Szabolcs Bíró librarian-
informatics teacher

biro.szabolcs@neumann-haz.hu
www.neumann-haz.hu

H-1014 Budapest
Színház street 1-3.
Telephone: 212-9220/220
Mobile: 30 641-3213

Cocoon-based web content generation with XML and XSL

Thanks to the increasing spreading of the XML and XSL languages on the web it has now become possible to use simple server applications which are capable of high-level processing. As these languages are available free of charge and they are platform-independent, these applications do not combine web servers and various scripts – such as CGI – and do not use trade-mark-protected technology (such as .Net or Enterprise Java Beans) either.

Multinational companies, publishers and content providers wish to produce content for the web in XML format, in order to be able to create (X)HTML, PDF, WML etc. formats therefrom as needed. The Cocoon product of the Apache project is an attempt to produce such a server.

Cocoon is a servlet – a framework program –, which presents the content marked up in XML on the web, using XML and XSL. Such novelties remain obviously only theoretically important as long as the technologies in question cannot be put to practice. Professionals are aware that XML analysts have been with us for a long time but their wider use has started only recently. XSL(T) style-sheet processors and XSL converters – which are capable of converting documents formatted in XSL into the content formats described above – have been developed for four years now. The XML Apache project received several such software products from developers for further open-source development but these have not had much revolutionising effect until now.

The work done till now was crowned by the Cocoon project as it had integrated the software elements prepared into a unified framework system, making it possible that full web pages be based on XML-XSL transformations. Such web pages possess incredible capabilities and rely on the W3C recommendations. The style and content are fully separated, they may consequently be transformed at any time to present information and documents in an entirely different manner. At the same time those creating content have a very simple task now because they could get rid of the "primitive and limited" solutions of web-based content description languages.

The solution mentioned may be apt for library application as well. As it is one of the basic tasks of libraries to preserve printed materials and modern tools are available to implement this task if we wish to preserve part of the printed collection within a digitisation project, there evolves a clear need for storing content in an "enduring" format. For this purpose XML is well suited; the Neumann House has been using this format from the very beginning of its activities, and it is going to use it from now for web development as well.

Some readers of this summary may murmur to themselves how complex the earlier simple web-based programming model has become. We should admit that it is becoming increasingly difficult to build web-based systems because they are accessed with a variety of tools from a variety of environments. Without adequate tools it is difficult to manage the complex solutions even today, and the situation is becoming more and more complex. It is time to start learning...