

Extending Vocabularies

The Rack and the Weeds

Liam Quin

Delightful Computing

liam@fromoldbooks.org

What

- Let's characterize customizations
- Suggest terminology for extensions
- Review what works (and what doesn't)
- **New!** How could we do this better?
 - shortcomings in process & outcome
 - some new ideas that might be dangerous



What's an Extension?

- A systematic variation in how a vocabulary is used.
More than a *personal customization*: that is, more than just a change in configuration that's not shared.
- Extending can include adding to, or subtracting from, the scope of a vocabulary.

Context, Situation, Scope

- The *context* in which a vocabulary is used: technical, social, business...
- The *situations* in which a vocabulary is used: the immediate goals...
- The *scope* encompasses all of these.

Not this sort of scope.



Why extend a vocabulary?

“Even simple problems benefit from thinking about the problem and careful planning”

– Erik Spiekermann

Why Extend...

- Save time (billboard view: fewer element choices means more freedom)
- Do something you couldn't do before
- Happiness and Security (Linux blanket)
- Compatibility (previous versions, software, trading partners...)

Extensionology

- Let's define some terminology to describe customizations and extensions;
- Then let's apply those terms to some widely-used vocabularies;
- After that let's look briefly at patterns that do and don't work well.

Planned-For Changes

- Grammar Hooks (%elephants;)
- Unchecked Islands (#ANY)
- Extension Names (X-Socks)

The Unplanned

- Altered Grammar
- Usage Conventions
- Unchecked Usage

Hybrid and Absorbed Extensions

- Ambiguous Markup (eg. chameleon xsd)
- New Vocabulary Features Adopted
- Usage Conventions Adopted (incl. subsets)
- Internal Formats *vs* Interchange



Kinds of Extension

- Functional: New or Changed Behaviour
- Semantic Coverage: New Meanings
- Explicit Extensions (e.g. namespaces)
- Implicit Extensions (folksonomy)
- Usage Conventions

Methods

- Add elements or attributes
- Add new content (e.g. translations)
- Adding new values to enumerations
- Subtractions
- Xreoles: combining vocabularies
- Adaptations: payback for the inflexible

Some examples

- TEI - explicit grammar: ODD Files for extensions
- DocBook, DITA - implicit grammar: relaxng
- JATS - extension elements plus implicit grammar (parameter entities)
- HTML - names (data-*), folksonomy (@class) and unchecked (custom elements)



Some problems & pitfalls

- HTML originally said the first unknown element ends the head. So, can never add new elements to the document head.
- XML at one point said parsers were to reject unknown versions; forbidding a 1.0 parser from reading a 1.1 document was a problem.

Avoiding the pitfalls

- DocBook's @version attribute allows
`version="5.2 with blue socks"`
- ARIA roles can be extended with a prefix, rather like data-* in HTML (but no central registry)

Dangerous Idea

- What about a markup language explicitly for describing extensions?
 - specify changes to grammars
 - include stylesheets & other resources
 - space for context & documentations & links to specs

Questions

The background of the slide features a low-angle shot of several trees, likely pines or cypresses, silhouetted against a bright blue sky filled with soft, white clouds. The trees' branches are dark and intricate, creating a stark contrast with the light sky.

Liam Quin

www.DelightfulComputing.com

liam@fromoldbooks.org