Defining the Possible Approaches to Optimum Metadata Management

Guhathakurta, Rahul

Veröffentlichungsversion / Published Version
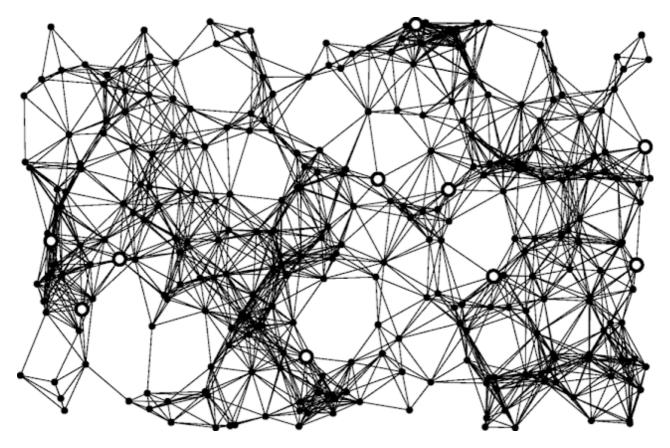Zeitschriftenartikel / journal article

Defining the Possible Approaches to Optimum Metadata Management

IG indrastra.com/2022/03/optimum-metadata-management.html

By Rahul Guhathakurta

In today's data management world, metadata is essential for gaining confidence in your data, making smarter decisions based on it, and unlocking the full potential of your data. Metadata, in simple words, is information about the data itself. Metadata can indicate if a piece of data is confidential *(such as an employee's personal information)*, financial *(such as a credit card number, or bank account number)*, or should be protected *(such as information about a customer's identity).*

Metadata has been the focus of a lot of recent work, both in academia and industry. As more and more electronic data is generated, stored, and managed, metadata generation, storage, and management promise to improve the utilization of that data. Data and metadata are intrinsically linked, hence the concept can be found in any possible application area and can take numerous forms depending on its application context.

However, it is found that metadata is often employed in scientific computations just for the initial data selection; at the most, metadata about query results are recovered after the query has been successfully executed and correlated. As a result, throughout the query processing procedure, a vast amount of information that may be useful for analyzing query results is not utilized. Thus, the data need *"refinements"*.

There are two distinct definitions of *"refinements"*. The first is the addition of qualifiers that clarify or enlarge an element's meaning. While such modifications may be necessary or even necessary for a particular metadata application, for the sake of interoperability, the values of such elements can be regarded as subtypes of a broader element.

The second type of refinement entails the declaration of specific schemes or value sets that define the value range for a particular element. Thus, indicating that a metadata value was chosen from a defined vocabulary or produced using a specific technique may make it far more helpful, particularly for automated processing. By relying on a common value set, semantic compatibility between apps can be increased.

The use of restricted vocabularies is another critical refinement technique that increases the clarity of descriptions and leverages the enormous intellectual capital invested by many domains to improve subject access to resources. For example, the **<u>Dewey Decimal Classification System</u>** provides a multilingual classification system that has been widely utilized in traditional library settings and can be extended to electronic materials as well. Additionally, hundreds of domain-specific thesauri and classification systems can be incorporated into the Web metadata framework to facilitate subject descriptions. By specifying the language to be used in a particular collection of metadata, programs can provide more cohesive search and browsing capabilities. Even if an application is not specifically built to make use of a classification scheme or thesaurus, users may benefit from the inherent coherence provided by such a scheme.

Also, there is a strong tendency for metadata creators to *"fill in all the blanks"*. When an element is available, it is desired that it be used in a description. Applications should be developed in such a way that it is clear that not every accessible piece is necessary for every resource type. Similarly, applications with their dashboards should aid key-users/end-users in selecting a suitable value for a given element *(wherever it is possible)* and to the degree that content production should include capabilities of metadata creation and addition, then only, the application can more accurately identify values for particular elements than the user.

However, please keep in mind that no single set of metadata elements will satisfy the functional requirements of all applications, and as the internet dissolves *"access barriers"*, it becomes increasingly critical to be able to traverse *"internet search and discovery barriers"* as well. This will be facilitated by application profiles, which will enable data science researchers to *"mix and match"* schemas as necessary.

Metadata management *(as a concept and as a tool)* is always going to be a critical component in the creation of more valuable information warehouses. Undoubtedly, the geopolitical policies, organizational agendas, and market pressures will keep on giving new shapes and formats to current and future information repositories, and at the same time generating new opportunities and niches. To fulfill these opportunities, the convergence of encoding formats and uniform semantics is a must.

About the Author

Rahul Guhathakurta (ORCID: **0000-0002-6400-6423**) is a strategic management consultant and is currently affiliated with Anaha Innovations — an Ahmedabad-based technology business incubator and private equity firm. Also, he is a primary investor in IndraStra Global — a US-based publishing company.