Revision Identifier를 이용한 파일 이력 추정 기법

max

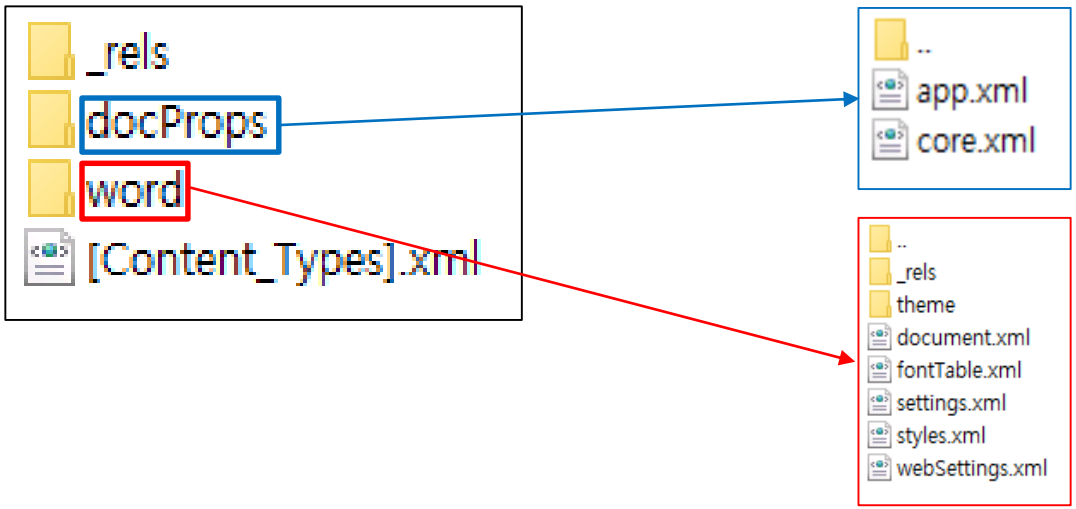
max.joun@gmail.com

Contents

Introduction

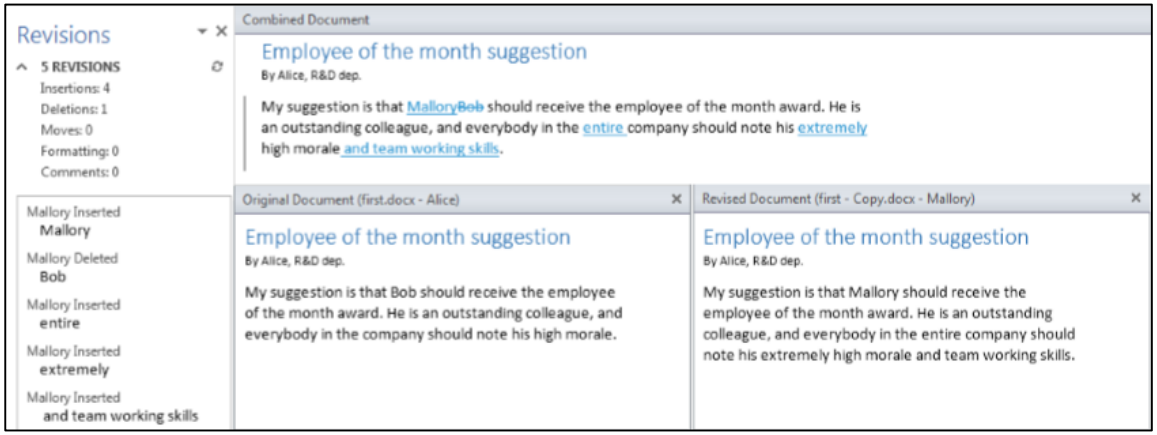
- Electronic documents are able to cause various illegal activities such as illegal copying, confidential leakage, and forgery of contract
- Forensic investigators face great difficulties in investigating and tracking the source of these illegal copies
- Currently, the behavior is tracked using document file contents and internal attribute information. (metadata)
  - It is already known information
  - Can be modified/deleted using tools
- Revision identifiers(RSID) can be one solution for these problems
  - Information leakage
  - Plagiarism, copyright

