# Automatically Documenting Program Changes

Ray Buse

**Wes Weimer**

Computer Science
*at the* UNIVERSITY *of* VIRGINIA

13 May 2011
13th CREST Open Workshop
*Automated Software Engineering '10*

ItaDoc

diffs

Commit Messages

De

# So Much Change

# Understanding Change
# Is Important



Developers



Managers



Evaluation

# Understanding change Remains Difficult



**Writing New Code**

**Modifying Existing Code**

**Understanding Code**
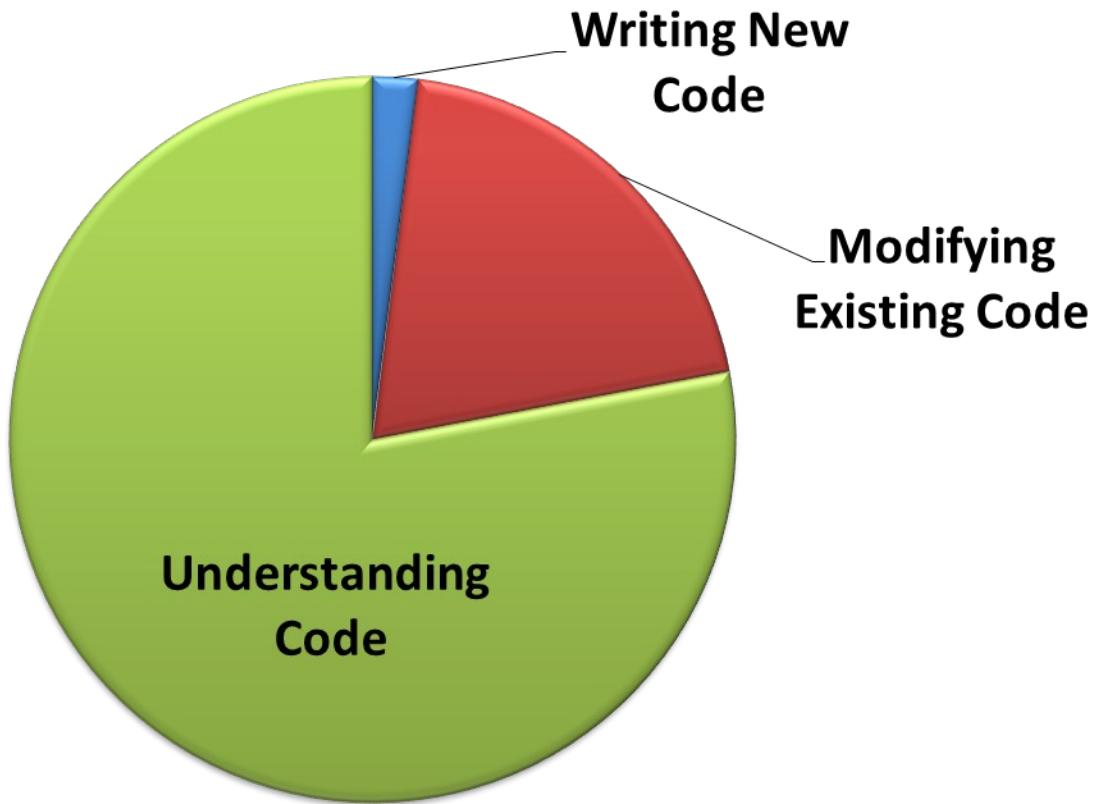
Peter Hallam. *What Do Programmers Really Do Anyway?*
Microsoft Developer Network (MSDN) – C# Compiler. Jan 2006.

# State of the Art: Diff

```
19c19 , 22
<    else return "";
---
>    else return pageParts[0];
>    // else return "";
```

JabRef Revision 3066

# State of the Art:
# Side-by-side

| # | revision 3065 by mortenalver, Tue Nov 25 20:22:56 2008 UTC | revision 3066 by mortenalver, Thu Aug 27 19:32:26 2009 UTC |
|---|---|---|
| | Line 1 | Line 1 |
| 1 | package net.sf.jabref.export.layout.format; | package net.sf.jabref.export.layout.format; |
| 2 | | |
| 3 | import net.sf.jabref.export.layout.LayoutFormatter; | import net.sf.jabref.export.layout.LayoutFormatter; |
| 4 | | |
| 5 | /** | /** |
| 6 | * Formatter that returns the last page from the "pages" field, if set. | * Formatter that returns the last page from the "pages" field, if set. |
| 7 | * | * |
| 8 | * For instance, if the pages field is set to "345-360" or "345--360", | * For instance, if the pages field is set to "345-360" or "345--360", |
| 9 | * this formatter will return "360". | * this formatter will return "360". |
| 10 | */ | */ |
| 11 | public class LastPage implements LayoutFormatter { | public class LastPage implements LayoutFormatter { |
| 12 | | |
| 13 | public String format(String s) { | public String format(String s) { |
| 14 | if (s == null) | if (s == null) |
| 15 | return ""; | return ""; |
| 16 | String[] pageParts = s.split("[\\-]+"); | String[] pageParts = s.split("[\\-]+"); |
| 17 | if (pageParts.length == 2) | if (pageParts.length == 2) |
| 18 | return pageParts[1]; | return pageParts[1]; |
| 19 | else return ""; | // If we didn't get two parts, it may be that only the number of pages is given. |
| 20 | | // Return the first part: |
| 21 | | else return pageParts[0]; |
| 22 | | //else return ""; |
| 23 | | |
| 24 | } | } |
| 25 | } | } |

# State of the Art:
# Commit Messages

- Free-form text which may describe
  - **What** the change was.
  - **Why** the change was made.

**Phex 3542**

Minor change

**jfreechart rev 3405**

(start): Changed from Date to long,
(end): Likewise,
(getStartMillis): New method,
(getEndMillis): Likewise,
(getStart): Returns new date instance,
(getEnd): Likewise.

**Jabref rev 2917**

Fixed NullPointerException when downloading external file and file directory is undefined.

Subject: An appeal for more descriptive commit messages

I know there is a lot going on but **please can we be a bit more descriptive when committing changes**. Recent log messages have included:
"some cleanup"
"more external service work"
"Fixed a bug in wiring" which are a lot less informative than others...

*http://osdir.com/ml/apache.webservices.tuscany.devel/2006-02/msg00227.html*

Toby,
Going forward, **could I ask you to be more descriptive in your commit messages?** Ideally you should state what you've changed and also why (unless it's obvious)... I know you're busy and this takes more time, but it will help anyone who looks through the log ...

*http://lists.macosforge.org/pipermail/macports-dev/2009-June/008881.html*

Sorry to be a pain in the neck about this, but could we **please use more descriptive commit messages?** I do try to read the commit emails, but since the vast majority of comments are "CAY-XYZ", I can't really tell what's going on unless I then look it up.

