

Flexible Collaboration over XML Documents

Claudia-Lavinia Ignat and Moira C. Norrie

Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
{ignat, norrie}@inf.ethz.ch

Abstract. XML documents are increasingly being used to mark up various kinds of data from web content to scientific data. Often these documents need to be collaboratively created and edited by a group of users. In this paper we present a flexible solution for supporting collaboration over XML documents. Merging of user work is based on the operations performed. A key to achieving flexibility for the definition and resolution of conflicts was to keep the operations distributed throughout the tree model of the document associated with the nodes to which they refer.

1 Introduction

XML is a popular format for marking up various kinds of data, such as application data, metadata, specifications, configurations, templates, web documents and even code. While some XML documents are generated automatically from systems, for example database exports, there are many cases where XML documents are created and edited by users either in raw text format or through special tool support. Our goal was to investigate concurrent editing over XML documents. A key issue in collaboration is merging versions of XML documents produced by different users.

Some state-based approaches for merging XML documents have been proposed in [12,2,9]. As opposed to state-based merging, where only the information about the states of the documents is used, the operation-based approaches keep information about the evolution of one document state into another in a buffer containing a history of the operations performed between the two states of the document. Merging is done by executing the operations performed on one copy of the document on the other copy of the document. Conflicts might exist between concurrent operations and an important issue is how to provide users support for the definition and resolution of conflicts. Merging based on operations offers good support for conflict resolution by having the possibility of tracking user operations.

Operation based approaches for merging XML documents have been proposed in [10] and [4]. These approaches did not deal with issues related to the definition and resolution of conflicts. They maintain a single history buffer where the executed operations are kept and it is difficult to select the operations that refer to a particular node in the document. This fact has limitations for the definition and resolution of conflicts referring to a unit of the document. These approaches adopt an automatic resolution of conflicts by combining the effects of

concurrent operations, but do not allow users in the case of conflict to manually select between versions of elements.

In this paper, we propose a flexible merging approach for XML documents where the history buffer containing the performed operations is distributed throughout the tree rather than being linear. We previously applied the distribution of history throughout a tree document model for the case of hierarchical text documents in the real-time [8] and asynchronous [6,7] modes of collaboration. In this paper we show how the association of operations with the structure of the document is used for the asynchronous collaboration over XML documents with a shared repository. The resolution of conflicts is simplified compared to the approaches that use a single history buffer, as conflicting operations that refer to the same subtree of the document are easily detected by the analysis of the histories associated with the nodes belonging to the subtree. Moreover, conflicts can be dynamically defined at different levels of granularity corresponding to the different levels of the document. The policies for merging can be automatically or manually defined by users.

The paper is structured as follows. In section 2 we describe the requirements for merging and editing XML documents. Section 3 presents the model of the document and the set of operations that we adopted. We then go on to describe in section 4 our approach for merging. In section 5 we present the flexibility of our approach for the definition and resolution of conflicts. We present concluding remarks in section 6.

2 Collaboration over XML Documents

In this section we are going to motivate the need for change management over XML documents and give a brief description of the issues to be considered for the editing of XML documents.

Consider the case of a research team in the field of computational chemistry that wants to publish the results of their experiments in XML documents. The XML format offers a number of advantages for computational chemistry, such as clear markup of input data and results, standardised data formats, and easier exchange and archival stability of data.

The results of the simulations may be modified several times by members of the team and various versions may be required at different times. Therefore, a version management system should offer support for the archiving of all versions and for the retrieval of past versions. Members may need to query not only the current value of data, but also past values of some elements of the simulation and process change information.

Concurrent editing of the documents containing the data results should be supported as simulations and the gathering of results can be performed in parallel by the members of the group.

Consider the case that two researchers concurrently edit the following part of an XML document: