# Using XMLSchema

## A Schema for Document Semeiotics

Dr. Steven Ericsson-Zenith
SEMEIOSIS RESEARCH
Sunnyvale, California
June 21st, 2005

# Some Relevant History

- Involved in the design of programming languages
    – Occam, Linda, Ease, MPI
- Spent time reviewing the SAP source :-(
- Oracle Alumni – Network Computer
- Spent several years in strategic development with Microsoft and licensing technology to them
- Never owned by anyone
- Not even Larry :-)

# General Interests

- Apprehension
- Meaning
- Formalization
- Engineering Practices
- Engineering Behavior
- Mathematical Foundations
- Prediction / Inference
- Deliberation
- Consensus
- Generalization
- Complexity

# PLATFORM

- SQL
- pl/pgSQL
- PHP/Smarty
- Python
- JavaScript
- Exim
- XML
- XHTML
- Apache
- OpenSSL
- PostgreSQL

- Not Java
- Not .NET

# Scenario
# Open Content Development

- Content development in highly creative environments
  - Semiconductor and software design
  - Media property development
  - Crisis and disaster management environments
- Document publishing
  - Multiple formats
  - Alternative forms
- Decision making
  - Pragmatics: what it all means

# Issues Addressed

- Immediacy
  - States: content origin and change
  - Declarations: draft, publish, rescind, delete
- Clarify Intent
  - Glossaries, not dictionaries
  - Concept distinction
  - References
- Formal Deliberation
  - Current and historical deliberation data
    - Goals
    - How to participate
    - Conclusions

# Subschemas

- Publisher
- Author
- Document Summary
- Immediacy
  - States
  - Declarations
- Document Structure
  - Annotated
- Glossary
- Concepts

- Bibliography
- Citations
- Recent Updates
- Syndication
- Formats
- Alternates
- Audience
- Journal
- Deliberation
- Provenance
- Include

# Schema Usage

- Document publishers like www.memeio.com
  - Document metadata presentation
  - Document summary
  - Alternates: an executive summary for example
- Aggregators like FirstSay (www.firstsay.org)
  - A view of recent and relevant content
  - A view of concept development
  - A view of recent analysis
- Applications
  - Executive descision
  - Creative development

# Schema and Instance Licensing

- Rights issues of instance discovery
  - What can I do with the data I find?
  - Who violates? Producer or consumer?
- Authentication
  - Web service usage rights.
- Relates to data privacy issues
  - Who is responsible for security of instances
  - What happens if an instance "escapes"
    - Is the mechanical aggregator liable?

# Observations

- Instance Validity / Data Validity
  - Assertions and constraints
    - Data may be 'valid' while instance is not
      - Assert e < e'
      - Schema specified type constraints
    - Instance may be valid while data is not
      - If e = x then e' required
      - Is the communication of rules
      - Informs using application of data validity while maintaining instance validity
  - Element valid iff e > e' || e < e' || e = e' etc...
- Data source requirements
  - Source must comply (state vs declaration)

# Observations

- Instance Design
  - Element / Attribute overloading
  - Why one and not the other?
  - Recursive Types: must be maxOccurs = 0
- Formalize instance trees
  - For data decompositon and sharing
- Keep it simple