- OOXML
  - In November 2006, Microsoft Office 2007 was released with full support for OOXML
  - New MS Office format – All data is stored as regular ZIP file
  - Structure
    - ✓ /docProps: core.xml, app.xml → metadata
    - ✓ /word: document.xml, settings.xml → revision identifier



Revision Identifier

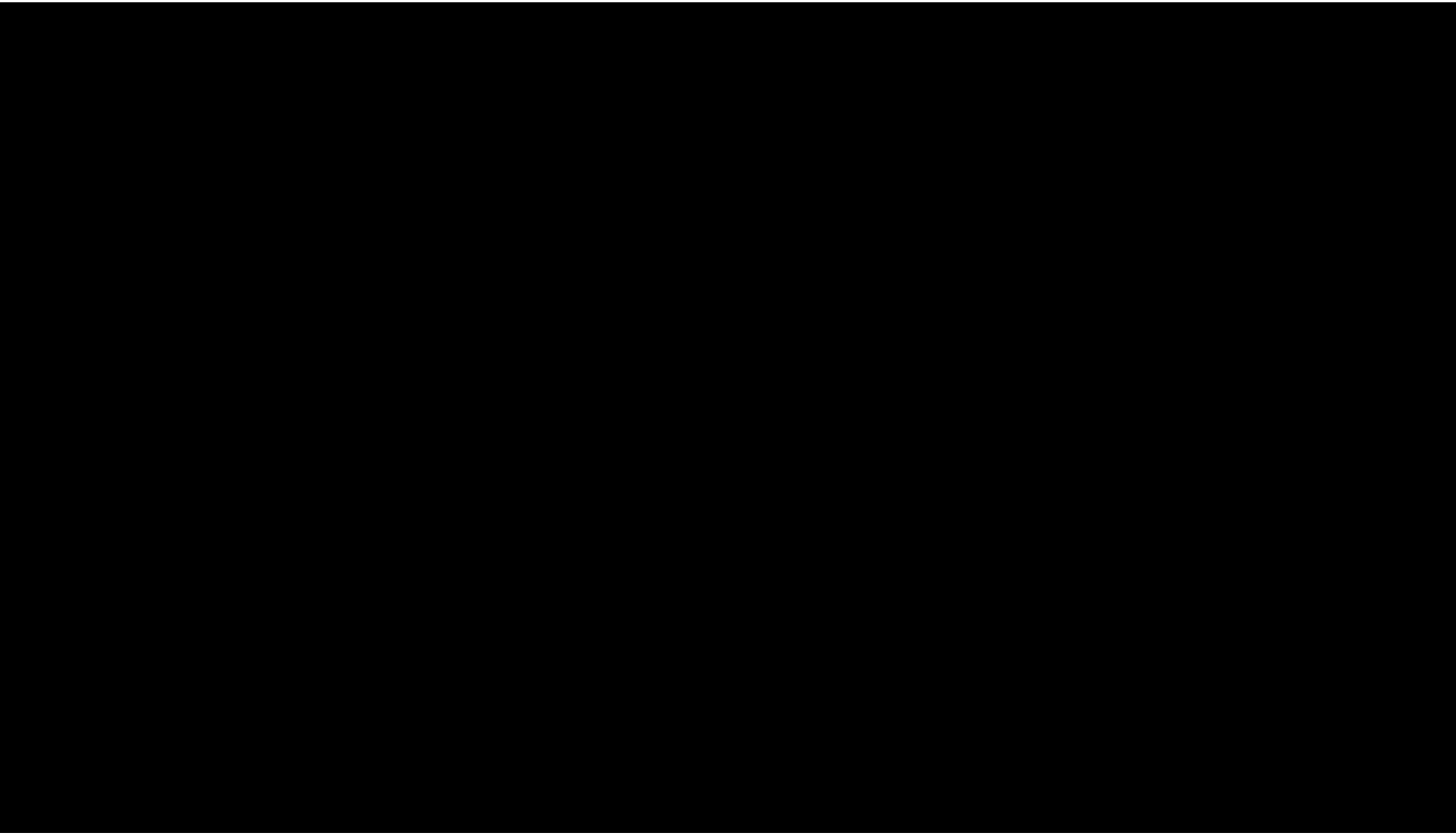
- Revision identifier Concept
  - 8-digit hexadecimal number in 32 bits
  - Effective and accurate way of merging two documents that origin from the same source