*http://osdir.com/ml/java.cayenne.devel/2006-10/msg00044.html*

# We Propose: DeltaDoc

Describes the observable EFFECT of a change

- Conditions that trigger the changed code
- How the change impacts functional behavior and program state

**Symbolic Execution** + **Summarization Transformations**

# DeltaDoc

Diff

```
19c19 , 22
<       else return "";
---
>       else return pageParts[0];
>       // else return "";
```

DeltaDoc

When calling LastPage format(String s)
   If s is not null and s.split ("[-]+").length != 2
      return s.split ("[-]+")[0] instead of ""

# DeltaDoc

Temporary removed the trade
routes from the game menu.

DeltaDoc

When calling FreeColMenuBar buildOrdersMenu
  No longer
    call JMenu.add(getMenuItem("assignTradeRouteAction"))

When calling FreeColMenuBar buildViewMenu
  No longer
    call JMenu.add(getMenuItem("tradeRouteAction"))

Freecol rev 2085

# The rest of this talk

- Approach: How DeltaDoc works.



SBSE!
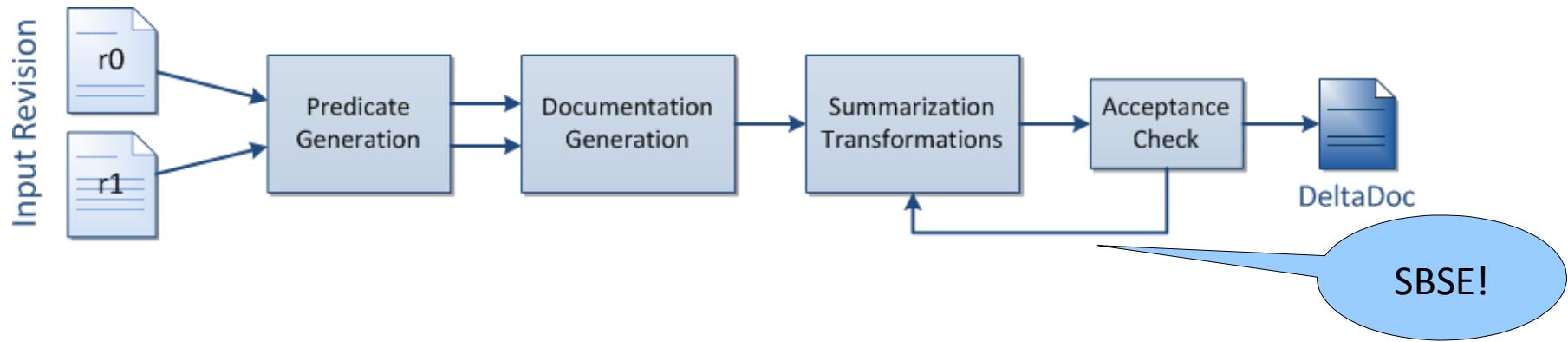
- Evaluation: Comparing DeltaDoc to Commit messages.
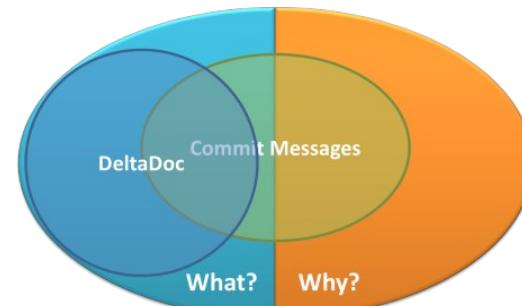
# DeltaDoc Architecture
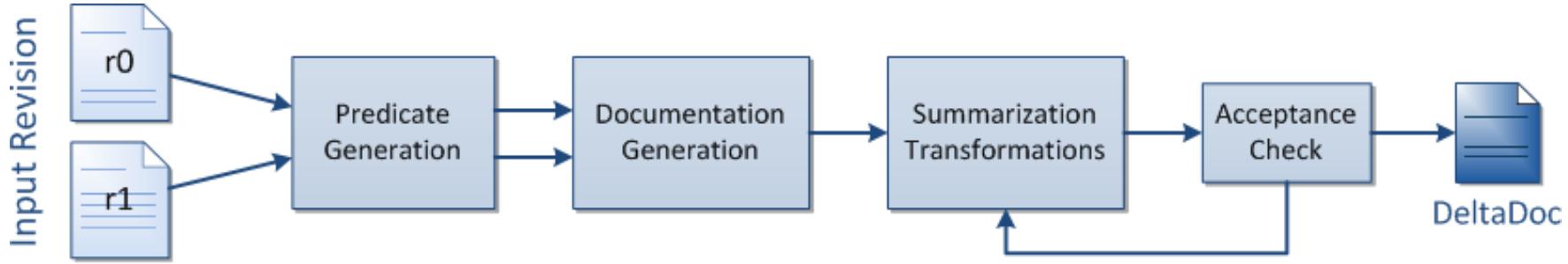
# DeltaDoc Architecture
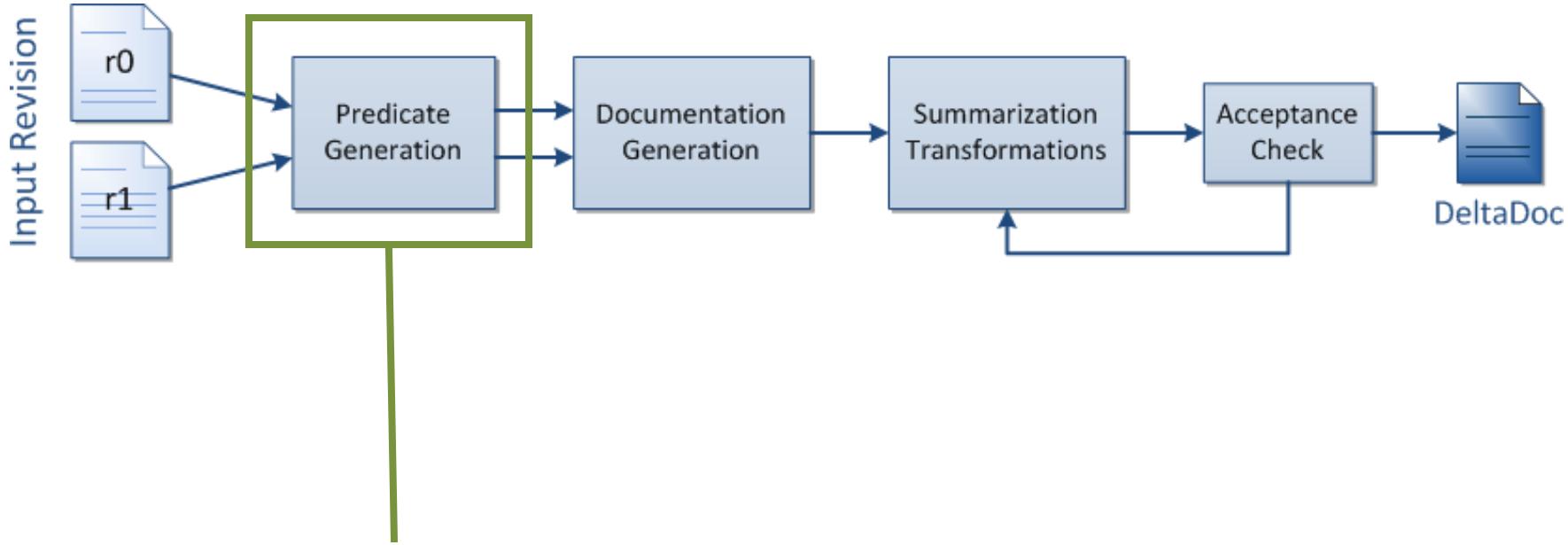


Compute symbolic path predicates for each statement.

