DB-Matcher

*Daten ---*

Programmiersprache: Visual C#

Ide: Visual Studio

Länge: ~1600 Zeilen

Oberfläche: CMD / Terminal

Plattform: Windows 10/11

Architektur: x64 (64bit)

Primärer Algorithmus: Levenshtein-Distance

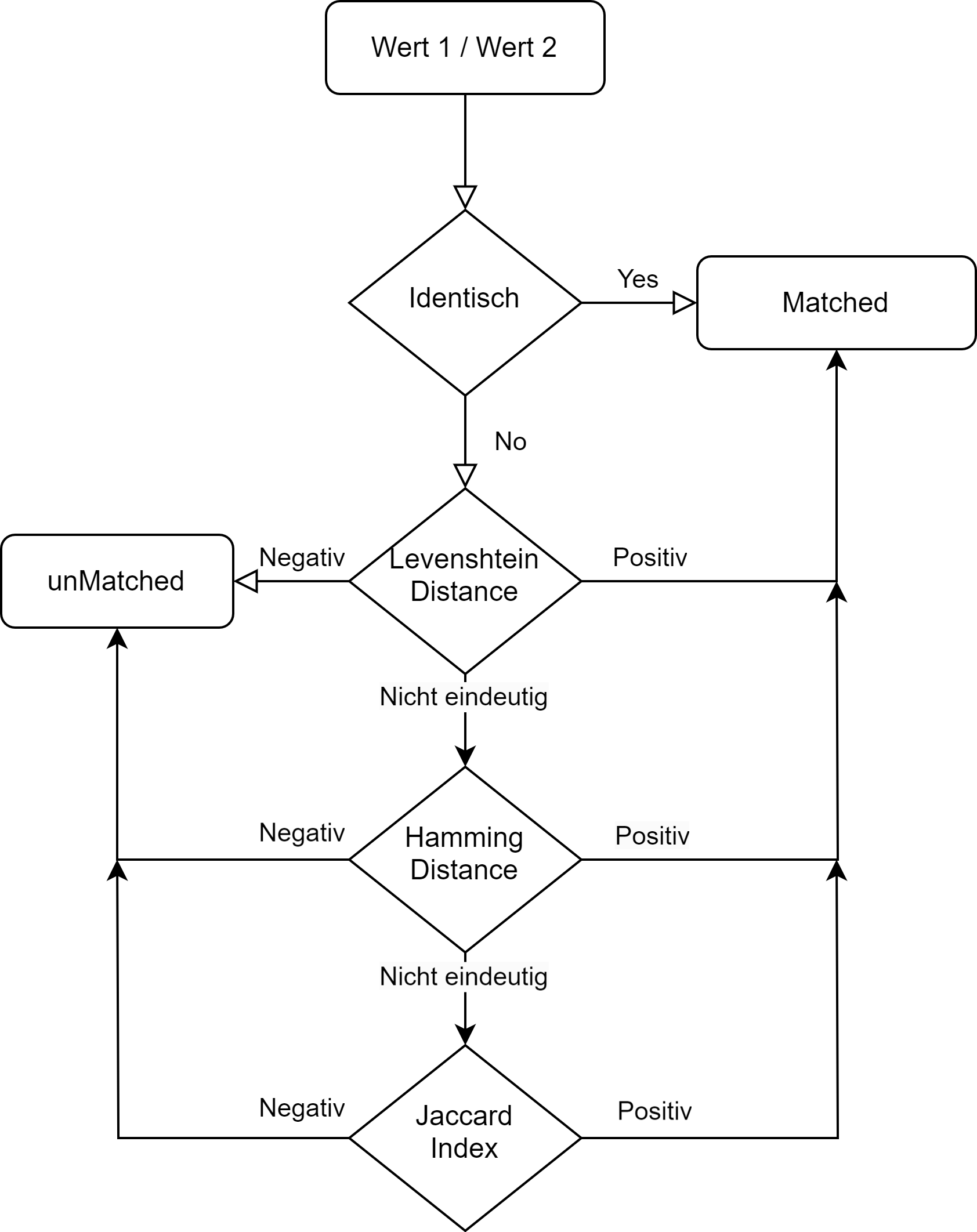
Sekundärer Algorithmus: Hamming-Distance

Tertiärer Algorithmus: Jaccard-Index

*Zusammenfassung ---*

Einsatzbereich: Zusammenführung zweier Datenbanken im Excelformat (\*.xlsx, \*.xls)

Basisfunktion: Der DB-Matcher berechnet, wie ähnlich ein Wert aus der primären Datenbank einem Wert aus der sekundären Datenbank ist. Er folgt dabei dem folgenden Algorithmus-Schema:



Zusatzfunktionen:

* Definierung einer eigenen Datenbank (mittels Wizard in DB-Matcher), in welcher bekannte Unterschiede (z.B. Firmenkürzel) gespeichert werden können. Diese Unterschiede werden beim Matching berücksichtigt. Die erzeugte Datenbank kann auch extern bearbeitet werden.
* Erstellen einer Sicherheitskopie der Excel-Datei um Datenverluste zu verhindern und den Komfort zu erhöhen, da die Ur-Datei weiter in Excel geöffnet bleiben kann, ohne dass es zu Datenstreamkonflikten kommt.
* Automatische dynamische Ermittlung der Datenbanken: Die wahrscheinlich gewünschten Tabellenbereiche werden als default-Wert bei der Eingabe bereitgestellt. Sie können entweder mit ENTER akzeptiert oder per Texteingabe geändert werden. Die ermittelten default-Werte verhalten sich dynamisch zu der geleisteten Eingabe und werden ich Echtzeit aktualisiert
* Fehlerhafte Eingaben werden erkannt und an den Nutzer rückgemeldet.
* Der eingegebene Pfad wird auf Echtheit überprüft
* Der DB-Matcher verfügt über mehrere Einstellungsmöglichkeiten, welche in einer Konfigurationsdatei gespeichert werden, sodass sie nicht bei jedem Start neu definiert werden müssen
* Mögliche Interrupt-Sequenz beim Starten von DB-Matcher
* Die Datenbanken müssen sich nicht auf dem gleichen Arbeitsblatt befinden.
* Die Resultate können ab einer beliebigen Spalte in ein beliebiges Arbeitsblatt geschrieben werden.
* Bei den Resultaten besteht die Möglichkeit, nicht nur die geprüften Werte zu schreiben, sondern zusätzlich die dazugehörenden Zellen mitzuübertragen
* Für Sicherheitsrelevante Datenbanken oder zur Registrierung in andere Prozesse, Programme oder Sicherheitssoftware wird automatisch die Prüfsumme (SHA256) vor und nach dem Matching berechnet und bereitgestellt.
* Ein Fortschrittsbalken wird in der untersten Zeile des Konsolenfensters in Kombination mit einer Zeitanzeige angezeigt. Der Fortschrittsbalken wird dynamisch an die Breite des Konsolenfensters angepasst. Die Zeitanzeige zeigt die berechnete Zeit, die der DB-Matcher vorraussichtlich noch für das Matching benötigt.
* Zu Diagnosezwecken wird die Anzahl an Array-zugriffen angezeigt
* Der Benutzer wird zu jeder Zeit durch das Programm geführt und jede Eingabe ist betitelt