Dataset

Suspicious text segment

- **11. location**
- **12. content**
- **13. length**

Suspicious Document

By XXX XX XX X

- **1.Title**
- **2.Author**
- **3. Structural data**
- **4. Content**
- **5.Length (The length can vary from a single paragraph to hundreds of pages)**
- **16. citations**

XXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXX``

. . . . . .

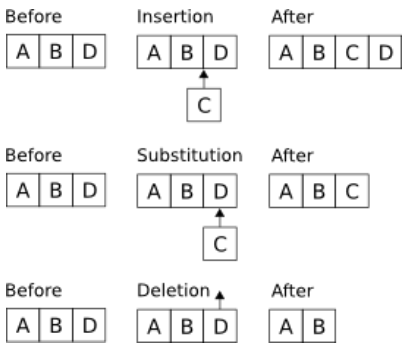
Reference
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXX
… …

14. modification methods (linked with specific text, e.g. insertion, removal, shortened, etc.)

15. Type of plagiarism (e.g.

(Almost) Complete Plagiarism: a section largely produced by copy-and-paste.
Obfuscation: a text passage which is more or less paraphrased.
Pawn Sacrifice: text from a cited source is used but is referred to somewhere else in the document.
Exacerbated Pawn Sacrifice: text is copied straight from a source and a correct reference is cited, but the reference is introduced with "'likewise"' suggesting that there is a similar statement but not equal text.
Shake and Paste: longer text sections, typically paragraphs, are taken and mixed from different sources.
Half Sentence Mending: short sentences or sentence fragments from a source have been used.
Translation Plagiarism: a text translated from a foreign language source which was more or less rephrased.

describing differences between text:



Original Document 1

By XXX XX XX X

- **6. Title**
- **7. Author**
- **8. Structural data**

XXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXX
XXXXXXXX

Original Document 2

Original Document 3

By XXX XX XX X

- **9. Content**
- **10. Length**

XXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXX
XXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXX``

. . . . . .