- Production rule
  - Random
  - Created/allocated when the document file is saved after modification

Revision Identifier

- RSID 생성 원리



- Revision identifier Concept (Con't)
  - Types of revision identifiers in OOXML documents

#	Name	Description
1	rsidR	used to track the editing session when the paragraph was added to the main document.
2	rsidRDefault	used for all runs in this paragraph which do not explicitly declare an rsidR attribute. This attribute allows consumers to optimize the locations where rsid* values are written in this document
3	rsidP	used to track the editing session when the paragraph's properties were last modified
4	rsidRPr	used to track the editing session when the glyph character representing the paragraph mark was last modified in the main document.

Revision Identifier

- File copy
  - Original file and copied file have always same RSID
  - Same RSID when the user copied to another storage
    - ✓ Other drive
    - ✓ USB
    - ✓ External Drive
    - ✓ Google Drive
    - ✓ Ndrive
    - ✓ OneDrive

Revision Identifier

- File copy [cont.]
  - Add contents in copied file
    - ✓ New RSID allocated to added contents
  - Modify contents in copied file
    - ✓ Allocated RSID value still remain
    - ✓ New RSID allocated to modified contents
  - Delete contents in copied file
    - ✓ Allocated RSID are not existed in "document.xml" file anymore
    - ✓ But, it remains in "settings.xml"

- Content copy
  - Default font
  - Changed font

#	Type	Effectiveness
1	Bold/Italic	O
2	Array	X
3	Color	O
4	Font Type	O
5	Font Size	O
6	Listing	X
7	Style	O
8	Char	X
9	Picture	X

- 버전 별 변경 사항

Version	Difference
MS Word 2007	At the end of the "document.xml" file, the <w: rsidSect> entry is always created and a new RSID value is assigned.
MS Word 2010	<W: rsidSect> which was always generated in MS Word 2007 is not generated.
MS Word 2013	<W: rsidSect> which was always generated in MS Word 2007 is not generated.
MS Word 2016	<W: rsidSect> which was always generated in MS Word 2007 is not generated.

- 문서 애플리케이션 별 변경 사항

Program	RSID exists
Hancom Office	O
MS Office Online	O
Office 365	O
Google Document	O
IOS MS Office	O
Libre Office	O
Open Office	X
Naver Office	X

Experiment

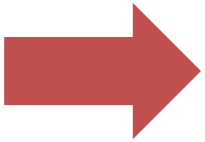
- Experiment setting
  - Windows 7, 8.1, 10, 10 Creator Update, MAC
  - Microsoft Word 2007, 2013, 2016

- Experiment list
  - Create, add, and delete characters
  - Save as, copied file and sending
  - Copy the original contents to another file

Experiment -1

- Add contents
  - When the user creates new word file with some sentences in a word file, "rsidR" and "rsidRDefault" are created
  - After adding more words in a word file, a new "rsidR" created

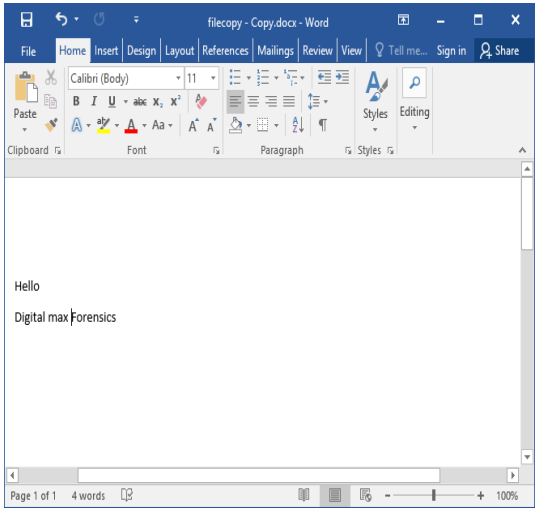
```
- <w:body>
  - <w:p w:rsidRDefault="00754511" w:rsidR="00491356">
    - <w:r>
      - <w:rPr>
          <w:rFonts w:hint="eastAsia"/>
        </w:rPr>
        <w:t>originalData</w:t>
      </w:r>
      <w:bookmarkStart w:name="_GoBack" w:id="0"/>
      <w:bookmarkEnd w:id="0"/>
    </w:p>
  + <w:sectPr w:rsidR="00491356">
  </w:body>
```



```
- <w:body>
  - <w:p w:rsidRDefault="00754511" w:rsidR="00491356">
    - <w:r>
      - <w:rPr>
          <w:rFonts w:hint="eastAsia"/>
        </w:rPr>
        <w:t>originalData</w:t>
      </w:r>
    - <w:r w:rsidR="00274E8E">
      - <w:rPr>
          <w:rFonts w:hint="eastAsia"/>
        </w:rPr>
        <w:t>added</w:t>
      </w:r>
      </w:r>
      <w:bookmarkStart w:name="_GoBack" w:id="0"/>
      <w:bookmarkEnd w:id="0"/>
    </w:p>
  + <w:sectPr w:rsidR="00491356">
  </w:body>
```

