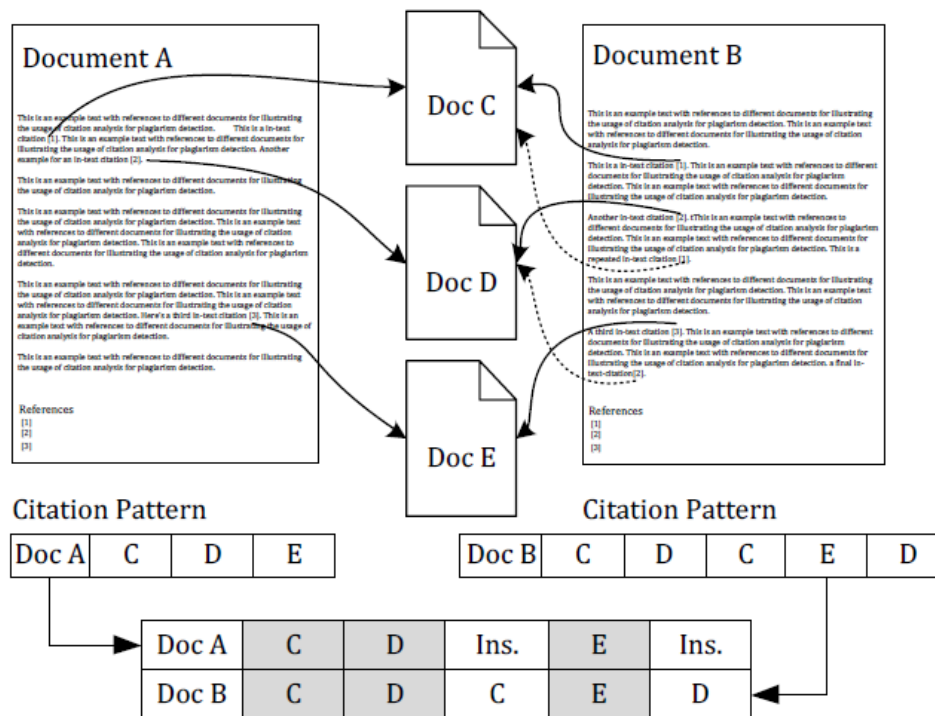


## Citation-Based Plagiarism Detection (CBPD)

“CBPD fasst Methoden zusammen, die Zitate und Quellenangaben verwenden, um Ähnlichkeiten zwischen Dokumenten zu erkennen und dadurch Plagiatismus zu identifizieren.”



Pattern Comparison (Levenshtein Dist.)

### Identifying citation patterns for CbPD

#### Mögliche Ausführungen des CBPD:

1. Bibliographic Coupling Strength  
Absolute Anzahl an gemeinsamen Quellen gibt Ausschluss darüber, ob es sich um ein Plagiat handeln könnte. Aber: Reihenfolge wird ignoriert.
2. Longest Common Citation Sequence  
„Copy and paste“- Plagiatismus wird erkannt, bei „shake and paste“ versagt jedoch das Verfahren.
3. Greedy Citation Tiling  
Eignet sich zusätzlich für „shake and paste“-Plagiatismus.
4. Citation Chunking

Erkennung von Ähnlichkeiten trotz Änderung der Reihenfolge. Textpassagen werden unterschiedlich markiert, je nach Ähnlichkeitsgrad. Dadurch kann auch die Übernahme von logischen Strukturen oder Ideen aufgedeckt werden.

Doc A: x, 1, 2, 3, x, x, 4, 5, x, x, x, x, x, x, 6, 7

Doc B: 3, 2, x, 1, x, x, 4, x, x, x, x, x, 5, 6, 7, x

#### 5. Chunk Comparism

Blöcke für beide Dokumente formen und miteinander vergleichen. Bei hoher Übereinstimmung erfolgt ein Treffer.

### Vorteile und Nachteile des CBPD:

- + Geeignet für Übersetzungen
- + Erkennt Ideen-Plagiatismus
- + Hat kein Problem mit veränderten Satzstrukturen
- Gleiche Themen beinhalten häufig ähnliche Quellen
- Häufig können nicht alle Zitierstile erkannt werden

→ Studien belegen, dass das CBPD Formen von Plagiatismus erkennt, welche bei textbasierten Verfahren untergehen.

#### Literaturverzeichnis:

Gipp, Bela und Meuschke Norman, "Citation Pattern Matching Algorithms for Citation-based Plagiarism Detection: Greedy Citation Tiling, Citation Chunking and Longest Common Citation Sequence"

<http://gipp.com/wp-content/papercite-data/pdf/gipp11c.pdf> [besucht am 15.05.2012]

Longest common subsequence problem solver (Python recipe)

<http://code.activestate.com/recipes/576869-longest-common-subsequence-problem-solver/>  
[besucht am 20.05.2012]

