



Amanda Vizedom <amanda.vizedom@gmail.com>

Why is the video property bound to creative work?

martin.hepp@ebusiness-unibw.org <martin.hepp@ebusiness-unibw.org>

Wed, Apr 9, 2014 at 3:23 AM

To: Laura Dawson <Laura.Dawson@bowker.com>

Cc: Vicki Tardif Holland <vtardif@google.com>, Jarno van Driel <jarno@quantumspork.nl>, W3C Web Schemas Task Force <public-vocabs@w3.org>

On 08 Apr 2014, at 18:44, Laura Dawson <Laura.Dawson@bowker.com> wrote:

> That's the job of any given ontology. In the book industry, we have a number of relationships defined already. I'm pretty sure in other media types those also exist, so as to communicate to vendors "this thing replaces that thing; this thing is derived from that thing; this thing has nothing to do with that thing". That's essential information that vendors (e.g. Amazon and Apple) need to have.

Well, I think the question in a *Web* ontology like schema.org is to find a good balance between the specificity of the property/relationship types (or type/class alike), and the ability of publishers to employ the distinctions properly, which is constrained by at least

1. their understanding of the definition of a property (or type/class),
2. their existing data structures (in particular local schemas), and
3. the fitness of the distinction to the context of the user (e.g. what is a common distinction in the library context - like book copy vs. book title - can be harder to understand and apply for users in other contexts - like car model vs. actual car, service template vs. service instance etc.).

So simply taking a set of relationship types from an existing standard is not always the best choice.

It really depends on

1. whether the degree of specificity is compatible with the local schemas of existing databases (i.e. that site owners can populate them with ease), and
2. whether the consumers of data (e.g. search engines) can reconstruct the distinction from contextual information or other signals.

If e.g. search engines can reconstruct a distinction from contextual information or other signals, it is not necessary to encourage publishers of data to spend resources on declaring the respective meta-data in mark-up.

In a Web vocabulary like schema.org, we should thus focus on those classes/types and properties that

1. are easy to apply reliably from existing data sources by typical site owners and
2. that cannot be reconstructed with ease by a consumer of the data (*).

Since adding type and structural information to Web data has costs, we should further center our efforts on such meta-data that

1. has a high information entropy (approximately: is useful, additional information) and
2. can be reliably provided by a large number of sites.

RDF and its proponents made, IMO, the mistake that they also expected such meta-data explicitly in the data representation that a client can often reliably reconstruct from the data alone. For example, data types for literals like the information that the literal "ABC123" is a string or that "2014-04-09" is likely a date. (The former has been fixed in RDF 1.1, I know ;-))

Many Web ontologies on the contrary (and maybe even GoodRelations in some of its branches) provide conceptual distinctions that cannot be reliably applied by site owners, either due to fact that the distinctions require a deep philosophical understanding, or because the existing databases available for powering a Web site simply do not support the distinction.

As for schema.org, I would say that as long as we are defining properties for a single schema.org type or a small set of types, this problem is a lesser issue, since the type provides context and people will more easily understand even distinctions (like author vs. editor for a book). But when we speak about properties at a more generic level, or even for Thing, we need to be super-careful.

Best wishes

Martin

(*) Of course, this depends on the power of the data consumer - Google et al. can likely do way more with messy, partly structured data than a smartphone app or browser extension. One untested and undeclared working assumption of the Semantic Web advocates is that data on the Web can ever be consumed in its raw form by relatively dumb clients. I have argued elsewhere [1] that structured data in Web content is rather "proto-data" that can be turned into usable data by heavy post-processing.

[1] <http://lists.w3.org/Archives/Public/public-vocabs/2013Oct/0293.html>

martin hepp
e-business & web science research group
universitaet der bundeswehr muenchen

e-mail: martin.hepp@unibw.de
phone: +49-(0)89-6004-4217
fax: +49-(0)89-6004-4620
www: <http://www.unibw.de/ebusiness/> (group)
<http://www.heppnetz.de/> (personal)
skype: mfhepp
twitter: mfhepp

Check out GoodRelations for E-Commerce on the Web of Linked Data!

=====

* Project Main Page: <http://purl.org/goodrelations/>

[Quoted text hidden]