classification system is an ontology, perhaps we can find warrant at work in the mechanism by which an ontology is invoked: the XML namespace. All Semantic Web standards rely on the principle of the XML namespace: the means of avoiding “collisions” between elements with the same name but different meanings by expanding them with a prefix that designates their origin, or “namespace” (W3C 2006).

By limiting and discriminating between different communities’ use of similar element and attribute names, XML namespaces provide a rough, machine-readable analogue to the concepts of warrant used in classification systems. But of course, the concept of warrant mutates considerably in the Semantic Web context. Instead of articulating the intellectual or institutional authority that justifies the inclusion or exclusion of a category in a classification system, the XML namespace provides a machine-readable manifestation of that authority, which is used to include or exclude Web resources in three main spheres of activity:

- Resource aggregation: namespaces, as well as crosswalks that map between similar elements in different namespaces, are used when harvesting resources across different domains, to ensure that only those resources that meet the search criteria are found;
- Data extraction: namespaces provide the means of identifying with accuracy the relevant fields and attributes of databases, enabling agents to extract data from the deep web with precision and selectivity;
- User-Resource interaction; as libraries discovered long ago, effective information use depends on translating the user’s needs into the language of the information system, a process that is frequently iterative and interactive in nature.

XML namespaces, therefore, provide a useful means of connecting library systems of metadata, classification and thesaural terms to a Semantic Web environment; namespaces for the Dublin Core, for subject heading lists, and for classification categories provide the means whereby the rigour of the intellectual structure of a given field of knowledge can be joined with the power of the search agent, allowing only those resources into a search set that satisfy the demands of this structure. While “warrant” in library science traditionally alludes to the principles by which concepts are admitted into the design of a classification or access system, “warrant” on the Semantic Web alludes to the principles by which Web resources can be admitted into a network of information uses.

But what effect is produced by the movement from human understanding to machine understanding, and from the movement from warrant principles that govern a classification scheme to the warrant principles that identify appropriate namespaces? The movement of computer technology into the process of reasoning and inference has raised suggestive issues for knowledge organization, and in particular for the complex judgments that regulate and administer the ongoing adaptations of warrant principles for large-scale
information access tools. To analyze those questions, we need a theory that specifically addresses oppositions between the artificial and the real, between the outside and the inside, and between artifice and spontaneity.

Derrida and the Logocentric View of Knowledge
Since its first appearance in 1967, Jacques Derrida’s *Of Grammatology* has raised powerful questions about the ways in which we implicitly rank various forms of discourse, various forms of expression. Derrida coins the terms “logocentrism” and “phonocentrism” to characterize what, for him, is Western culture’s lamentable tendency to devalue writing in favour of speech. Derrida argues that linguistics, as pioneered by Ferdinand de Saussure, proceeded from a phonocentric bias which led Saussure to treat the spoken word as the primary signifier of the signified, with the written word as the exterior copy: “the voice, producers of the *first symbols*, has a relationship of essential and immediate proximity with the mind. … The written signifer is always technical and representative” (Derrida 1976, 11). For Saussure, Derrida argues, the purity of spoken language is always fighting the corruption and debasement of writing. Derrida maintains that this battle has been waged ever since Plato’s *Phaedrus*, and that the battle was always already lost. The supposed preeminence of speech over writing relies on a series of binary oppositions that were suspect from the beginning: presence and absence, signifier and signified, exterior and interior. As these oppositions collapse under Derrida’s analysis, so too do the logocentric assumptions of language as the articulation of presence.

Derrida’s deconstruction of phonocentrism may seem far removed from concerns of warrant on the Semantic Web. But the Semantic Web, and indeed the Web as a whole, have forced us to reopen many questions in our culture about what constitutes a “real” communication as opposed to an artificial one. Dave Weinberger (2002) challenges many of our assumptions about the Web, especially the assumption that it has no space: the Web, he claims, is an environment of lived space rather than measured space, and one in which human contact is the norm rather than the exception (32). And at first glance, the Semantic Web seems to defy the logocentric bias which, according to Derrida, pervades Western society and intellectual thought.