# DeltaDoc Architecture



When **X**,
 Do **Y**
Instead of **Z**
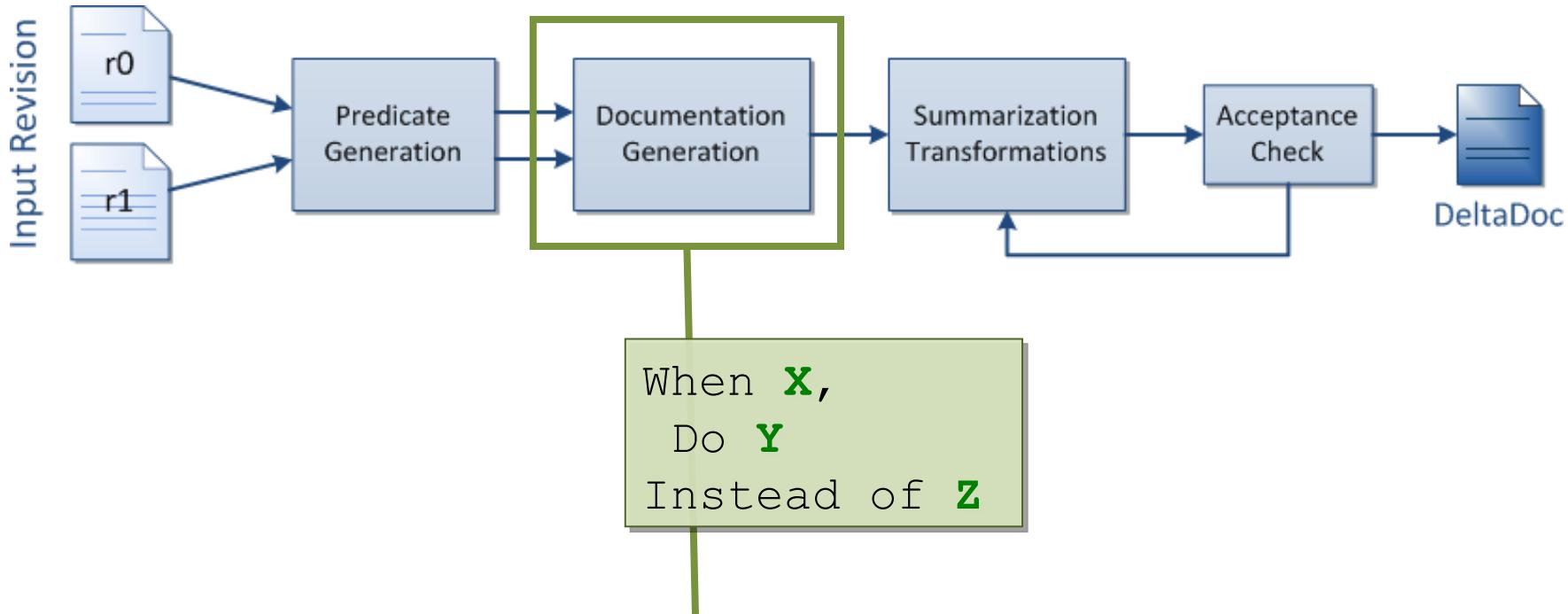
Identify statements that have been added,
removed, or have a different predicate.

# DeltaDoc Architecture
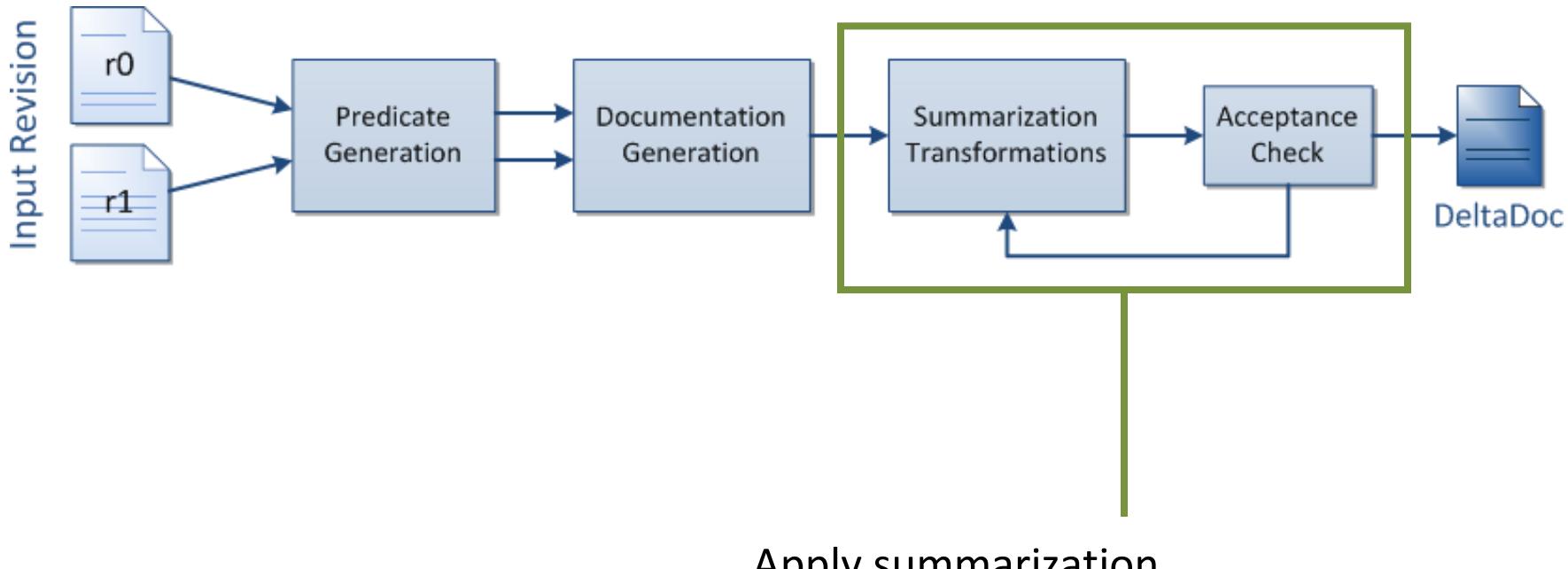


Apply summarization transformations until result is sufficiently concise.

