

Incremental Updates in Structured Documents

Greger Lindén

Ph. Lic. Thesis

Department of Computer Science, University of Helsinki

Report C-1993-19, Helsinki, February 1993

142 pages

Abstract

We present incremental parsing, incremental translation, and lazy translation for structured documents. The logical structure of a document is described by a context-free grammar. Views of the document are defined by annotating the grammar. An annotated grammar defines a translation from the logical structure to a document view. The translation can be inverted giving a translation from the view to the logical structure. A document instance is a parse tree (a logical tree) over the context-free grammar. A document view is a parse tree (a view tree) over an annotated grammar. The textual view of a document is the catenation of the string leaves of the view tree.

We study modifications in the different document representations and how to update them efficiently. First, we give an incremental parser for a textual view. The parser efficiently updates a view tree after a modification in a textual view. The parser is an extension of an algorithm given by Murching et al. We also present an incremental translator, based on syntax-directed translation. The translator transforms a modified logical tree into a view tree or vice versa. The transformation is done incrementally, minimizing update work. Finally, we present a lazy translator that transforms a logical tree into a view tree by postponing unnecessary work. Lazy translation is combined with incremental updates: a lazy view is incrementally updated. We show that the complexity of the algorithms vary substantially.

CR Categories and Subject Descriptors:

- D.3.4 [Programming languages]: Processors – parsing, translator writing systems and compiler generators
- I.7.2 [Text processing]: Document preparation – format and notation, languages and systems
- H.2.3 [Database management]: Languages – data description languages (DDL)

General Terms: Algorithms, Languages.

Additional Key Words and Phrases: incremental parsing, incremental syntax-directed translation, lazy syntax-directed translation, structured text