A Practitioner’s Approach to Evolving and Remodeling Large-Scale WWW Sites

Paris A. Zafiris, Nektarios P. Georgantis, George E. Kalamaras, Sotiris P. Christodoulou and Theodore S. Papatheodorou
High Performance Computing Lab, Computer and Informatics Engineering Dept., University of Patras, Rion, Patras, 26500, Greece
{paz, npg, gek, spc, tsp}@hpclab.ceid.upatras.gr

Abstract
This paper addresses the issues of evolving large-scale WWW sites, in order to migrate to more flexible forms of developing and managing content and applications. We first present an integrated process model for the life-cycle of a WWW site. Next, we focus on specific steps of this model to (i) examine a layered, next-generation architecture for large WWW sites, providing for content re-use and easier implementation of services and applications and (ii) to discuss the activities of remodeling our existing content infrastructure and applications, according to the specifications of this target architecture. In our case study, we apply the new architecture, through the remodeling activities, in the evolution of the official site of the Hellenic Ministry of Culture.

Keywords: Architectures, design, development, evolution, hypermedia, methodology, models, process, remodeling, systems, web

1. Introduction
It is widely recognized that most Web site installations, even large-scale ones, are still using obsolete technologies for content and applications management, as well as insufficient processes for carrying out their operation and maintenance. A simple example is the widespread coupling of content and applications through HTML, that is posing numerous problems for maintainability and extensibility. Therefore, a speedy evolution and the establishment of more systematic approaches for managing information, applications and operations is considered especially important.

In this work we are addressing the issues of major, transparent evolution of large scale WWW infosystems, an activity which involves significant remodeling of their core components. By the term “evolution” we may refer to
    • the regular maintenance of the site, which can be perfective, adaptive, corrective, speculative, as in other software systems [15]
    • a more extended restructuring and enhancement. This not only prevents the degradation of the site, but allows existing installations to maintain acceptable levels of manageability, to safely avoid the dangers of insufficient or immature technologies and to prepare for the future of online computing. Obviously, it is an action that should be performed using a careful and thorough approach, at least once in the life of any Web Site and even more often at large installations with increased complexity and potential points of failure.

In this paper we emphasize on the second case. Since this kind of evolution occurs at less frequent intervals, it is akin to migrating to a next generation of the site and requires serious remodeling of the overall architecture, information infrastructure (infostructure), applications/services, system and user interfaces.

Remodeling, as a core step for a major evolution process, includes
    • the definition of a next-generation site architecture, especially focusing on the management of content and applications
    • the migration procedures from the existing state towards the new environment

In our approach, we first present an integrated process model for the life-cycle of large-scale Web sites, consisting of design, implementation, maintenance and evolution steps. We then focus on two specific steps of the process model:
    (i) the introduction of a layered architecture for viable content management and easier deployment of application/services on-top-of the integrated content. Key notion is the proper (re)modeling of information, in different repositories, and its migration to a more streamlined “form” (including structure and semantics), which allows for easier maintenance and expansion and is the basis for easier applications and services development.
    (ii) the actual remodeling of content repositories and applications into more flexible forms, according to the specifications of the target architecture. This is done in a systematic fashion, so that existing large sites can continue their operation and perform their evolution without sacrificing functionality.