# Example Change

```
String sayHello(String name)
{
    String ret = "";

    if(name != null)
    {
        if(System.Lang == ENG)
            ret = ret + "Hello ";
        else
            ret = ret + "Bonjour ";

        ret = ret + name;
    }

    return ret;
}
```

```
String sayHello(String name)
{
    String ret = "";

    if(name != null)
    {
        if(System.Lang == ENG)
            ret = ret + "Hi ";
        else
            ret = ret + "Bonjour ";

        ret = ret + name;
    }

    return ret;
}
```

# Example
# Predicate Generation

| Stmt | Symbolic | Path Predicate |
|---|---|---|
| **String** ret = ""; | ret = "" | **true** |
| name != **null** | name != **null** | **true** |
| System.Lang == ENG | System.Lang = ENG | name != **null** |
| ret = ret + "Hi "; | ret = "" + "Hi " | name != **null** && System.Lang = ENG |
| ret = ret + name; | ret = "" + "Hi " + name | name != **null** && System.Lang = ENG |
| **return** ret; | **return** "Hi " + name | name != **null** && System.Lang = ENG |

# Example
# Change Enumeration

| Stmt | Path Predicate |
|---|---|
| return "" | name == null |
| return "Hello " + name | name != null && System.Lang = ENG |
| return "Bonjour " + name | name != null && System.Lang != ENG |

| Stmt | Path Predicate |
|---|---|
| return "" | name == null |
| return "Hi " + name | name != null && System.Lang = ENG |
| return "Bonjour " + name | name != null && System.Lang != ENG |

# Example
# Generated Documentation

```
When calling sayHello(String name)
  If name != null AND System.Lang == ENG
    return "Hi " + name Instead of "Hello " + name
```

# Summarization: Lossless

- ## Re-arrange terms *(code hoisting)*

```
If P and Q, Do X
If P and Q and R, Do Y
```

→

```
If P and Q,
   Do X
   If R,
      Do Y
```
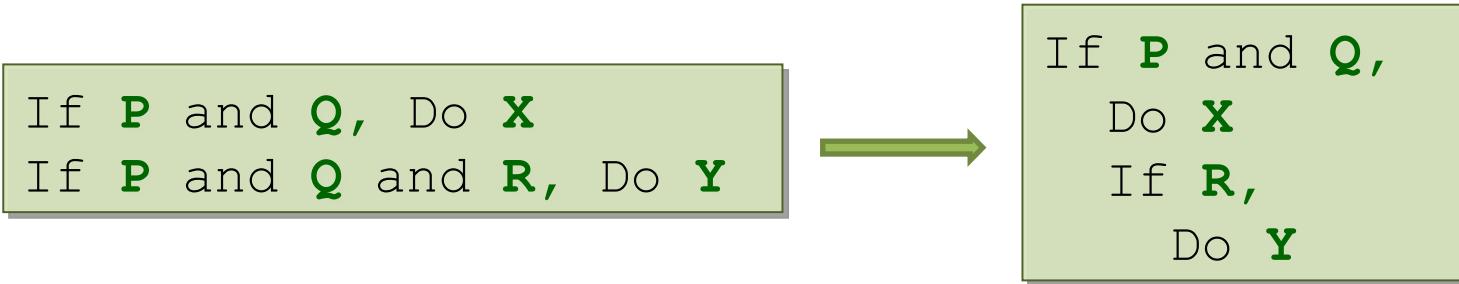
- ## Simplification *(common subexpression)*

```
If iter.hasNext
    call iter.hasNext
    call otherFun
```

→

```
If iter.hasNext
    call otherFun
```
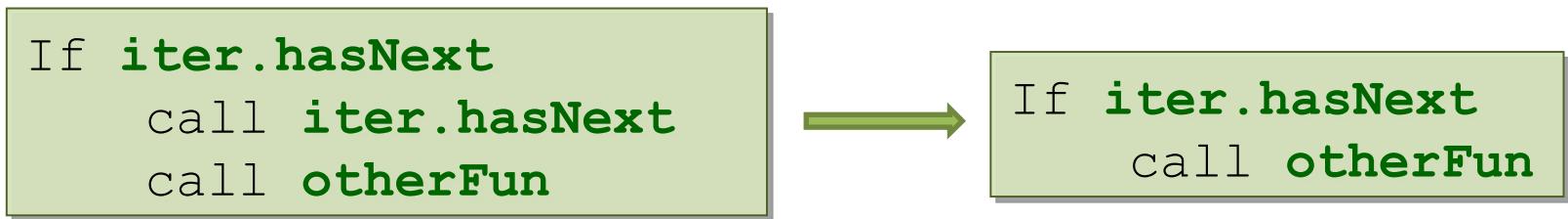
# Summarization: Lossy

Remove irrelevant path predicates.

```
If s != null and a is true
and b is true and c is true
 return s
```

