Applying the semantic web to the Arnamagnæan manuscript collection

Project funded by the Nordunet2 research programme

Gardar Gudgeirsson - Raqoon ehf - gardarg@raqoon.com - http://www.raqoon.com

Project participants

- Stofnun Árna Magnússonar á Íslandi
- Det Arnamagnæanske Institut
- Raqoon ehf

Project overview

- The project seeks to investigate the applicability of **semantic web technologies** for the manipulation of digital library data

- Manuscript data from the Arnamagnæan institutes is used as proof of concept

- The project is funded by the **Nordunet2 research programme**

NORDUNET2

RAQOON

Project benefits

- Allows end-users to browse the Arnamagnæan digital collection through the web in more intelligent and useful ways than before

- Enables computers to process knowledge that up until now has only been available on paper (or inside the heads of scholars)

RAQOON

The semantic web – what is it?

- The information on the world-wide-web is mainly intended for consumption by humans

- The "semantic web" is an effort to make the web more friendly towards machines, with the ultimate aim of making the data more accesible for humans

The semantic web – architecture

⊕ Tim Berners-Lee (the inventor of the web) has proposed this as an architecture for the semantic web:
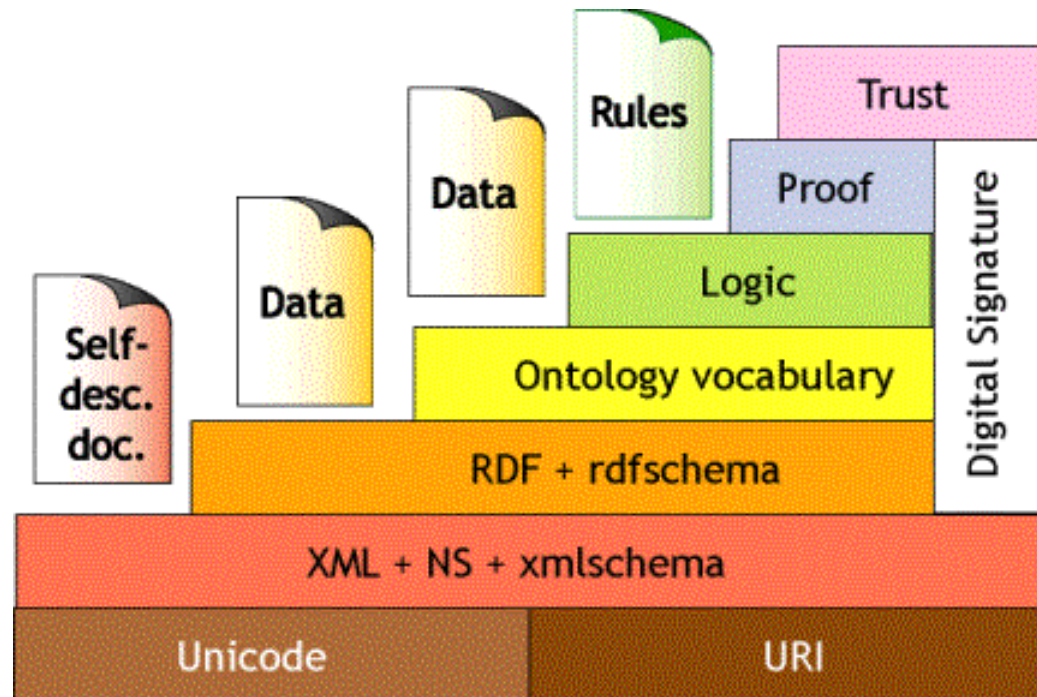
The Arnamagnæan manuscript collection

- Around 3000 items
- Mostly Icelandic manuscripts but also Norwegian, Danish and Swedish
- Stored in Denmark and Iceland

RAQOON

The Arnam agnæan digital collection

- Current collection…
  1. Digital images of manuscript pages
  2. Electronic catalogue (XML)
  3. Text archive (XML)
- This project adds…
  - XML metadata describing the content of digital images (MIML)
  - XML metadata describing the semantics of the collection as a whole (MSML)

M M L

- **M**anuscript **I**mages **M**arkup **L**anguage
- XML dialect to describe the contents of manuscript images
- Enables the creation of advanced user interfaces
- Serves as a metadata layer above the digital images, providing access for the semantics layer

M SM L

- **M**anuscript **S**emantics **M**arkup **L**anguage
- Describes the semantics of the manuscript collection **as a whole**
- MSML is split in two parts:
  - Core ontology (of which there is only one)
  - Ontology instances (of which there can be many)
- Uses DAML+OIL (Darpa Agent Markup Language+Object Inference Language) to model the semantics
  - DAML+OIL is RDF, which in turn is XML

