Factoring Differences for Iterative Change Management

Software Development Laboratory
<SDML>

Michael L. Collard

Huzefa Kagdi

Jonathan I. Maletic

KENT STATE UNIVERSITY

<SDML>

Introduction

- Large commits are problematic due to the broad impact on a system
- Small, incremental changes minimize the impact of a change simplifying comprehension, integration, and testing
- However, during development changes are often committed as a complete package, not as a series of small, incremental changes
- The path of development changes is often indirect and not the clear path that is desired

KENT STATE UNIVERSITY

<SDML>

Solution

- Factor a large change into prime factors based on syntactic and documentary elements
  - Deleted items
  - Changes to functions
  - Changes dealing with a new attribute
- Scalable for large commits/systems

KENT STATE UNIVERSITY

<SDML>

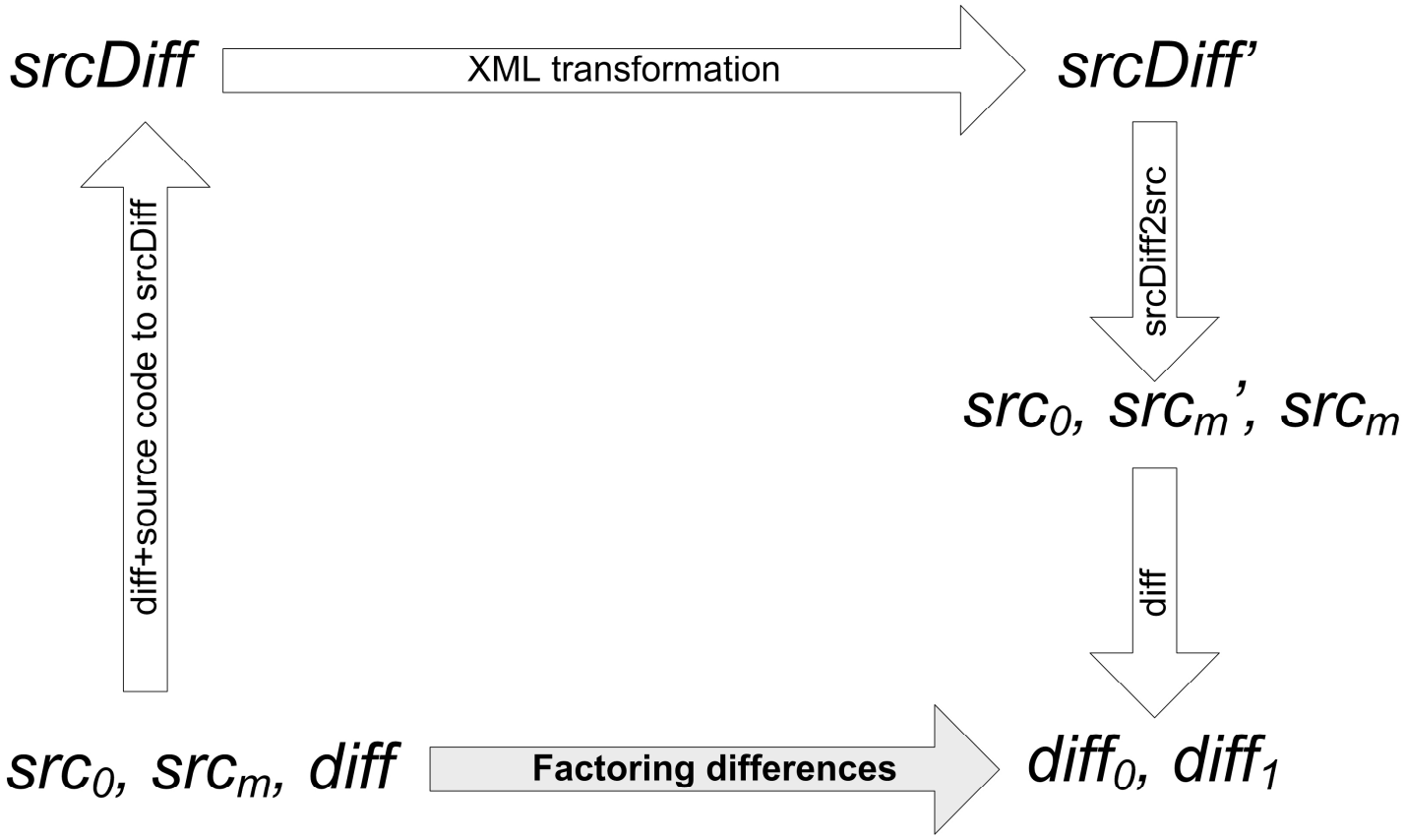
Framework

- srcML [IWPC'01,'03] an XML representation that inserts selective AST information as markup
  - Complete preservation of original text
  - Compound format for multiple files
- srcDiff [ICSM'04] an XML representation of multiple versions of a source-code file and their differences
  - Complete preservation of original and modified versions
  - Compound format for multiple files

KENT STATE UNIVERSITY

<SDML>

Factoring Process



KENT STATE UNIVERSITY

<SDML>

Example Change

Original file

```
/*

*/

int a() {

        return 1;
}
```

Modified file

```
/*
        f.cpp

        John Doe
        jdoe@scam.org
*/

int a() {

        return 2;
}

void b() {
   }
```

KENT STATE UNIVERSITY

<SDML>

Comment Change

src:comment

```
/*
        f.cpp

        John Doe
        jdoe@scam.org
*/

int a() {

        return 1;
}
```

KENT STATE UNIVERSITY

<SDML>

Added Function Change

diff:new[src:function]

```
/*

*/

int a() {

        return 1;
}

void b() {
   }
```

KENT STATE UNIVERSITY

<SDML>

Return Value Change

src:function[src:name='a']//src:return/src:expr

```
/*

*/

int a() {

        return 2;
}
```

KENT STATE UNIVERSITY

<SDML>

Toolset

- *src2srcml*, *srcml2src*
  - code to/from srcML
- *src2srcdiff*
  - code to srcDiff
- *diffver*
  - extracts original and modified versions
- *difffact*
  - produces a factor of the change using XPath to select region

KENT STATE UNIVERSITY

<SDML>

Evaluation

- Searched KDE repository
  - “refactoring” and “API”
  - Found 32 changesets
- Selected revision 473657 from 2005-10-24 with changes to Kate and a corresponding API change
- Manually mapped small changes to corresponding file and line changes
- Used our toolset to extract prime factors from changes to a particular file

KENT STATE UNIVERSITY

<SDML>

KDE Change Set

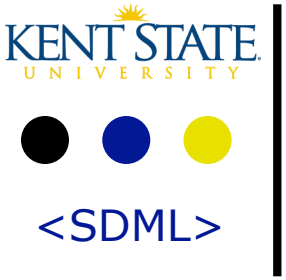
- *Kate* text
  - 26 files, 567 KB, 101 KB gz
- srcDiff document:
  - 2.1 MB, 153 KB gz
  - Conversion speed ~3 seconds
- Factored into updated comments and deleted comments on entire change set with each factoring completed in seconds

KENT STATE UNIVERSITY

<SDML>

Issues

- Complexity of XPath expressions
- Scalability
- Granularity
- Integration with version-control systems
- Prime factors



Software Development Laboratory

<SDML>

www.sdml.info