

In put ValiD ation

Chris Cosner
Stanford University Press

Why Validate Input?

Specific

- ONIX
- XML to partners
- Website
- Catalog

General

- Security
- **Clean data**
- **Clean exports**
- **Sanity**

Data types

int

real

text

char

date

boolean

timestamp

interval

time

▪

▪

▪

Data types

Good data

01/08/2009

Literary Theory -- Russian
(212) 234-5678

herrdrprof@wichtiguni.de

35.00

<I>Tom Sawyer</I>, by Mark
Twain

value lists

Evil data

Jan09

Lit theory-rUssia
1 212234-567

herrdrprof.w uni.de

35

Tom **Sawyer**, by Mark
Twain

free-form text

Pattern Language

Characterized by

1. Isolating common problems
 2. Describing solutions
 3. Helping you approach problems logically
 4. Allowing for multiple solutions to the same problem.
- See http://en.wikipedia.org/wiki/Pattern_language

CHECKS Pattern Language

by Ward Cunningham

- Selected patterns
 - Whole Value
 - Exceptional Value
 - Meaningless Behavior
 - Deferred Validation
 - [Implications]
 - Instant Projection
 - Hypothetical Publication
 - Diagnostic Query

CHECKS Pattern Language

by Ward Cunningham

- **Whole Value** [human readable/writable]

CHECKS Pattern Language

by Ward Cunningham

- **Exceptional Value** [other, fuzzy, unexpected]

CHECKS Pattern Language

by Ward Cunningham

- **Meaningless Behavior**
 - Inappropriate cases
 - large text fields
 - multiple inputs (which one did it?)
 - Also manifests as
 - silent rejection
 - buttons that do nothing
 - navigation prison
 - Helped by
 - highlighting incorrect or missing values
 - other visual cues

CHECKS Pattern Language

by Ward Cunningham

- Selected patterns
 - Deferred Validation [stay out of my way]

CHECKS Pattern Language

by Ward Cunningham

- Selected patterns
 - Implications
 - Instant Projection
 - Hypothetical Publication
 - Sandbox [do what you like]

CHECKS Pattern Language

by Ward Cunningham

- Selected patterns
 - **Diagnostic Query** [show all]

Filtering

- One more...
 - Filtering [replace/remove]

Suggestions

- Document your data structures.
 - Data dictionary
 - ER diagrams
- Random audits. What if there is a problem with X?
 - how does it get filled initially?
 - how can it be modified?
 - where does it end up?
 - where are the logs for this piece of data?
 - what validation methods are imposed on input and export?
- Match good data rather than attempting to define every possible bad.

Links and Further Suggestions

If you give me your business card or email address, I can email the slides and text to you directly. They should also eventually be available online with the rest of the conference materials.

Comments or Questions?

ccosner@stanford.edu