Logocentrism and the Semantic Web
Berners-Lee’s grandiose vision of the Semantic Web rests upon an “interoperable infrastructure” of “global standards,” just as the Web we know today rests upon the widespread adoption of the URL, the Hypertext Transfer Protocol, and HTML. For Berners-Lee, the Semantic Web’s suite of standards, including RDF, RDF-Schema, OWL and XML Namespaces, holds a similar relation to ontologies and Web objects that warrant principles occupy in relation to classification schemes and classified documents. In both cases, the relationship between document and ordering scheme rests upon a generative principle of justification and verification that governs, not just how the ordering scheme is created, but how it is subsequently interpreted and revised. The Semantic Web, however, takes these diffuse principles and articulates them as formal, machine-readable standards. In so doing, the Semantic Web appears to have taken the acts of computer programming, standards creation and machine processing—acts which rarely appear to the popular imagination as primary creative acts—and placed them at the very center of an emergent Web information environment. At the heart of Web interactions lie documents in the form of standards and protocols that govern the behaviour of Web agents. Just as warrant is
the principle that lends coherence to our semantic knowledge structures, Semantic Web protocols are designed to lend coherence to our interactions on the Web: by resolving ambiguities, aggregating facts from different sources, and finding meaning and pattern in data stores too massive for humans to interpret without assistance.

However, before we start hailing the Semantic Web as the world’s first truly post-structuralist information architecture, we should remember that these standards are designed to remain largely out of sight. Indeed, RDF in its XML form is so complicated and difficult to read that its widespread adoption depends largely on the design of applications that can render it all but invisible. Much of the Semantic Web rhetoric suggests that Web agents would adopt a whole new incarnation of “presence”: they would propose solutions to current problems (Berners-Lee & Hendler & Lassila 2001) or prompt users with suggestions (Passim 2004, 8). And because the data structures and underlying ontological assumptions of the various SW standards would remain largely out of sight, they would remain largely outside the user’s awareness.

Conclusion: The Library Community, Warrant, and the Future of the Web

The prospect of using tools such as DDC, LCC, UDC and other schemes and thesauri as a basis for the Semantic Web is certainly intriguing, and certainly deserves further study. But the library community has something to offer besides its classification systems, based as they are upon physical documents in centrally-organized physical repositories. Greenberg suggests that libraries and library systems grew out of a collaborative spirit (2007, 207), and this collaborative spirit comes in particularly handy around issues of warrant. The binary nature of computer data forces automated systems, even those with sophisticated inference capabilities, into a naive position in relation to their retrieval practices and their knowledge structures. Because the namespace principle attempts to resolve ambiguity to heighten the performance of search agents, it forces the agent to take seriously a series of oppositions that librarians have learned, through experience, to treat with some irony: the opposition of presence to absence, of internal and external, of fundamental equivalence and fundamental difference. Librarians have traditionally responded to these complexities by employing a diverse collection of strategies, based on different principles of warrant, ranging from the literary warrant, through to philosophic warrant and to the principles of cognitive authority.

Furthermore, library systems have historically developed within the context of constant, sometimes acrimonious debate from different stakeholders in the information community. From cataloguers who offer suggestions and comments on problematic subject headings, to users who object to the knowledge organization they encounter in libraries, to librarians who develop special schemes for specific interests and communities, to international bodies of professionals committed to the slow and deliberate process of consensus building, the library community airs its issues of warrant in a combination of speech and writing, in physical and digital form, in a way that invites community participation on many levels. As such, library information practice has evolved a complex discourse around questions of warrant that provide a subtlety and richness to knowledge organization that the Semantic Web has not yet attained. If the Semantic Web is to develop further as a new environment for library services, its systems and agents will need to find new ways of incorporating the satisficing strategies that have developed in the library world to deal with the incongruities of information environments.
References