

XML Topic Maps (XTM) for Digital Libraries

Roxana Grigoras
Cashmere-int Workshop

Content



- Situation in many digital libraries
- XML
- Topic Maps
- XTM
- Case Study
- Earlier work

Curent situation



- Historically, libraries have served as centralized repositories of information.
- e.g. books, films, scientific papers
- Catalogues are designed for giving access to these repositories.

- Catalogues are structured using simple XML.
- e.g. simple xml digital library record:

- `<?xml version="1.0" encoding="ISO-8859-1"?>`
- `<?xml-stylesheet type="text/xsl" href="example1.xsl"?>`
- `<lib>`
- `<item>`
- `<identifier>omn:1</identifier>`
- `<creator>Guenther Specht, Michael G. Bauer</creator>`
- `<title>OMNIS/2 - A Multimedia Meta System for existing Digital Libraries</title>`
- `<subject>`
- `<publisher>Springer Verlag, Berlin Heidelberg</publisher>`
- `<date>2000</date>`
- `</subject>`
- `</item>`
- `</lib>`

Topic Maps (TMs)

- TM - index of information overlay which can be constructed separate from a set of resources, identifying instances of subjects and relationships within the set of resources
- its purpose - to convey knowledge about resources through a superimposed layer (or map) of resources
- TM captures the subjects of which resources speak, and the relationships between subjects, in a way that is implementation-independent.

- key concepts – *topics*, *associations* and *occurrences*
- *topic* – resource within the computer that stands in for (or “reifies”) some real-world object; they can have names
- Topics can also have *occurrences* - information resources considered to be relevant in some way to their subject.
- They can participate in relationships (*associations*), in which they play roles as members.

TMs



- Topics have three kinds of characteristics: *names*, *occurrences* and *roles* played as members of associations.
- the assignment of such characteristics - valid within a certain scope (context)
- TMs can be merged.
- Merging can be made at the discretion of the user or application, or may be indicated by the TM's author while being created.

- TMs help navigating enormous amount of information; it provide a simple and flexible data model to capture knowledge and a standard interchange syntax that makes knowledge accessible across applications.
- *easy to search information* with

- the standard XML interchange format for TMs
- purpose – adapting the TM standard (ISO 13250) for use with XML and the Web
- Provides an alternative XML-based syntax for expressing TMs and clarifies certain concepts.

NZETC (New Zealand Electronic Text Centre)



- Collection of this Digital Library can be searched by topics like author, work, projects, etc.
- completely TM – based
- <http://www.nzetc.org>:

Search

Site search

Browse Collection By...

Authors

Works

Projects...



Victoria University of Wellington History



The Print History Project



Golder Poetry



Kōtare



Welcome to the New Zealand Electronic Text Centre's free online archive of New Zealand and Pacific Islands texts and heritage materials, which offers an ever expanding set of images and full-text books, manuscripts and journals. All resources are fully searchable and delivered through an open source framework, and users may choose from multiple formats for download or online browsing. Please read further for more information on the [NZETC's mission](#) and digitisation activities.

Te Ao Hou Online

The NZETC has recently assisted the National Library of New Zealand to make Te Ao Hou available online (external site). Te Ao Hou / The New World (1952-76) was a bilingual quarterly published by the Maori Affairs Department, and printed by Pegasus

War History Project



Published by the New Zealand Department of Internal Affairs in the decades after the end of the war,

this fifty volume series covers all areas of

Case Study

- The website is a digital library, providing access to a couple of hundred digitised books and manuscripts.
- The source material for the site is a collection of TEI (Text Encoding for Interchange) XML files, each of which is an encoding of a source object (i.e. a book).
- Most of the topic map is harvested from these files using XSLT.

- Each book, chapter, subsection, figure, author, publisher, etc, is represented by a topic, names are harvested from headings and captions in the text, and the containment hierarchy is represented by associations.
- These associations are used to generate tables of contents, as well as to provide "next" and "previous" links between web pages.

[HOME](#)[ADVANCED SEARCH](#)[ABOUT](#)

Site search

Browse Collection By...

