Integrating Document and Workflow Management Systems

Lerina Aversano, Gerardo Canfora, Andrea De Lucia, Pierpaolo Gallucci

Faculty of Engineering - University of Sannio
Palazzo Bosco Lucarelli, Piazza Roma - 82100 Benevento, Italy
(aversano / canfora / delucia / gallucci)@unisannio.it

Abstract

A critical point for developing successful information systems for distributed organisations is the need for integrating heterogeneous technologies and tools. This paper reports on an experience of integrating two key enabling technologies, namely workflow and document management.

1. Introduction

Information and communication technology (ICT) is a primary enabler of virtual organisations, as peoples and institutions in a network make substantially more use of computer-mediated channels than physical presence to interact and cooperate in order to achieve their objectives. In particular, two primary enabling technologies are workflow management and document management. One of the main advantage of Workflow Management Systems (WfMSs) is moving the focus from the automation of single process activities, through traditional information systems, to the overall management and improvement of the business processes, through the integration of different software technologies. In addition, the last generation of WfMSs leverage the Web as enabling infrastructure, thus allowing a higher level of coordination and control among the geographically distributed teams and individuals that take part in a business process. Document management systems complement workflow management as they focus on the management of the documents developed and exchanged by the subjects taking part in a business process.

In this paper we reports on an experience in achieving integration of workflow and document management. The work presented in this paper is being developed and experimentally validated within the technology-transfer research project LINK, aiming at transferring innovative technologies to the Public Administration (PA) and Small and Medium software Enterprises [1].

2. GIANO: a document management system

The first subsystem developed within the LINK project is a document management system, called GIANO, that combines database and information retrieval technologies with hypertext systems and mark-up languages.

The document space consists of XML files, whereas a relational database is used to store metadata on the documents. Different Document Type Definitions (DTDs) are used to define the structure of different document types. For each DTD a table in the GIANO database exists whose attributes correspond to metadata and particular tags of the DTD that identify items for which searching the database is more efficient than searching the document space; examples of such attributes include title, author, and creation date. An additional table in the GIANO database associates each document to the set of its keywords (where available).

The document retrieval in GIANO is achieved through two refinement levels, as depicted in Figure 1. At the first level a class of documents (i.e. a DTD) is selected and a database query on the discriminative tags of the DTD is made to select a subset of documents in the document space. This defines a new document space where the second refinement is applied. Indeed, the information retrieval subsystem searches the XML documents in this space according to a context-sensitive boolean query expressed on the DTD tags. The final set of relevant documents is retrieved together with a measure of their relevance with respect to the query. The retrieved documents are presented to the user in hypertext format. The user can select a document to browse and a format visualisation style (i.e. an XSL - eXtensible Style Language - style sheet), thus achieving a customised document view.

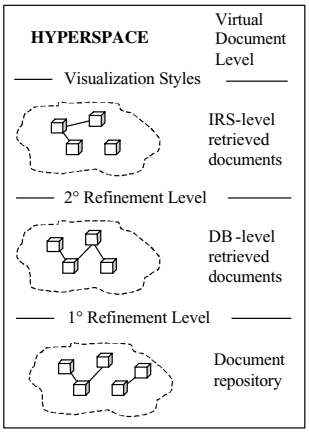

Figure 1: GIANO document levels

Figure 2 shows the architecture of GIANO. The system is composed of the GIANO client, the GIANO server, the

Proceedings of the IEEE Symposia on Human-Centric Computing Languages and Environments (HCC'01)
0-7695-0474-4/01 $17.00 © 2001 IEEE

repository, and external components, such as the Web browser and Web server, the relational Data Base Management System (DBMS), and the Information Retrieval System (IRS). The Client component provides the facilities for user accounting, DTD selection, and query composition and refinement. The GIANO client communicates with the GIANO Document Manager to load the list of available DTDs. The GIANO client also communicates with the Web Server to load the selected DTD. The metadata and the DTD tags are used to compose the two levels of queries. The metadata-level query is sent to the Document Manager to recover from the DB the list of documents that satisfy the query. The context-sensitive query is sent to the Information Retrieval Manager to access the Information Retrieval System and retrieve the final set of documents. The Hypertext Generator is the component that produces the HTML index of the retrieved documents including the link to the corresponding XML files, together with information about the XSL file-sheet used to visualise and browse them.

