A Hierarchy of Metadata Elements for Business Intelligence Information Resource Retrieval

Neil Foshay, Department of Information Systems, St. Francis Xavier University, Antigonish, Canada

Todd Boyle, Department of Information Systems, St. Francis Xavier University, Antigonish, Canada

Jacob Mather, Department of Information Systems, St Francis Xavier University, Antigonish, Canada

ABSTRACT

Effectively managing information resources is an important activity contributing to the competitive advantage of modern organizations. Organizational knowledge workers must be able to search for pertinent information quickly and effectively. This research identifies the relative usefulness of the metadata elements associated with the Dublin Core metadata standard for the effective retrieval of three different information resources – structured business intelligence reports, structured spreadsheet reports, and unstructured reports in formats such as Word and PowerPoint. A survey of knowledge workers was conducted to determine the relative usefulness of the metadata elements for each of the three information resources and to develop a framework outlining where metadata tag requirements differ between such resources. Overall, the study and resulting framework emphasize the need for system developers and database management personnel to be cognizant of the type of information resources being used, and ensure that search metadata elements that are appropriate for these specific resources are in place.

Keywords: Business Intelligence, Dublin Core, Information Retrieval, Knowledge Workers, Metadata

INTRODUCTION AND MOTIVATION

Information is a strategic asset for today’s organizations. The ability to effectively capture, organize and exploit information resources can provide a competitive advantage, and is critical to today’s information and technology rich organizations (Zheng, Fader & Padmanabhan, 2012). Organizations build business intelligence (BI) applications to meet the information needs of knowledge workers (Ramakrishnan, Jones, & Sidorova, 2012; Kiron, Shockley, Kruschwitz, Finch & Haydock, 2012). Business intelligence
applications provide users with direct ‘ad hoc’ access to structured data, as well as to prebuilt reports, scorecards and dashboards based on this data (Chaudhuri, Dayal & Narasayya, 2011). In addition, it is common in organizations for individuals to export data generated from business intelligence applications (and other data sources) into structured documents (such as spreadsheets) for further analysis and manipulation and to create unstructured documents (such as presentations and documents in tools such as Microsoft PowerPoint and Word) that analyze and interpret this data (Roberts, 2010).

As a result, there is currently an ever-expanding pool of heterogeneous information resources that need to be effectively managed in order to be reusable and by extension, valuable to the organization. The need for large volumes of information has resulted in challenges for many organizations as an overwhelming volume of data, reports and documents are created on an ongoing basis (Davenport & Snabe, 2011). In a seeming paradox, the more information that gets created in an organization, the less it is used, as users and potential users become overwhelmed by the sheer volume (Bruce, 2005). Addressing this issue requires an effective information retrieval mechanism: a powerful and easy to use method that allows individuals in an organization to search for and retrieve the information resources they need.

However, modern information technology (IT) departments are struggling with how best to identify information resources and to organize and catalog them in a way that they can easily be retrieved. An effective search mechanism requires the right metadata (Efron, 2011). Metadata can be thought of as data ‘tags’ associated with an information resource that describe the characteristics of the resource (Hurst, 2010).

This study seeks to determine the feasibility of employing a standard set of metadata elements to aid in the retrieval of three different BI-oriented information resources, specifically: (1) structured BI generated reports; (2) structured reports implemented in tools such as spreadsheets; and (3) unstructured reports from Word, PowerPoint, and similar applications. The objectives of this study are to:

1. Develop a hierarchy of metadata elements used by knowledge workers when searching for structured-BI, structured non-BI, and unstructured information resources;
2. Identify the most (and least) useful metadata elements for information retrieval in a Business intelligence context by knowledge workers; and
3. Explore how the relative usefulness of metadata elements to knowledge workers vary on the type of information resource (i.e., structured-BI, structured non-BI, unstructured) to be retrieved.

**RELATED WORK**

**Business Intelligence**

Business intelligence applications are a relatively new class of information system that create, package and present information tailored to the needs of different information consumers (Isik, Jones & Sidorova, 2011). BI applications support a wide variety of use cases, including viewing pre-developed reports, executing ad hoc queries and performing knowledge discovery by interacting with data in real time through sophisticated user interfaces, such as Online Analytical Processing (OLAP) or statistical analysis tools (Gessner & Scott, 2009). Business intelligence applications generally obtain data from a data warehouse (O’Leary, 2011). A data warehouse consists of a set of processes that obtain data from other systems, typically operational system from within an organization (but sometimes from external sources), and cleanse, integrate, transform and load it into a database that is designed specifically to support query and reporting (Huang, Liu, & Chang, 2012).

The output from BI applications can be structured or unstructured (Airinei & Berta, 2012). Structured and unstructured documents
Social Capital in Technology-Based Firms
www.igi-global.com/chapter/social-capital-in-technology-based-firms/107407?camid=4v1a