→

```
If s != null
 return s
```

# Summarization: Effects

Random search, greedy prioritization of mutation operators.
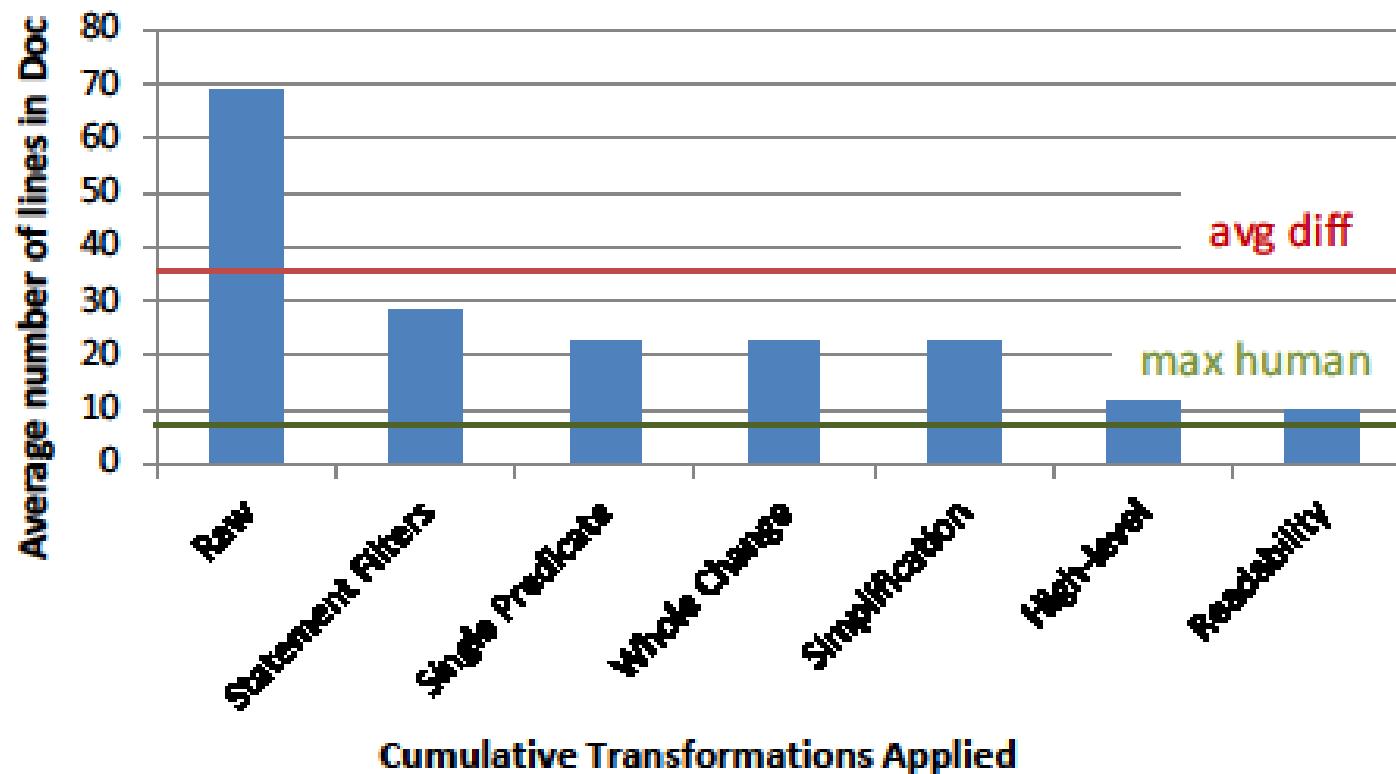
# Evaluation
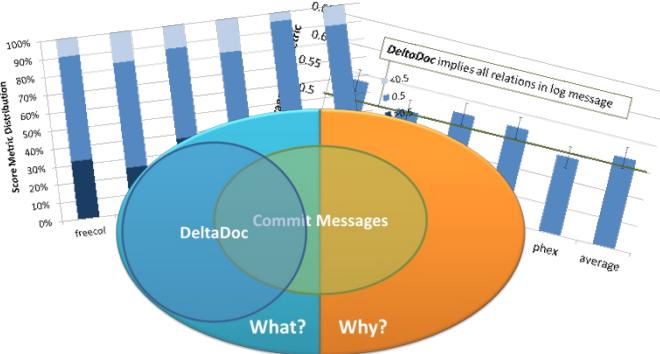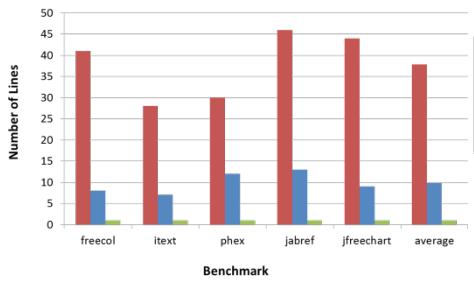


**Size**

**Content**

**Quality**

*(Human study: 16 graduate students)*

# Benchmarks

| Name | Revisions | Domain | kLOC | Developers |
|------|-----------|--------|------|-----------|
| FreeCol | 2000-2200 | Game | 91 | 33 |
| iText | 3800-4000 | PDF utility | 200 | 14 |
| jFreeChart | 1700-1900 | Presentation | 305 | 3 |
| JabRef | 2500-2700 | Reference Management | 107 | 28 |
| Phex | 3500-3700 | File Sharing | 177 | 13 |
| **Total** | **1000** | | **880** | **91** |
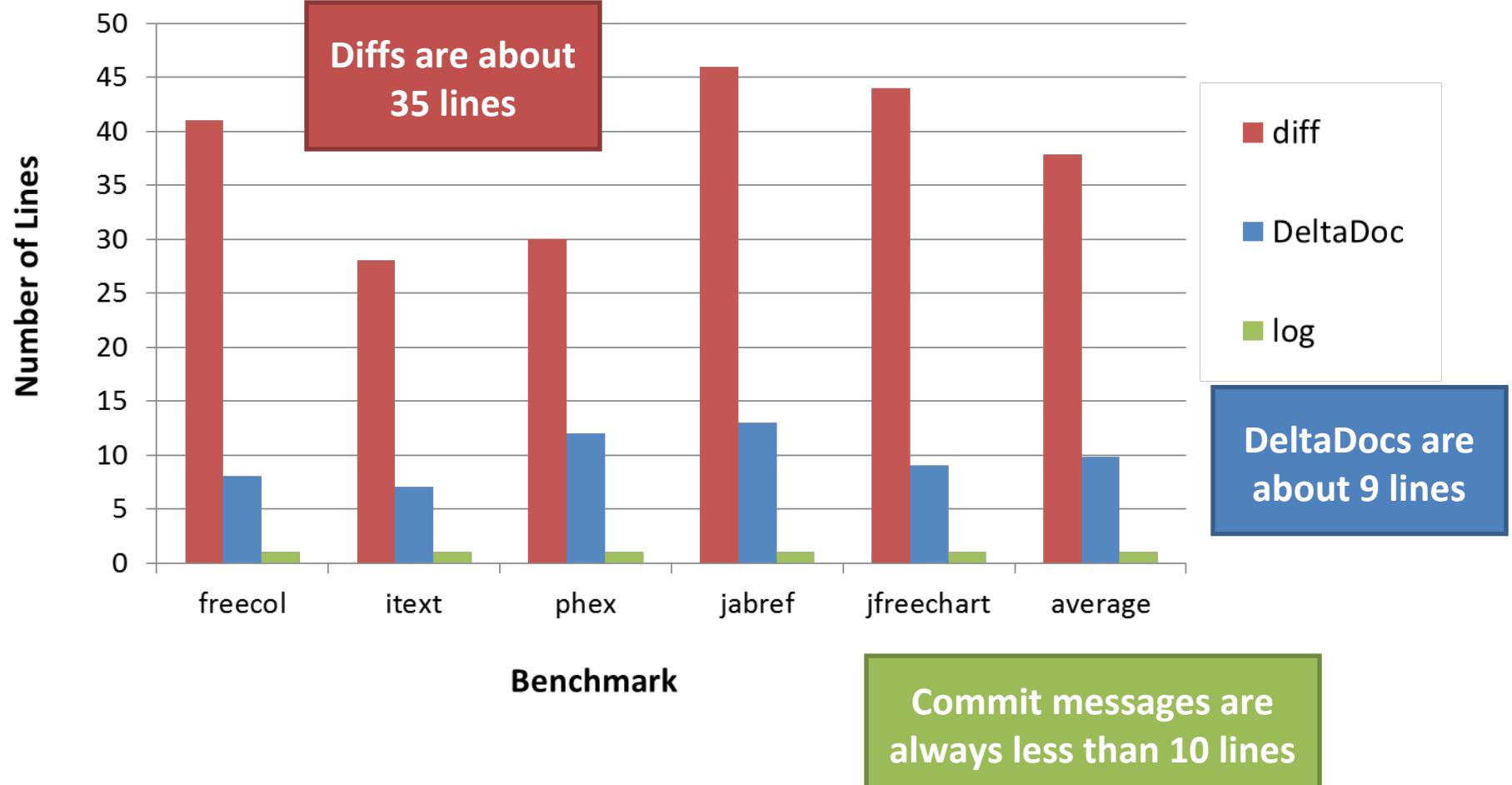
# Size Comparison