[Authors](#)[Works](#)[Projects...](#)

Victoria University of Wellington History



The Print History Project



Golder Poetry



Kōtare



AUTHORS

A B C D E F G H K L M N P R S T U W

A

Agnew, Alastair (author of 2 texts)

Aitchison, Jo (author of 3 texts)

Aitchison, Johanna (author of 2 texts)

Alessio, Dominic (author of The Great Romance, by The Inhabitant)

Allan, Guy (author of GUY ALLAN)

Allan, Rob (author of ROB ALLAN)

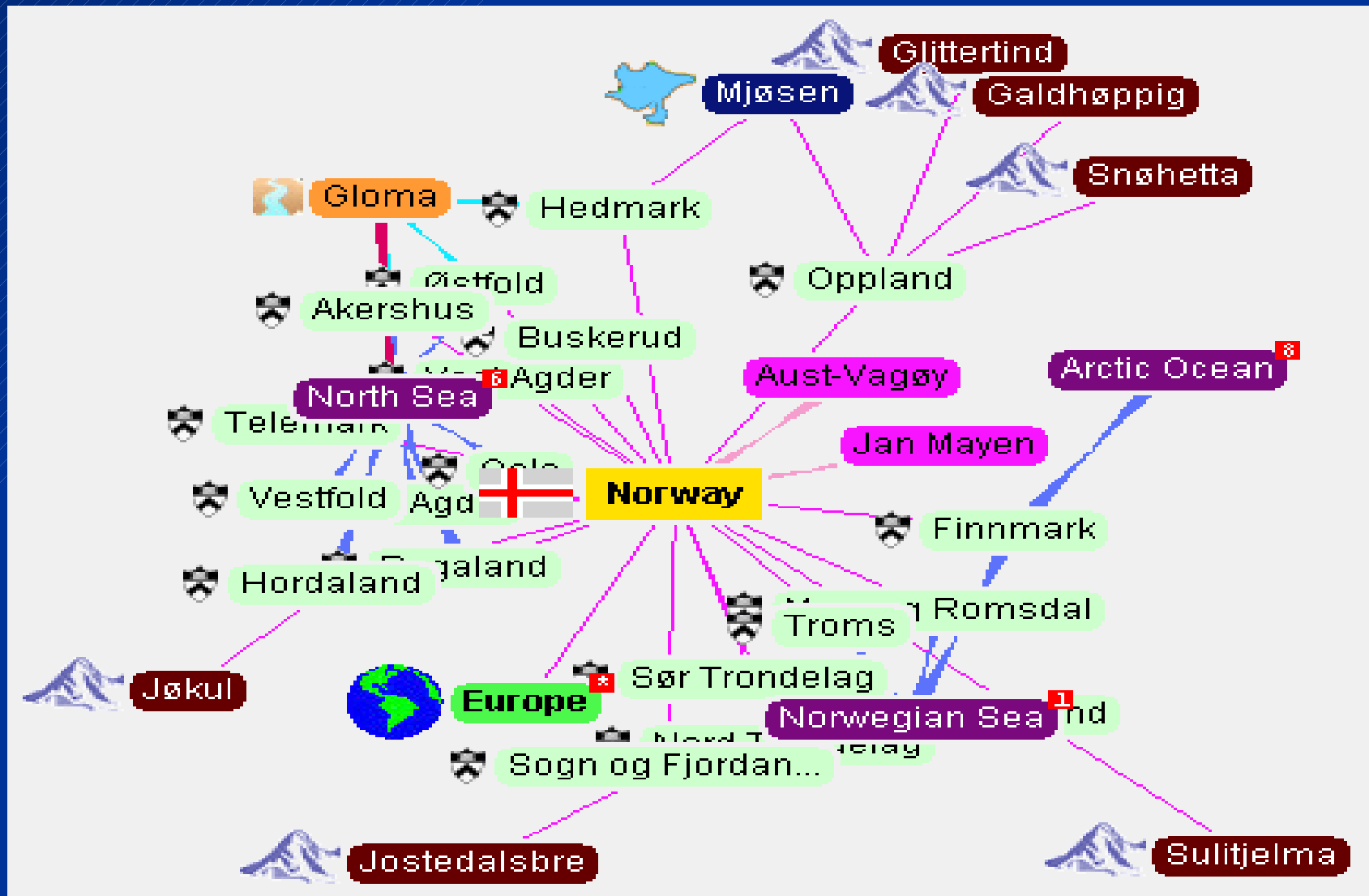
Allpress, Susan (author of 2 texts)

Anderson, Barbara (author of 3 texts)

Anson, T. V. (author of The New Zealand Dental Services)



Part of a topic map: geographic information displayed with Vizigator



Earlier work



- In earlier work we described how traditional XML data entries could be transformed with XSLT in RDF/XML.
- This technique could be used in future work to transform traditional XML data entries of digital libraries (like ex. from slide 4) to XTM data entries, which would allow to navigate the digital library on topics like author, field, resource type (book, paper, film, etc.).

Bibliography



- <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>
- <http://www.xml.com/pub/a/2002/09/11/topicmaps.html?page=3>
- <http://www.nzetc.org/tm/scholarly/tei-NZETC-About.html>
- <http://www.isotopicmaps.org/sam/sam-xtm/>
- <http://xml.coverpages.org/topicMaps.html>

Thank you 😊