

Comp 336/436 - Markup Languages

Fall Semester 2017 - Week 11

Dr Nick Hayward

Contents

- XML - XPath & XSLT tests
- functions
- tests
- Metadata typology and uses - part I

XML - XPath details - functions - total values

- additional mathematical functions
 - e.g. *sum()*
- use *sum()* to total all values in a selected node set
- use the following pattern to *sum()* values
 - *add sum(*
 - *add path to required node set with values*
 - *add)*
- use *sum()* with division to get initial average
 - use *count()* for total in node set
 - use *format-number()* to structure output for rendering

XML - working example - ancient sites - total values

XML

```
<dimensions>
  <width type="average" unit="metre">230.360</width>
  <height type="original" unit="metre">146.59</height>
  <height type="current" unit="metre">138.75</height>
</dimensions>
```

XSL

```
...
<xsl:value-of select="sum(./height) div count(./height)"/>
...
```

- Demo - Ancient Sites I4

XML - XPath & XSLT tests - functions - total values

Exercise - part 14

- update your XSL stylesheet
 - *add an option to total some values, and then output the average, in your XML*
 - add new values to XML, if necessary
 - *add output to rendered document*
- test stylesheet with XML file

~ 10 minutes

XML - XPath details - functions - extras - node functions

- many useful extra node functions
- `name (node-set)`
 - *returns name of first node in specified node set*
- `name ()`
 - *returns name of current node*
- `id (id-str)`
 - *use with unique IDs specified in DTD &c.*
 - *returns all elements with an ID equal to `id-str`*
 - *get only certain elements with a given ID...*
 - *e.g. all first pages from each chapter...*
- full list of the functions in XPath Version 1.0
 - *XPath Version 1.0 functions*

XML - XPath details - functions - extras - string functions

- many useful extra string functions
- `contains(str1, str2)`
 - *returns True if `str1` contains `str2`*
 - *otherwise returns False*
- `string-length(str1)`
 - *returns the number of characters in `str1`*
- `string-length()`
 - *returns the number of characters in the current node*
- `normalize-space(str1)`
 - *returns `str1` with all leading and trailing white space removed*
 - *sequences of white space replaced with a single space*
- `normalize-space()`
 - *performs same action on current node*

XML - XPath details - functions - extras - boolean functions

- boolean functions as well, e.g.
- `not (expression)`
 - *returns `True` if expression evaluates to `False`*
 - *returns `False` if expression evaluates to `True`*
- further details on functions for XPath 2
 - *XPath 2 functions*

XML - working example - ancient sites - add some images

XSL

```
...  
<img>  
  <xsl:attribute name="src">  
    <xsl:value-of select="image[@type='jpg' and @size='thumb']/@url"/>  
  </xsl:attribute>  
</img>  
...
```

- Demo - Ancient Sites I5

XML - XPath & XSLT tests - functions - add some images

Exercise - part 15

- update your XSL stylesheet
 - *output some thumbnail images*
 - *wrap a link (anchor element) around at least one thumbnail image*
 - *add output to rendered document*
- test stylesheet with XML file

~ 10 minutes

XML - XPath & XSLT tests - conclusion

Exercise - Working Demo

- tidy up the code
- add some headings and structure to HTML output
- add some CSS styling
- test rendering and output

~ 10 to 15 minutes if necessary

XML & Semantic Web

- semantic web as a broad grouping of organised data
- different XML-based recommendations
- RDF (Resource Description Framework)
 - *XML based text format*
 - *supports resource description and metadata apps*
- GRDDL (Gleaning Resource Descriptions from Dialects of Languages)
 - *introduces markup based on existing standards*
 - *declare XML document with data compatible with RDF*
 - *add linking to algorithms - usually as XSLT*
 - *used to extract data from document...*
- OWL (Web Ontology Language)
 - *semantic markup language for publishing and sharing ontologies*
 - *OWL is a vocabulary extension of RDF*
 - *derived from DAML+OIL Web Ontology Language*
- SPARQL
 - *query language for RDF*
 - *used to express queries across diverse data sources*

Metadata typology and uses - intro

the main goal of metadata, whether considered as a theory or implemented in a functional system, is to make data useful...

