Meta Data Validation for Linked Datasets in HTML

Introduction

Hypertext Markup Language [HTML] has an internal structure including a formal <meta> tags section in the <head> of the document. However, meta data can be placed in the <body> of the document. To image or expose the meta data in an HTML document, certain assumptions have to be made about the meta data text, in structural context.

- The GRDDL Assumption
- Other People's Data (aka The External Authority Assumption)
- The Fourier Assumption (Redaction)

The GRDDL Assumption

Gleaning Resource Descriptions from Dialects of Languages [GRDDL] makes the assumption that a Schema (a Collection of meta data terms) is declared in the <ln> tag).

The "profile" attribute of the <head> element is a link to a readable text version of the Schema (Collection).

And furthermore, that the prefix to term names in the <meta> tags is bound to the name space URI of the Collection. In aggregate, the <meta> tag names are an enumeration of the Collection Members in use in the document instance.

Other People's Data

Certain tags in the <body> of an HTML document rely upon external authority for semantic meaning. The existential assumption implies that they *all belong to the same Schema (Collection)* as the other meta data in the document, the Schema given in the <head> profile attribute, and exactly one of the Schema links <ln>.

The veracity of these elements is another matter all together, and external authorities may differ as to the meaning of members of different Vocabulary Encoding Schemes. In fact, the concept of linked data means that an <owl:Thing> in one Vocabulary Encoding Scheme is <owl:sameAs> an <owl:Thing> in a different Vocabulary Encoding Scheme.

At least these HTML elements fall into the category of having an external authority, although other elements are candidates.

HTML Element	*Example Meta Data Element URI*	*Content*
<abbr>	http://purl.org/dc/terms/bibliographicCitation	Abbreviations
<acronym>	http://purl.org/dc/terms/bibliographicCitation	Acronyms
<address>	http://purl.org/dc/terms/bibliographicCitation	Addresses
<cite>	http://purl.org/dc/terms/bibliographicCitation	Citations, References
<dfn>	http://purl.org/dc/terms/bibliographicCitation	Definitions
<rust>	http://purl.org/dc/terms/bibliographicCitation	Redactions – see below [RUST]

The Fourier Assumption (Redaction)

Meta data exists in both visible <body> and hidden <head> styles in an HTML document. The method of data collection has a bearing on the migration of the document meta data from one structural section to the other.

A Fourier Transform and Inverse Fourier Transform give representations of the same "image" in time space and frequency space. The act of reading text is a scan of the time space spectrum. On the other hand, looking at a photograph is a scan of the (light) frequency space spectrum. Both activities are pattern matching, in their own way.

Meta data has width, in both time space and frequency space, regardless of whether it is hidden by style. This leads to spin resonance phenomena and spin coupling in subdivisions of a "thing". In spectroscopy, one recognizes a pattern of "things", then integrates the group to evaluate the pattern in proportion to other integrated groups. Vocabulary Encoding Schemes are the assignment of the integrated value of the group to one Code/Number and so hiding the width in a different way because all patterns integrate to Unity (value=1 'whole group of things'). The Fourier Transform is just a simple way for an Signal Processing Engineer to express the idea that virtual space is largely a matter of Computer Science *seeing* "things" the same way that *look* quite different.

If one were to move an encoding scheme from the <head> to the <body>, there is no problem because the group integral has one value. Moving an encoding scheme from the <body> to the <head> might just be a problem, if the integral value is over a different range of <owl:Thing> component parts.

The relation to redacted information is this:

Information can be redacted by leaving it out of the document <body>, in which case it manifests by being a member of the <head> profile Collection, but not enumerated in a <meta> tag.

Meta data in the <body>, even in (style) hidden attributes, is still "exposed" because it can manifest as multiple profile Collection assignments, possibly redundant and possibly previously unseen (re-identification in the case of Personally Identifiable Information). Zeroing the meta data width is a fundamentally different operation than moving the meta data from <body> to <meta>.

This is a perfectly natural consequence of one meta data class having one unique valued Vocabulary Encoding Scheme where in fact some members are subdivisions, partial sums of other members. The Fourier Assumption is that <u>every member of a Vocabulary Encoding Scheme is both unique valued, and represents a conceptual peer group</u>.

[GRDDL] W3C

[RUST] The <rust> tag (**RUST** = **R**edact **U**nless **S**tatic **T**ext) has the effect of grouping the enclosed information with the (abstract) information not in the document to begin with, or to put it another way, the enclosed information is has no <meta> *enumeration* of the <head> profile yet retains *membership*.