Experiment -1

- Add/Delete contents which are copied
  - Even though the user delete or modify copied contents, RSID still remains




```
- <w:body>
  - <w:p w:rsidRDefault="001B61E1" w:rsidR="001E0D96">
    - <w:r>
        <w:t>Hello</w:t>
      </w:r>
    </w:p>
  - <w:p w:rsidRDefault="00153CE2" w:rsidR="00153CE2">
    - <w:r>
        <w:t xml:space="preserve">Digital </w:t>
      </w:r>
    - <w:r w:rsidR="00BF39C1">
        <w:t xml:space="preserve">max </w:t>
      </w:r>
    <w:bookmarkStart w:name="_GoBack" w:id="0"/>
    <w:bookmarkEnd w:id="0"/>
    - <w:r>
        <w:t>Forensics</w:t>
      </w:r>
    </w:p>
  - <w:sectPr w:rsidR="00153CE2">
      <w:pgSz w:w="12240" w:h="15840"/>
      <w:pgMar w:gutter="0" w:footer="720" w:header="720" w:left="1440" w:bottom="1440"
        w:right="1440" w:top="1440"/>
      <w:cols w:space="720"/>
      <w:docGrid w:linePitch="360"/>
    </w:sectPr>
</w:body>
```

```
- <w:body>
  - <w:p w:rsidRDefault="00153CE2" w:rsidR="00153CE2">
      <w:bookmarkStart w:name="_GoBack" w:id="0"/>
      <w:bookmarkEnd w:id="0"/>
    </w:p>
  - <w:sectPr w:rsidR="00153CE2">
      <w:pgSz w:w="12240" w:h="15840"/>
      <w:pgMar w:gutter="0" w:footer="720" w:header="720" w:left="1440" w:bottom="1440"
        w:right="1440" w:top="1440"/>
      <w:cols w:space="720"/>
      <w:docGrid w:linePitch="360"/>
    </w:sectPr>
</w:body>
```

```
- <w:rsids>
    <w:rsidRoot w:val="00B20787"/>
    <w:rsid w:val="00153CE2"/>
    <w:rsid w:val="001B61E1"/>
    <w:rsid w:val="001E0D96"/>
    <w:rsid w:val="00B20787"/>
    <w:rsid w:val="00BF39C1"/>
  </w:rsids>
```

```
- <w:rsids>
    <w:rsidRoot w:val="00B20787"/>
    <w:rsid w:val="00153CE2"/>
    <w:rsid w:val="001B61E1"/>
    <w:rsid w:val="001E0D96"/>
    <w:rsid w:val="003C57F4"/>
    <w:rsid w:val="00B20787"/>
    <w:rsid w:val="00BF39C1"/>
  </w:rsids>
```

Experiment -2

- Save As
  - New RSID added in “settings.xml”

document.xml

```
- <w:body>
    + <w:p w:rsidRDefault="00746B1D" w:rsidR="00F41FE6">
    + <w:p w:rsidRDefault="00DB52AF" w:rsidR="00DB52AF">
    + <w:sectPr w:rsidR="00DB52AF">
  </w:body>
```