- Garoufallou, E., Greenberg, J., *Metadata and Semantics Research: 7th International Conference*. MSTR. 2013.
- scholars and practitioners need to ascribe types to things
- classify kinds of *things*
 - *enables better understanding of variety and complexity...*
- need to typify metadata
 - *also metadata schemas, metadata uses, metadata elements, metadata values...*
- **types of metadata** includes different concepts
 - *types of metadata standards*
 - *types of metadata values*
 - *types of metadata functions*
 - *and types of metadata elements*

Metadata typology and uses - which types?

- common metadata usage is *Dublin Core*
 - *often perceived as a catch-all solution...*
- many different options available for metadata, e.g.
 - *metadata for preservation*
 - *metadata for intellectual property rights*
 - *METS*
 - *metadata for geospatial information*
 - *AACR*
 - *metadata for describing music*
 - *FRBR, LOM, RDA, RDF...*
- often use a mixture of models, domains, uses...
 - *disparate parts forming our understanding of metadata*

Metadata typology and uses - typify

- need *criteria* to typify something
 - *and potentially everything in a given domain...*
- simple term *metadata* and simple definition, *data about data*
 - *encompass a complex technical and intellectual infrastructure*
 - *used to manage and retrieve digital objects in different digital contexts*
 - *within different digital information systems and services*
- criteria often matches sheer diversity of digital information management systems
- could be applied to different metadata levels, e.g.
 - *records, agents, elements, schemas...*

Metadata typology and uses - classify metadata

- metadata may describe different attributes or properties of information objects
 - *giving them meaning, context and organisation in a standardised way*
- many uses and dimensions of metadata
 - *has now led to construction of a very broad typology*
- use obvious criteria to typify metadata
- might classify metadata, mixing different levels, e.g.
 - *systems, practices, models, schemas, elements, records and trends*
- mix different levels with seven criteria, e.g.
 - *criterion 1: way of creation...*
 - *criterion 2: moment of creation...*
 - *criterion 3: way of storage...*
 - *criterion 4: level of structure...*
 - *criterion 5: purpose of the metadata...*
 - domain independent metadata
 - domain dependent metadata
 - *criterion 6: application...*
 - i.e. what's metadata used for?
 - *criterion 7: level of standardisation*

Metadata typology and uses - types and systems

- often consider metadata types to help understand associated metadata systems
- need to distinguish different types of metadata
 - *to help categorise metadata*
- classification can mix different levels to categorise metadata, e.g.
 - *record level, element level or schema/element set level*
- we often consider the following
 - *general versus specialist*
 - *minimalist versus rich*
 - *hierarchical versus linear*
 - *structured versus unstructured*
 - *machine generated versus human authored*
 - *embedded versus detached*
- **surface** metadata
 - *information that can be gathered by machines and converted into metadata*
 - *process of gathering it is often known as **screen scraping***
- *other types of metadata*
 - *including keywords, Google, tags, user created metadata, &c.*
 - *self-description mechanisms for social tagging and new trends*
 - *avoid keyword stacking*
 - *Kartus, E. 2006.*

Metadata typology and uses - types and vocabularies

- purpose of any metadata, whether element or model
 - *describe a property or set of useful properties*
 - *for a given information resource or object*
- many current metadata systems include an inherent multidimensional nature
 - *such structures often called metadata vocabularies*
- a few useful definitions
 - *metadata models/formats help express properties of a resource, e.g. a subject*
 - *metadata schemas = metadata formats encoded in a standardised machine readable markup language*
 - *metadata scheme = a set of rules or terms for encoding the value of a particular metadata term*
 - *both schemas and schemes involve metadata elements*

Metadata typology and uses - formats and schemas

- in the 1990's, common to speak about *metadata formats* or *metadata models*
 - to refer to a set of properties
 - expressed and defined in a standardised way
 - served to describe a digital information object
- each description constituted a metadata record
 - applying that metadata format to a particular object
- grouped metadata formats along a continuum of growing structural and semantic richness
 - used to identify shared characteristics of each grouping
 - patterns from paths of the given metadata formats
 - Dempsey, L. & Heery, R. 1998.