**Diffs are about 35 lines**

**DeltaDocs are about 9 lines**

**Commit messages are always less than 10 lines**

# Content Comparison

DeltaDoc

Commit Messages

Relational Form

Relational Form
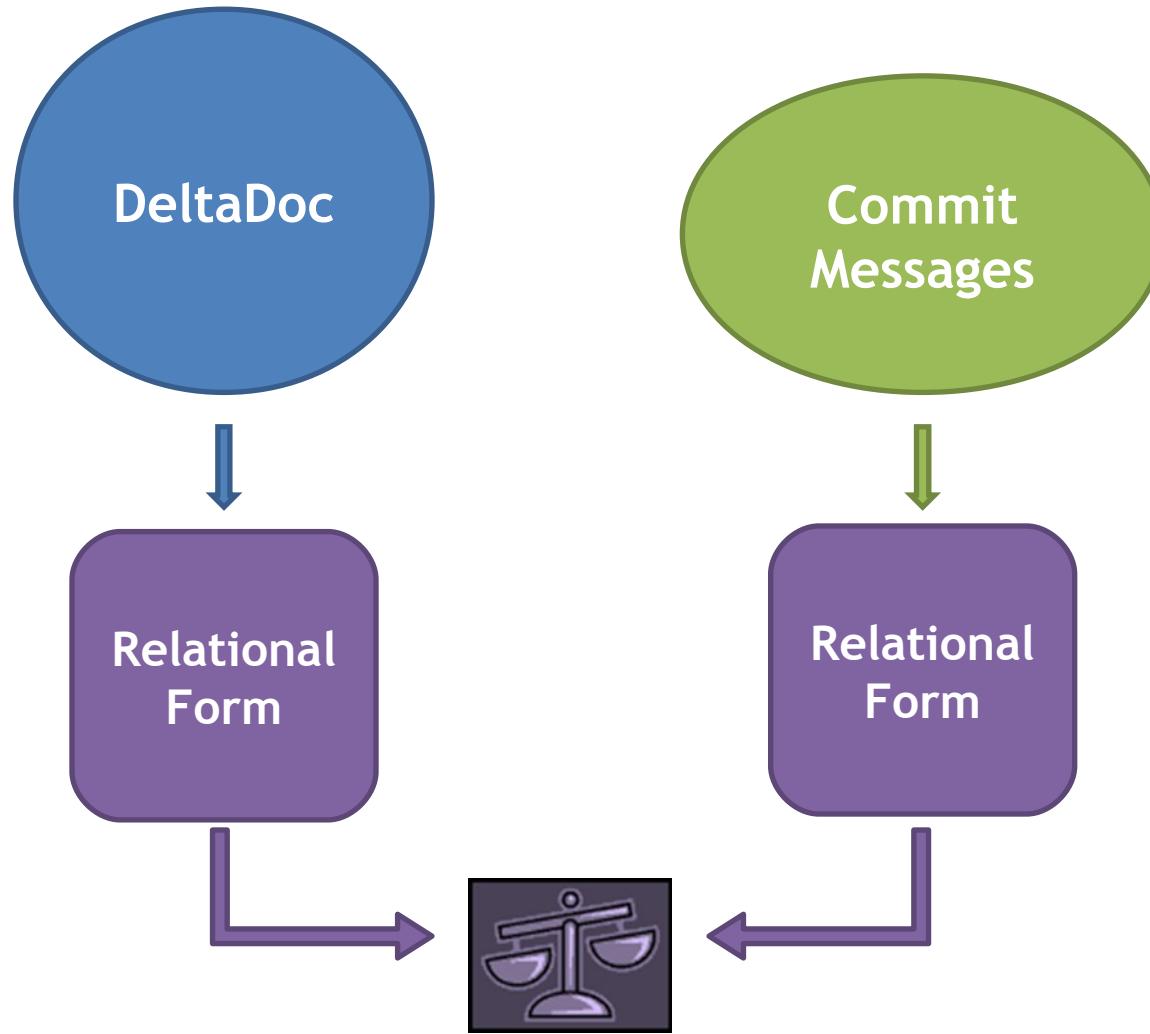
# Score Metric

- Conservatively assume only relations from commit messages are important.

- Reward precision.

- Used 16 human annotators to validate.

- Score of 0.5 indicates that the DeltaDoc contained all the information in the commit message.