=

```
- <w:body>
    + <w:p w:rsidRDefault="00746B1D" w:rsidR="00F41FE6">
    + <w:p w:rsidRDefault="00DB52AF" w:rsidR="00DB52AF">
    + <w:sectPr w:rsidR="00DB52AF">
  </w:body>
```

settings.xml

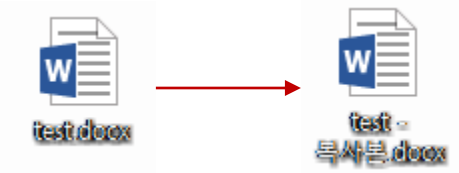
```
- <w:rsids>
    <w:rsidRoot w:val="00F01E50"/>
    <w:rsid w:val="005B4D37"/>
    <w:rsid w:val="00823006"/>
    <w:rsid w:val="00F01E50"/>
```

≠

```
- <w:rsids>
    <w:rsidRoot w:val="00F01E50"/>
    <w:rsid w:val="005B4D37"/>
    <w:rsid w:val="00823006"/>
    <w:rsid w:val="00CB7EEC"/>
    <w:rsid w:val="00F01E50"/>
```

Experiment -2

- Copy & Paste file
  - Same RSID value in both “document.xml” and “settings.xml”



document.xml

```
- <w:body>
    + <w:p w:rsidRDefault="00293001" w:rsidR="00F97F1E">
    + <w:sectPr w:rsidR="00F97F1E">
  </w:body>
```

=

```
- <w:body>
    + <w:p w:rsidRDefault="00293001" w:rsidR="00F97F1E">
    + <w:sectPr w:rsidR="00F97F1E">
  </w:body>
```

settings.xml

```
- <w:rsids>
      <w:rsidRoot w:val="00CC41CE"/>
      <w:rsid w:val="00293001"/>
      <w:rsid w:val="00CC41CE"/>
      <w:rsid w:val="00F97F1E"/>
  </w:rsids>
```

=

```
- <w:rsids>
      <w:rsidRoot w:val="00CC41CE"/>
      <w:rsid w:val="00293001"/>
      <w:rsid w:val="00CC41CE"/>
      <w:rsid w:val="00F97F1E"/>
  </w:rsids>
```

Experiment -2

- Sending through messenger, e-mail, etc.


```
- <w:p w:rsidRDefault="000E39E3" w:rsidR="002764DB">
   - <w:r>
      - <w:rPr>
            <w:rFonts w:hint="eastAsia"/>
        </w:rPr>
        <w:t>SENDING</w:t>
    </w:r>
    <w:bookmarkStart w:id="0" w:name="_GoBack"/>
    <w:bookmarkEnd w:id="0"/>
  </w:p>
- <w:sectPr w:rsidR="002764DB">
```

=

```
- <w:p w:rsidRDefault="000E39E3" w:rsidR="002764DB">
   - <w:r>
      - <w:rPr>
            <w:rFonts w:hint="eastAsia"/>
        </w:rPr>
        <w:t>SENDING</w:t>
    </w:r>
    <w:bookmarkStart w:id="0" w:name="_GoBack"/>
    <w:bookmarkEnd w:id="0"/>
  </w:p>
- <w:sectPr w:rsidR="002764DB">
```

```
- <w:rsids>
      <w:rsidRoot w:val="00506824"/>
      <w:rsid w:val="000E39E3"/>
      <w:rsid w:val="002764DB"/>
      <w:rsid w:val="00506824"/>
  </w:rsids>
```

=

```
- <w:rsids>
      <w:rsidRoot w:val="00506824"/>
      <w:rsid w:val="000E39E3"/>
      <w:rsid w:val="002764DB"/>
      <w:rsid w:val="00506824"/>
  </w:rsids>
```

Experiment -3

- Copy all the contents to new file (w/ default setting)

```
- <w:body>
    + <w:p w:rsidRDefault="00746B1D" w:rsidR="00F41FE6">
    + <w:p w:rsidRDefault="00DB52AF" w:rsidR="00DB52AF">
    + <w:sectPr w:rsidR="00DB52AF">
  </w:body>
```

≠

```
- <w:body>
    + <w:p w:rsidP="00BE29D1" w:rsidRDefault="00BE29D1" w:rsidR="00BE29D1">
      <w:p w:rsidP="00BE29D1" w:rsidRDefault="00BE29D1" w:rsidR="00BE29D1"/>
    + <w:p w:rsidRDefault="00E86E59" w:rsidR="00E86E59" w:rsidRPr="00BE29D1">
    + <w:sectPr w:rsidR="00E86E59" w:rsidRPr="00BE29D1">
  </w:body>
```