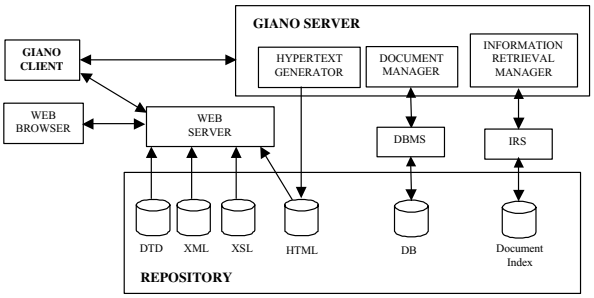

Figure 2: GIANO architecture

3. Integrating GIANO into a WfMS

In our project, an evaluation of market workflow technologies was performed to select the more appropriate platform for the automation of the administrative processes of a customer PA department [1]. In particular, we used a customised and simplified version of the DESMET method [2] for the evaluation of software engineering technologies. We selected Ultimus Workflow (http://www.ultimus1.com), a Web-based workflow management system running on Microsoft Windows NT.

The integration of GIANO into the workflow prototype has been achieved at two different levels: document level and user interface level. Integration at the document level first requires to combine document and workflow analysis and modelling. The information collected during the analysis of the business processes needs to focus on the document structure and life cycle, in addition to the workflow of the process activities. We used UML to build static and dynamic models of each document type involved in the process, in particular UML class diagrams have been used to model the structure of the documents and the relationships among them, while the document life cycles have been formalised through state diagrams. UML class diagrams have been used to semi-automatically generate the DTDs, the corresponding metadata tables in the DB, and to identify hypertext links between documents. UML state diagrams are needed in addition to the DTD and the metadata table schema to produce the form-based user interfaces for the workflow activities corresponding to document creation/evolution (see Figure 3). Scripting functions have been generated to produce/load the XML documents from/in the fields contained in the HTML forms.

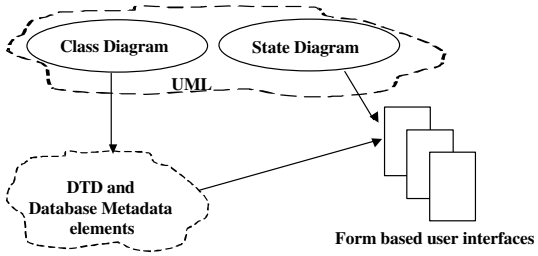

Figure 3: DTD, metadata, and form production

The second issue concerns a web-based reengineering of the Giano user interface and its integration into the Ultimus workflow client. To map the graphical objects of the old user interface onto objects of the new interface we used the guidelines of the MORPH methodology [3].

In the old version of GIANO, the communication between the client and the server was based on TCP/IP sockets. Therefore, to avoid changing the GIANO server, we also needed to reengineer the client-side java communication layer. In particular, we implemented a DLL that is loaded by the workflow engine and used to wrap the GIANO server.

4. Concluding remarks

The work reported in this paper has addressed the problem of creating an information infrastructure and services for distributed and virtual organisations, and particularly the integration of two key enabling technologies, namely workflow and document management. We have presented a case study where integration has been achieved by combining several approaches, including software engineering, hypertexts, and visual languages.

References

[1] L. Aversano, G. Canfora, A. De Lucia, and P. Gallucci, "Business Process Reengineering and Workflow Automation: A Technology Transfer Experience", *The Journal of Systems and Software*, to appear.

[2] B. Kitchenham, "DESMET: a Method for Evaluating Software Engineering Method and Tools", *Technical Report 96-09*, Dep of Computer Science, University of Keele: Staffordshire, UK, 1996.

[3] M. Moore and S. Rugaber, "Using Knowledge Representation to Understand Interactive Systems", *Proc. of 5th Intern. Workshop on Program Comprehension*, Dearborn, MI, 1997, pp. 60-67.