# Results



DeltaDoc implies all relations in log message
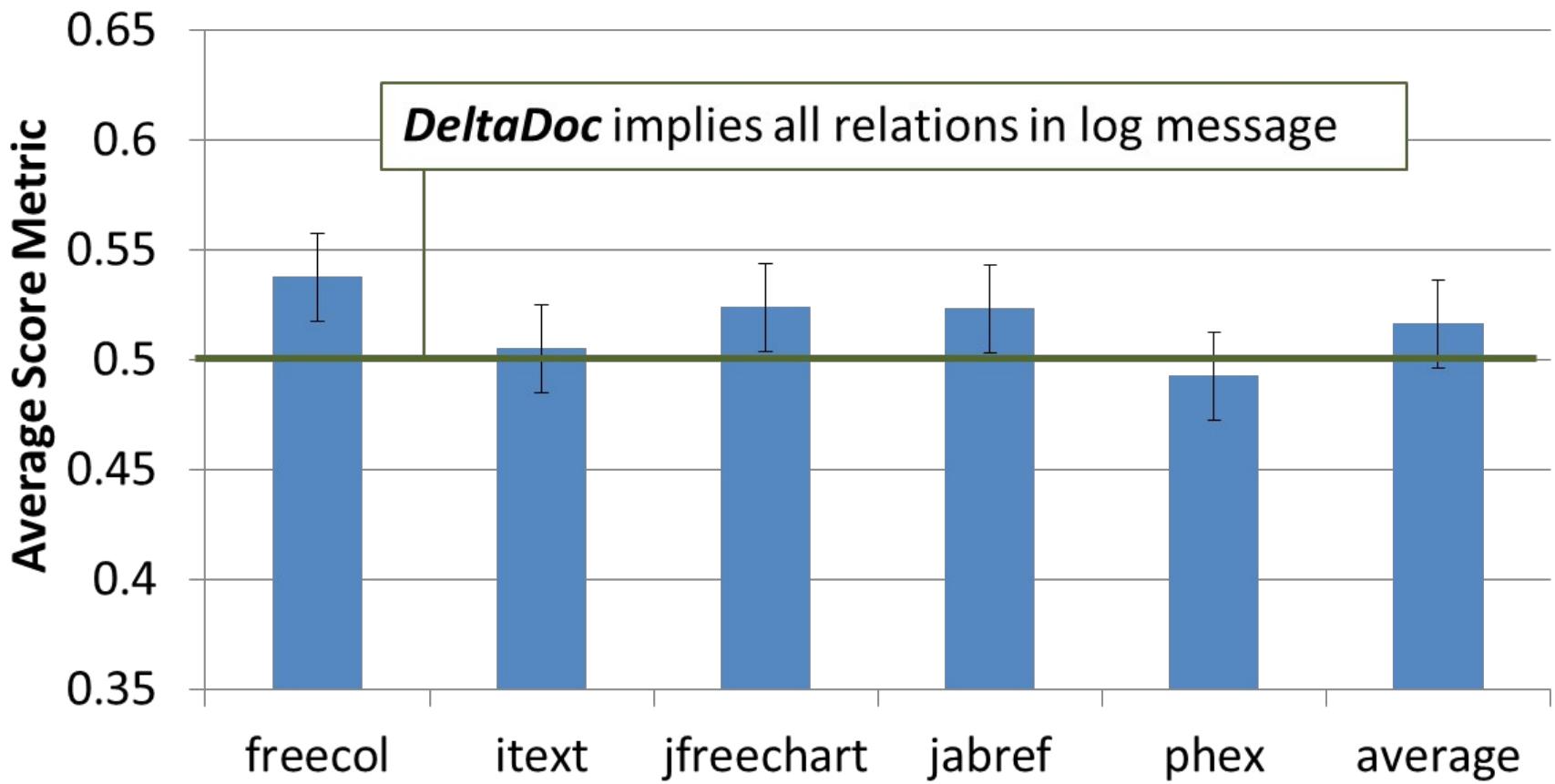
# Results
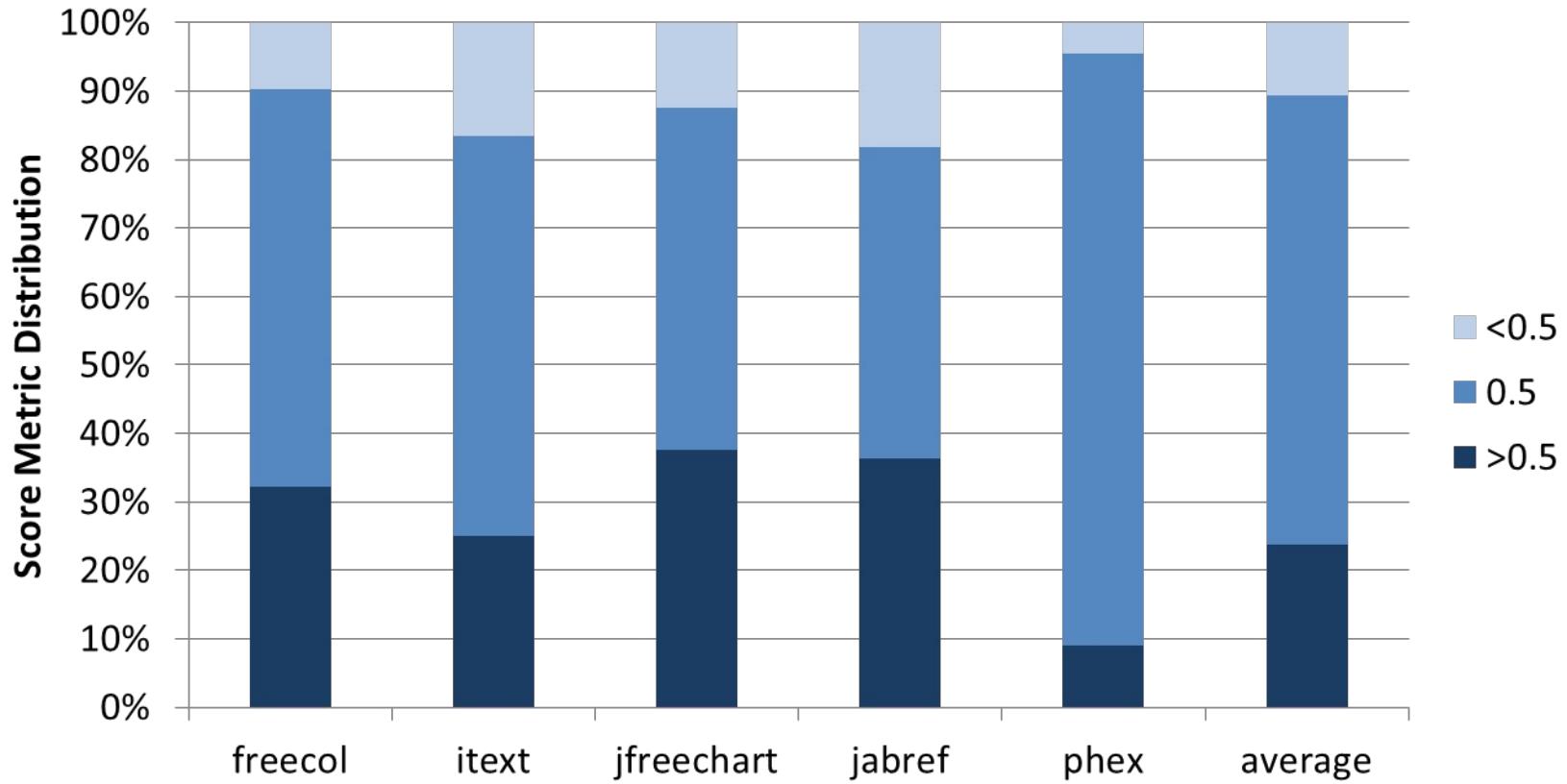
# Results

