Collaborative Editing

M. Tech. Seminar Report

Submitted in partial fulfillment of the requirements
for the degree of

Master of Technology

by

V Sai Saketh Ram
Roll No: 13305R010

under the guidance of

Prof. Ganesh Ramakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology, Bombay
Mumbai

December 1, 2014

Contents

Abstract

Real-time group editors allow a group of users to modify the same document at same time from geographically different sites connected by communication networks. Consistency maintenance is one of the most significant challenges in design and implementation of these type of systems. In this report we try to study the various techniques adapted for consistency maintenance during collaborative editing. Most importantly a technique called operatinal transformation and algorithms using this technique are studied.

1 Introduction

1.1 Background

Researchers have many software options to write papers,But fewer to collaborate with others. Two main Reasons for less use of versioning systems are Usability problems and Need for users to resolve conflicts. We will compare google docs with other solutions. And also extensions to google docs to overcome its drawbacks.

1.2 Existing solutions

Current existing solutions for collaborative editing are-

- E mail
- CVS and SVN

1.2.1 E mail

- purely sequential
- conceptually simple
- others cannot contribute
- sent to a single collator

1.2.2 CVS and SVN

- single server maintains repository
- check out,check in concepts
- user must have client software installed
- resolving conflicts
- only text based files

1.3 Reasons for still using older Techniques

- Usability,update and commit concepts
- user accounts have to be managed
- need to resolve conflicts

1.4 Google Docs

- Author edits a doc on google repository.
- uses simple browser editor developed by AJAX
- There is also viewer category
- Changes to docs are automatically transmitted to server(every 30 seconds)
- Because of high frequency of updates,conflicts are unlikely
- Doc can be saved in various formats like PDF,HTML

1.5 Shortcomings of Google Docs

[3]

- Editor problems
- Academic requirements
- Offline support
- Docs reside on google server.So no gaurantee of security
- Text based only
- Hard to control output layout
- Vendor independence(cannot use favorite editors)

1.6 Extending Google Docs

[3]

- Automatic conversion to LATEX
  - Importing google docs
  - Initial cleansing
  - Canonicalization
  - Latex transformation
  - Security issues
- Synchronizing offline documents

2 Operational Transformation

2.1 Introduction

Consistency maintenance is significant challenge in any collaborative ediotor. Most innovative technique for maintaing it is Operational Transformation. We concentarte on integrative review of all OPT techniques and their issues,algorithms and achievements[1]

2.2 Preliminaries

In this section some basic concepts,definitions and terminologies are introduced. Operational transformation technique allows two type of operations called local operations and remote opeartions. Local operations are nothing but operations that are genearted by local site. Where as remote opeartions are those operations taht are generated at some other site other than local site. Local opeartions are executed immediately with out any changes but remote opeartions have to be transformed against existing operations before execution.
Now we introduce some defnitions.

Def 1:causal ordering relation [1]

$$\rightarrow$$

Given two operations Oa and Ob,generated at sites i and j,then

$$Oa \rightarrow Ob$$

,if (1)i=j and generation of Oa happened before the generation of Ob or (2)i is not equal to j and execution of Oa at site j happened before the generation of Ob or (3)there exists an opeartion Ox,such that

$$Oa \rightarrow Ox$$

and

$$Ox \rightarrow Ob$$

.

Def 2:Dependent and Independent operations
Given two operations Oa and Ob,(1)Ob is dependent on Oa if

$$Oa \rightarrow Ob$$

.(2)Oa and Ob are independent(or concurrent), expressed as

$$Oa \parallel Ob$$

,if neither

$$Oa \rightarrow Ob$$

nor

$$Ob \rightarrow Oa$$

2.3 Principles of consistency

- convergence(all copies are identical)
- Casuality preservation
- Intention preservation(effect of execution of operation O is same as intention of O)

2.4 List of algorithms

[1]

- GROVE
- REDUCE
- JUPITER
- adOPTed

2.4.1 GROVE approach

- Replicated architecture(for good responsiveness)
- Locally executed and then braodcasted
- solution consists of 2 components
    - state-vector time stamping scheme for precedence property
    - dOPT(distributed) algorithm for convergence
- dOPT algorithm requires the transformation function to satisfy the following property
    - $$O_a \circ O'_b = O_b \circ O'_a$$
- dOPT algorithm
- unsolved dOPT puzzle
- Reason:above TP has to be applied only when two operations are in same state
- If they are not in same state? dOPT did not handle that !!

2.4.2 REDUCE approach

- Fully didtributed and Replicated as in GROVE
- Linear history buffer is maintained instead of log
- Also handles 3rd consistency principle(Intention Preservation)
- To address that it uses IT/ET and pre/post conditions for applying those
- specifications
- GOT(generic operational transformation algorithm)

2.4.3 JUPITER approach

- No broadcasting among clients
- only 2-way between client and server

2.4.4 adOPTed approach

- It requires additional properties from GROVE for transformation function to satisfy
  - $$O_a \circ O'_b = O_b \circ O'_a$$
  - $$T(T(O, O_a), O'_b) = T(T(O, O_b), O'_a)$$
- TP1 ensures unique vertex labelling
- TP2 ensures unique edge labelling

2.4.5 treeOPT algorithm

[2]

one of the keys to customization is structured representation of the document.
In this section we summarize treeOPT algorithm.
It applies operational transformation mechanism recursively over different levels of document.
Advantages of using this algorithm are

- better efficiency
- possibility of working at different granularity levels
- improvement in semantic consistency

This algorithm uses tree representation of the document.
Disadvantages of linear representation of document are

- keeps single history of operations
- Less efficient as whole history needs to be scanned for every operation(Though they are working on different sections of the document)

3 Conclusion

Operational Transformation,most important technique for maintaining consistency in collaborative editing is studied. And also algorithms that use this technique are studied.Google Docs problems has been identified and some extensions are proposed for better usage of google docs.

Acknowledgements

I would like to thank my guide, Prof. Ganesh Ramakrishnan for the consistent directions he has fed into my work.

References

[1] Chengzheng Sun,Clarence (skip) Ellis. *Operational Transformation in Real-Time Group Editor:Issues,Algorithms and Achievemnts.* Proceeding CSCW '98 Proceedings of the 1998 ACM conference on Computer supported cooperative work Pages 59-68

[2] Claudia-Lavinia Ignat,Moria C.Norrie. *Customizable Collaborative Editor Relying on treeOPT Algorithm.* Proceeding ECSCW'03 Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work Pages 315-334

[3] Stijn Dekeyser,Richard Watson.The University of Southern Queensland, Australia. *Extending Google Docs to collaborate on Research Papers* In Proceedings of the ACM conference on CSCW'06, pages 207 to 217.