Metadata typology and uses - schema/scheme

- DCMI Glossary definition for *schema/scheme*

any organization, coding, outline or plan of concepts. In terms of metadata, a systematic, orderly combination of elements or terms.

- when a declaration of metadata terms is represented in XML or RDF schema language
 - *it might be better considered strictly a schema*
- schemas are machine-processable specifications
 - *define structure and syntax of metadata specifications in a formal way*
- difficult to think of a single metadata element or a single metadata value
 - *in isolation, without broader context*
- current Semantic Web approach
 - *considers schemas and schemes*
 - *assumes they are formalised in a machine-understandable way*
- *schemas* are sets of metadata elements and rules
 - *previously stated for a particular purpose*
- *schemes* are a set of rules for encoding information
 - *supports a specific information domain*
- schemas and schemes are commonly expressed in XML/RDF vocabularies
 - *easier to read and use*
- also form a perceived complete metadata infrastructure
 - *for current digital information services*
 - *becoming Metadata Vocabularies*
- metadata classification starting from metadata vocabularies
 - *metadata schemas, metadata schemes, and metadata elements*

Metadata typology and uses - types of schemas

- current metadata schemas include
 - *not only the semantics of an entire element set*
 - *but also encoding of elements and structure with a markup language*
- two crucial criteria to take into account
 - *to categorise the complexity and variety of current metadata*
 - **granularity** and the **information domain**

Metadata typology and uses - types of schemas - granularity

- draw up a typology of metadata schemas
 - *based on hierarchical relations between different sets of metadata*
- extensible nature of metadata, e.g.
 - *global metadata schemas:*
 - e.g. Dublin Core, which is the best known example
 - *local metadata schemas:*
 - e.g. SEPIADES, Europeana Semantic Elements...
 - *container metadata schemas:*
 - container architectures
 - e.g. RDF, METS, ONIX, MARC, &c.
 - *conceptual metadata schemas*
 - e.g. CIDOC-CRM

Metadata typology and uses - types of schemas - domain

- an example of domain specific is *Web-based information*
 - *sectioned in a vertical way*
- there are global search engines, and very large digital libraries like Europeana
- increasingly, there are also different portals
 - *digital libraries*
 - *and other digital information systems and services*
- locations where the objects you can find belong to a category
 - *either subject-oriented or type-oriented*
- a few domain specific examples
 - *metadata for Cultural Heritage*
 - *metadata for geographic and geospatial information systems*
 - *metadata for educational information systems*
 - *metadata for digital preservation/curation*

Metadata typology and uses - types of schemas - domain examples

metadata for Cultural Heritage

- often applied to cultural objects and visual resources
- in this huge information context, many early developed metadata schemas
- schemas range from those influenced by the librarian's domain
 - *to those built from archives and museums*
 - *or arts and architecture domains...*
- for example,
 - *traditional cataloging standards converted into schemas - MARCXML, MODS...*
 - *metadata standards for Finding Aids - Encoding Archival Description (EAD)*
 - *representation of digital texts - Text Encoding Initiative (TEI)*
 - *metadata schemas/standards for digital visual arts - CDWA...*

Metadata typology and uses - types of schemas - domain examples

metadata for geographic and geospatial information systems

- metadata in this particular domain has a clear protagonist, and almost standard
- *Content Standard for Digital Geospatial Metadata (CSDGM)*
 - *originally adopted for the Federal Geographic Data Committee in 1994*
 - *and then revised in 1998*
- its main elements were embraced by the international community
- adopted through the ISO metadata standard *ISO 19115*