- Copy all the contents to new file (w/ bold/font/size changed)

```
- <w:body>
    + <w:p w:rsidP="009B4DF4" w:rsidRDefault="00746B1D" w:rsidRPr="008A6880" w:rsidR="00F41FE6">
    + <w:p w:rsidRDefault="00DB52AF" w:rsidR="00DB52AF">
    + <w:sectPr w:rsidR="00DB52AF">
  </w:body>
```

=

```
- <w:body>
    + <w:p w:rsidP="00951615" w:rsidRDefault="00951615" w:rsidRPr="008A6880" w:rsidR="00951615">
      <w:p w:rsidP="00951615" w:rsidRDefault="00951615" w:rsidR="00951615"/>
    + <w:p w:rsidRDefault="00F95D75" w:rsidR="00F95D75" w:rsidRPr="00951615">
    + <w:sectPr w:rsidR="00F95D75" w:rsidRPr="00951615">
  </w:body>
```

Experiment -3

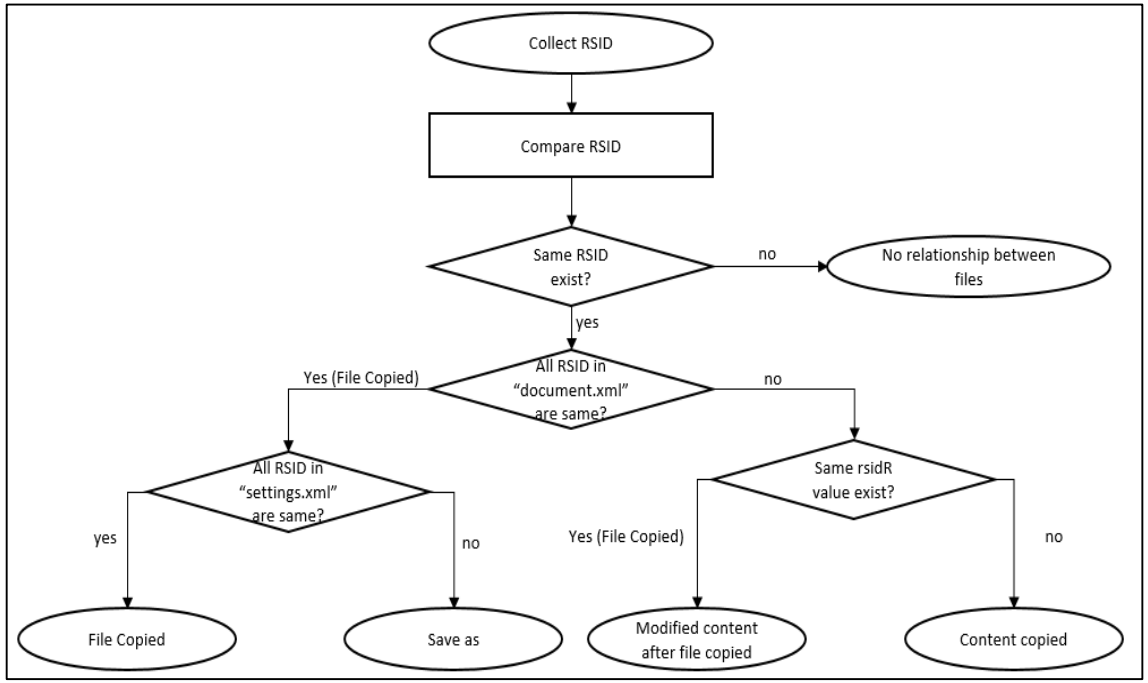
- Copy partial contents to new file (w/ bold/font/size changed)
  - If the content's font, size, etc has been changed, the copied file has same "rsidRPr"

```
- <w:body>
    + <w:p w:rsidP="00951615" w:rsidRDefault="00951615" w:rsidRPr="008A6880" w:rsidR="00951615">
      <w:p w:rsidP="00951615" w:rsidRDefault="00951615" w:rsidR="00951615"/>
    + <w:p w:rsidRDefault="00F95D75" w:rsidRPr="00951615" w:rsidR="00F95D75">
    + <w:sectPr w:rsidRPr="00951615" w:rsidR="00F95D75">
  </w:body>
```

```
- <w:body>
    - <w:p w:rsidP="004251EC" w:rsidRDefault="004251EC" w:rsidR="009E6030">
        - <w:r w:rsidRPr="008A6880">
            - <w:rPr>
                <w:rFonts w:hint="eastAsia"/>
                <w:b/>
              </w:rPr>
              <w:t>디지털 포렌식</w:t>
          </w:r>
          <w:bookmarkStart w:name="_GoBack" w:id="0"/>
          <w:bookmarkEnd w:id="0"/>
      </w:p>
    + <w:sectPr w:rsidR="009E6030">
  </w:body>
```