Architecture – Metadata view

RAQOON

| User Interface | Web pages **(HTML)** | Interactive graphics **(SVG/XML)** |

XSLT Transformation Engine

Images (JPEG) | **MIML** (XML) | **MASTER** (XML) | **Other metadata** (XML)

MSML (DAML+OIL/RDF/XML)

Stylesheet 1 (XSLT/XML)

Stylesheet 2 (XSLT/XML)

⋮

Stylesheet n (XSLT/XML)

RAQOON
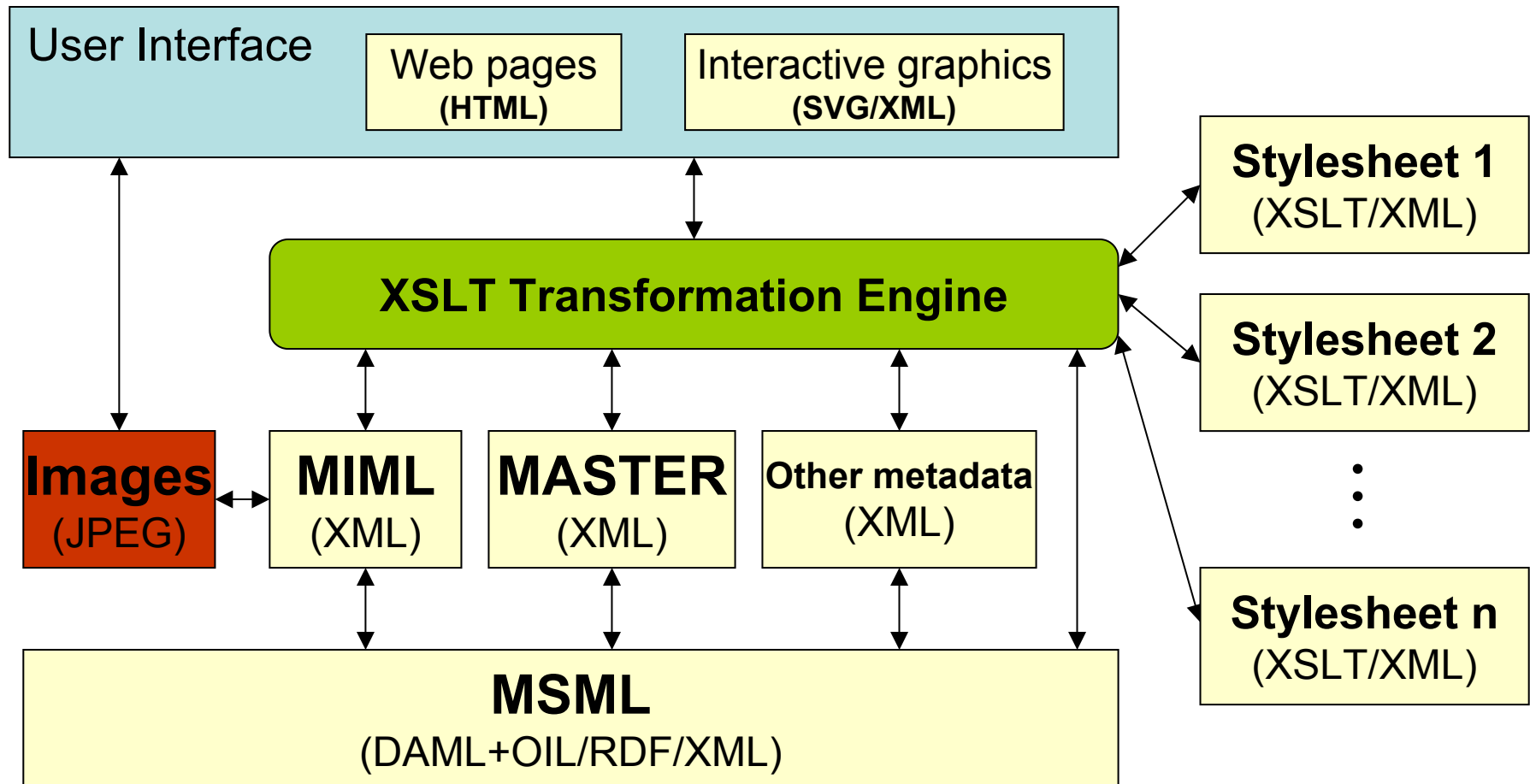
Architecture – System view

Web Browser with SVG plugin

↕ **HTTP**

Tomcat servlet engine

Struts MVC framework

Xerces XML parser

Xalan XSLT transformer

↕ **Java API**

X-Hive/DB native XML database

| MIML | MASTER | MSML | Stylesheets |
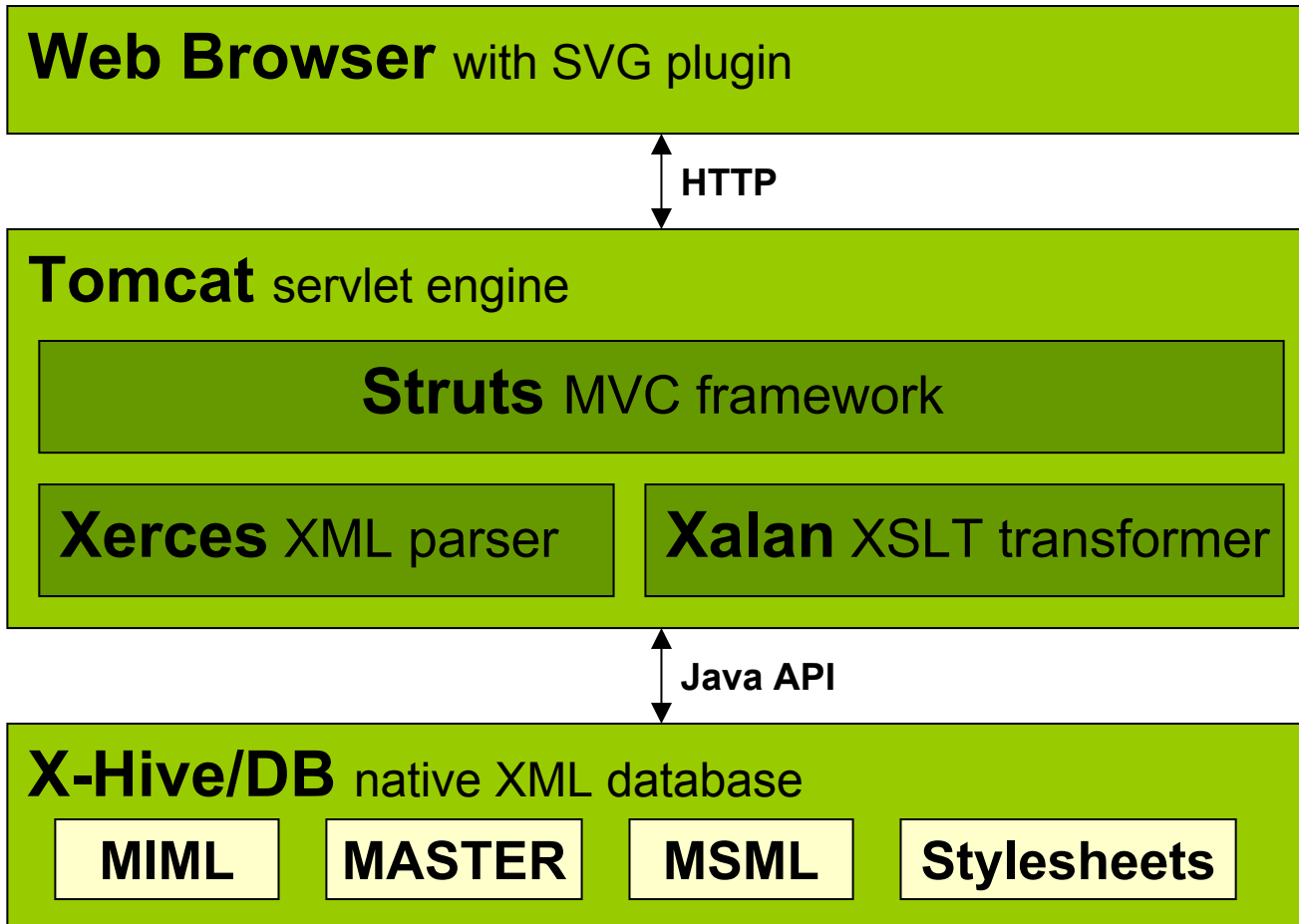
Software Demonstration

This demonstration will show:

- How MIML data (stored in a native XML database and transformed into SVG for visualisation) augments images of manuscript pages

- How MSML data (also stored in a native XML database and transformed into SVG for visualisation) is useful for intelligently linking manuscript related information with other information sources (in this case an ancient map)

Problems encountered during the project

- Bridging the gap between the technological people (Raqoon) and the manuscript people (the two Arnamagnæan institutes) was hard
- We may have tried to climb to high in the semantic web architectural stack
  - DAML+OIL might be overkill, RDF could have sufficed for the project
- Non-technical people find it hard to work with MSML (MIML is ok)

Dissemination of results

- Project ends on the 1st of June with a grand opening
- For the Arnamagnæan institutes this project means...
    - ...a much bigger audience for Icelandic manuscripts
    - ...a rich toolset for creating Internet exhibitions
- For Raqoon this project means...
    - ...technological leadership in advanced digital libraries
    - ...opportunities in other areas, such as knowledge management