Metadata typology and uses - types of schemas - domain examples

metadata for educational information systems

- another big domain to analyse specific metadata schemas and standards
 - *the educational and learning environment*
- *Learning Objects Metadata (LOM)*
 - *developed by the IEEE Learning Technology Standards Committee*
 - *the main metadata schema in this field*
- learning objects used and reused for educational purposes
 - *a particular type of digital information object*
 - *requiring particular types of metadata schemas*
 - *helps describe their education-specific properties*
- *Learning Objects Metadata* has a more complicated structure than many other schemas
 - *integrates different kinds of metadata element*
 - *e.g. descriptive, administrative, and technical elements*

Metadata typology and uses - types of schemas - domain examples

metadata for digital preservation/curation

- digital preservation and curation is a broad domain and concept
- implies much more than metadata
 - *a broad range of systems and infrastructures*
 - *designed to preserve the usable life of digital objects*
- preservation metadata is a domain
 - *supports all processes associated with digital preservation*
- primary examples for preservation metadata as a specific domain
 - *Open Archival Information Systems (OAIS (ISO 14721))*
 - *PREMIS (Preservation Metadata: Implementation Strategies)*

Metadata typology and uses - types of schemas - domain

- domain approach to digital information
 - *increasingly manifest throughout digital information services*
- besides big metadata schemas based upon a specific domain
 - *there are many application specific profiles*
- these are data elements drawn from other metadata schemas
 - *then combined together*
 - *and used for more specific or even local application domains*
 - *e.g. for public sector information*
- also find application profiles based on one schema
 - *but tailored for a particular information community*
 - *e.g. Dublin Core Education Application Profile*

Metadata typology and uses - types of schemes

- schemes describe a particular way to encode *metainformation* describing a given resource
 - *similar to Anglo American Cataloging Rules (AACR), Cataloging Cultural Objects (CCO)...*
- schemes are also sets of terms or vocabularies
 - *only concern the possible values for a particular metadata element*
- *metadata schemes* are a range of values
 - *values that might be provided for an assertion about a resource*
 - *e.g. Date-time formats, authority lists, controlled vocabularies...*

Metadata typology and uses - types of schemes - examples

- *authority list* is a metadata scheme
 - *could be applied to encode the values of a metadata element*
 - *or term in a metadata schema*
- e.g. if we're dealing with a resource's authorship
 - *the element `DC.Creator` in Dublin Core*
 - *or `Author` in TEI Header*
 - ...
- large-scale metadata content standards such as *Cataloging Cultural Objects (CCO)*
 - *include general instructions applicable to a metadata model or a set of its elements*
- metadata schemes are specific vocabularies
 - *devoted to the values of a particular metadata element*
- encoding schemes provide contextual information
 - *or parsing rules that aid in the interpretation of an element value*
- such contextual information may take different forms, e.g.
 - *controlled vocabularies, formal notations, or parsing rules*
- schemes might also be called *value spaces*
 - *forming the set of possible values for a given type of data*

Metadata typology and uses - types of schemes - domain

- as with schemas, schemes may be classified as well
- *general purpose schemes*
 - e.g. *universal classifications or generic subject heading lists...*
- *specific purpose schemes*
 - e.g. *traditional metadata schemes*
 - *thesauri, other vocabularies applied to a specific information domain...*
- many, many vocabularies encoded as metadata schemes...
 - e.g. *metadata vocabularies for cultural heritage*
 - *Union List of Artist Names (ULAN)*
 - *Arts and Architecture Thesaurus (AAT)*
 - ...
- schemes could be
 - *traditional thesauri*
 - *classification schemes*
 - *and other knowledge organisation systems...*
- schemes could also be
 - *subject-based metadata like ontologies and folksonomies*
 - *traditional vocabularies encoded for the Semantic Web*
 - e.g. *following a formal schema like Simple Knowledge Organization System (SKOS)*

Metadata typology and uses - types of schemes - ontologies

- most of the ontologies we find, for example, are also domain oriented
- they divide the realm of knowledge that they represent into:
 - *individuals*
 - *classes*
 - *attributes*
 - *relations*
 - *and events*