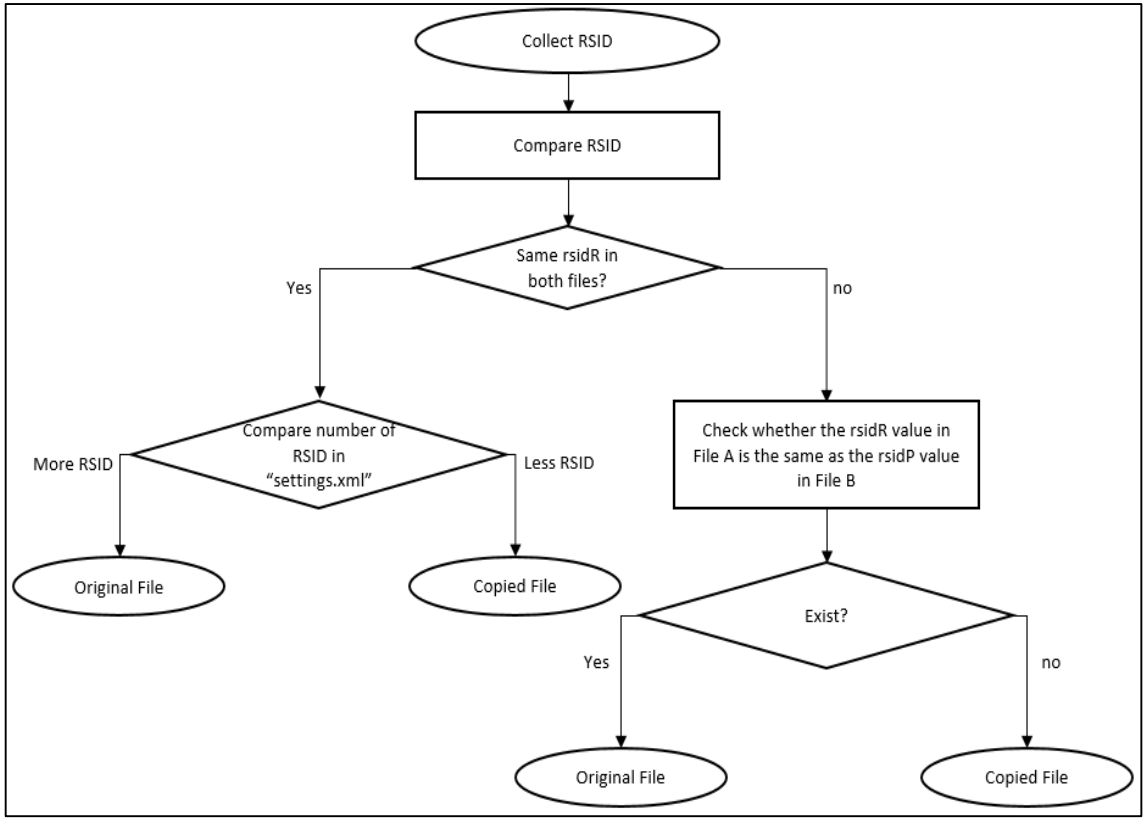
Forensic analysis method

- Method to investigate document file leakage



Forensic analysis method

- Method to distinguish between original file and copied file



Conclusion

- It is new method to investigate document leakage
- It is possible to compare the original file and copied file
- More rules exist to track the user behavior

Question and Answer