About 89% coverage.
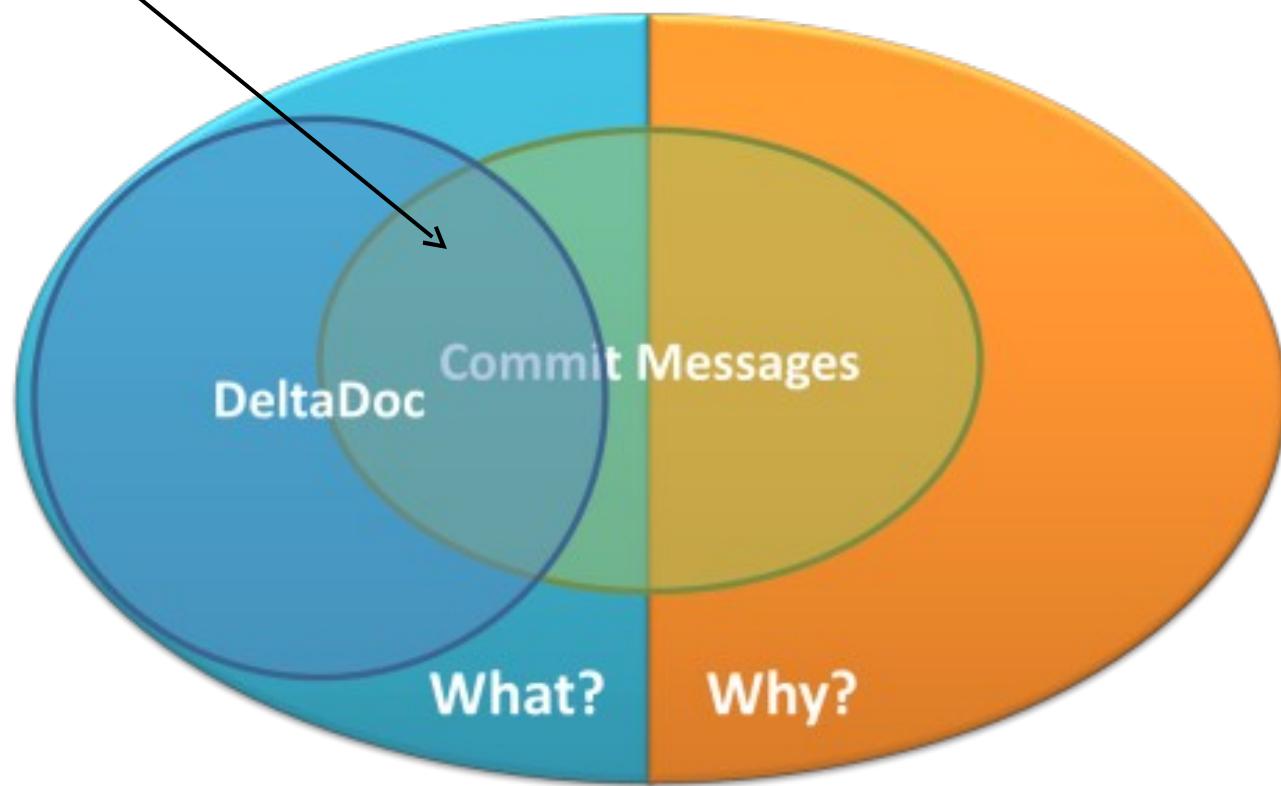
# Qualitative Evaluation

"very useful" "highly useful" "would be a great supplement" "definitely a useful supplement" "can help make the logic clear" "often easier to understand" "more accurate" "easy to read" "provides more information"



Preferred Commit Message 37%

Preferred Deltadoc 46%

No Preference 17%
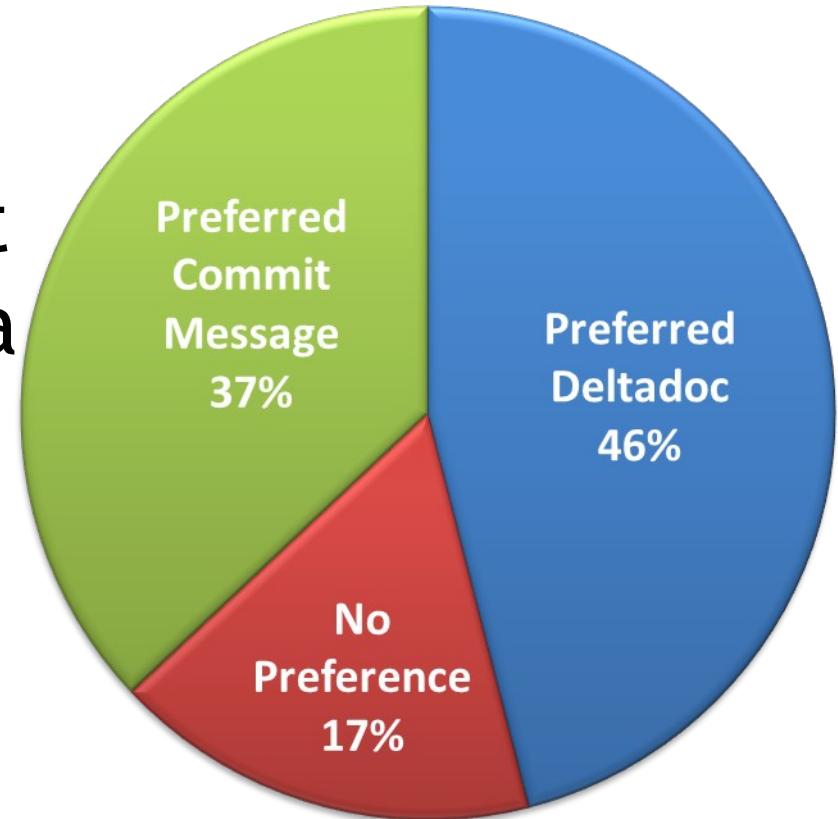
# DeltaDoc

- Cheap
  - Can be computed in about a second on average.
- Suitable for quick adoption
  - Can supplement or replace many existing commit messages.
- Structured
  - Suitable for search.
- Does not address "why"
- Search-based simplification?

# Questions?

# Relational Form

| Type | Arity | Name |
|---|---|---|
| | unary | is *true* |
| | | is *false* |
| logical | | or |
| | | and |
| | | equal to |
| | | not equal to |
| | binary | less-than |
| | | greater-than |
| | | less-than or equal to |
| | | greater-than or equal to |
| | | is empty |
| | | call / invoke |
| | unary | **return** |
| | | **throw** |
| programmatic | | assign to |
| | | element of |
| | binary | inserted in |
| | | implies |
| | | instance of |
| | | added |
| edit | unary | changed |
| | | removed |

# Relational Form Example

| getPriceForBuilding() > getOwner().getGold() | has an insufficient amount of gold |
|---|---|

| getPriceForBuilding() > getOwner().getGold() | ? > gold |
|---|---|
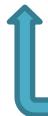
# Score Metric

- Conservatively assume only relations from commit messages are important.

- Reward precision.

- Used 16 human annotators to validate.

- Score of 0.5 indicates that the DeltaDoc contained all the information in the commit message.

# Example Score = 0.5

Commit Message

```
no need to call clear()
```

DeltaDoc

When calling PdfContentByte reset()
    If stateList .isEmpty(),
        No longer call stateList .clear()

iText Rev 3837

# Example Score > 0.5

Commit Message

Commented unused constant

DeltaDoc

removed field : EuropePanel : int TITLE_FONT_SIZE

Freecol rev 2054

# Example Score < 0.5

Commit Message

Fixed bug: content selector for 'editor' field uses ',' instead of 'and' as delimiter.

DeltaDoc

When calling EntryEditor getExtra()
   If ed.getFieldName().equals("editor")
      call contentSelectors .add(FieldContentSelector)

JabRef Rev 3111