# Aufgabenbeschreibung: Differenz-Report

Ziel dieser Aufgabe ist die Erstellung eines Programms zur Darstellung von Abweichungen von einer Datei zu einer anderen Datei.

## Allgemeine Beschreibung

Die zu vergleichenden Dateien können beliebigen Dateiformates sein. Wichtig ist eine weitere Datei je Inhaltsdatei, welche die Struktur der Inhaltsdatei beschreibt. Für diese Strukturdatei muss ein festes Format vorgegeben werden.

Der Vergleich der Inhaltszeilen soll anhand von Schlüsseln[[1]](#footnote-1) erfolgen. Folgende Szenarien sind dabei denkbar:

* Eingabedatei A enthält einen Schlüssel x; Eingabedatei B enthält ebenfalls Schlüssel x
  + Die Inhalte der jeweiligen Dateien unterscheiden sich (Flag: c)
  + Die Inhalte der jeweiligen Dateien unterscheiden sich nicht (Flag: i)
* Eingabedatei A enthält einen Schlüssel x; Eingabedatei B enthält Schlüssel x nicht (Flag: d)
* Eingabedatei B enthält einen Schlüssel x; Eingabedatei A enthält Schlüssel x nicht (Flag: n)

## Zielvorstellung Strukturdatei

Für die Strukturdatei ist ein festes Format erforderlich. Über die Strukturdateien werden die eigentlichen Inhaltsdateien beschrieben. Über dieses Verfahren ist es möglich auch zur Laufzeit Inhalte bisher unbekannter Struktur bzw. Dateiformates als Eingabedatei zu nutzen.

In der Strukturdatei muss daher beschrieben werden, von welchem Format die zugehörige Inhaltsdatei ist (im ersten Schritt werden nur Textdateien verarbeitet), damit das Programm entscheiden kann, wie es die Inhaltsdateien lesen muss.

## Ausgabe

Die Ausgabedatei, der Differenz-Report, sollte folgendermaßen aufgebaut sein:

* Anzahl identischer Zeilen
* Anzahl gelöschter Zeilen
* n Gelöschte Zeilen
* Anzahl neuer Zeilen
* n Neue Zeilen
* Anzahl veränderter Zeilen
* n Veränderte Zeilenpaare inklusive Kennzeichnung der Veränderung je Feld (true/false)

Zur Verdeutlichung bitte die Beispieldatei „output.txt“ beachten.

## Beispiel

An folgendem Beispiel soll die Funktionsweise verdeutlicht werden:



Die beiden Eingabedateien „inputA.txt“ und „inputB.txt“ enthalten die zu vergleichenden Inhalte.

In der XML-Datei „strukturA.xml“ ist der Aufbau der Eingabedatei „inputA.txt“ beschrieben.  
Hier ist hinterlegt **wie** (Textdatei, Trennzeichen ist in diesem Fall ein Semikolon) die Eingabedatei zu lesen ist und **was** in den Dateien enthalten ist.

In der XML-Datei „strukturB.xml“ ist der Aufgbau der Eingabedatei „inputB.txt“ beschrieben.  
Die Inhalte sind in beiden Dateien (sowie Strukturdateien) sind grundsätzlich die gleichen, lediglich die Sortierung der Inhalte sowie das verwendete Trennzeichen sind geändert.

Zuletzt wird in der „output.txt“ exemplarisch ein Ergebnis dargestellt. Hier werden die o.g. Flags zur Kennzeichnung genutzt.  
Bei den **C**hanges werden die betroffenen Inhalte aus beiden Eingabedateien untereinander aufgeführt. In einer vorgestellten Zeile wird über die false/true Kennzeichnung dargestellt in welchen Spalten sich die Inhalte unterscheiden. Die Zeilen werden zudem nicht einfach aus den jeweiligen Eingabedateien übernommen sondern sortiert und anschließend in die Ausgabedatei geschrieben.

1. Ein Schlüssel kann auch eine Kombination aus Feldern einer Inhaltszeile darstellen. [↑](#footnote-ref-1)