Metadata typology and uses - types of schemes - folksonomies

- semantically weaker *folksonomies* does not usually have a domain orientation
 - *they do not have any kind of control*
- a *folksonomy* is a record with labels, tags or keywords
 - *used by many people on the Web*
 - *usually without a particular purpose or initial structure*
- a *folksonomy* will be more general
 - *dynamic in its growth and application*
- some experiences using social tagging within a specific domain
 - *e.g. Steve Museum - a systematic research project*
 - *considers how social tagging can best serve the museum community and its visitors*
 - *Steve Museum Project*
- in a distributed networked environment - rapid scheme changes
 - *managing semantics of changes vital to functioning and utility of schemes*

Metadata typology and uses - types of metadata - part I

- many traditional classifications of metadata types are based on metadata elements
 - *assuming that every metadata schema has elements of similar types*
- allows fair comparison with known points of comparison
- almost every metadata handbook distinguish the following types of metadata
 - **Descriptive** metadata elements
 - **Structural** metadata elements
 - **Administrative** metadata elements
- types and functions of metadata elements classified into broad categories
 - *descriptive, structural, and administrative*
 - *they do not have well-defined boundaries and often overlap*
- different types blend into one another when using a specific metadata schema

Metadata typology and uses - types of metadata - part 2

- in general, metadata schemas and standards include these types of metadata elements
- e.g., **METS** packages structural, descriptive, administrative, and other metadata
 - *with an information object or digital surrogate*
 - *indicates types of relationships*
 - *e.g amongst different parts of the current complex information object*
- so, any metadata schema can classify their elements as different types of metadata
- a traditional categorisation using this pattern
 - *Dublin Core elements division*
 - *distinguishes three groups for its metadata elements*
- *Dublin Core groupings*
 - *Content elements - Title, Subject, Description, Source, Language, Relation, and Coverage*
 - *Intellectual Property elements - Creator, Publisher, Contributor, Rights*
 - *Instantiation elements - Date, Type, Format, and Identifier*

Metadata typology and uses - user needs - part I

Setting metadata loose, through the internet and their widening user base, has in some cases resulted in new user contexts for existing metadata. The biodiversity domain provides an excellent illustration of how metadata which previously might have had a strong local focus and user community can attract a wider interest on the internet. International collaborative efforts, such as the European Network for Biodiversity Information (ENBI), are currently building large metadata repositories by aggregating metadata of local institutions through portals. Existing metadata are thus repurposed in the context of the international research on global warming and its impact on biodiversity.

- Dempsey, L and R Heery. *Metadata: A current view of practice and issues*. Journal of Documentation, 54(2). PP.145–172. 1998.

Metadata typology and uses - user needs - part 2

- user needs regarding cultural heritage usually defined in general and vague terms
- museums, libraries, and archives have struggled to identify user needs
 - *lots of digitised material to access, not sure how it's being accessed...*
- user needs have attracted attention on a research level and within individual projects
- application of research outcomes and recommendations in the field remains problematic
- metadata practitioners rarely have chance to start from scratch
 - *issues with reviews, surveys, legacy requirements and standards...*
- metadata services rarely built from scratch...

Demos

XML & XSLT - Part 2 - Functions

- Ancient Sites - total values - part 14
- Ancient Sites - node functions - part 15

References

- Content Standard for Digital Geospatial Metadata
- Dempsey, L and R Heery. *Metadata: A current view of practice and issues*. Journal of Documentation, 54(2). PP.145–172. 1998.
- Dublin Core
- Europeana Semantic Elements
- Garoufallou, E., Greenberg, J., *Metadata and Semantics Research: 7th International Conference*. MSTR. 2013.
- Kartus, E. *Types of metadata*. University of Melbourne. 2006.
http://www.infodiv.unimelb.edu.au/metadata/add_info.html
- Learning Objects Metadata
- Open Archival Information Systems
- PREMIS
- Simple Knowledge Organization System
- Steve Museum Project
- XPath Version 1.0 functions
- XPath Version 2 functions