

# Abstract

## German

In medizinischen Systemen, ist es oft nötig, Patienten ohne eindeutigen Identifier zu einem Krankheitsverlauf zu integrieren. Dies ist vor allem bei epidemiologischen Studien und bei Studien notwendig, bei denen Überlebensraten betrachtet werden sollen. Auch für die internationale Zusammenarbeit im Biobank-Bereich sind zusammenhängende Patientenakten nötig. Phonetische Vergleichsalgorithmen, Vergleichsalgorithmen für Zeichenketten und Tippfehlerkorrekturen sind bei Namenmöglich, um Patientenakten zu kombinieren. Doch sind nur der Vorname und der Nachname nicht ausreichend, um eine Entscheidung zu treffen. Weitere Parameter sind nötig um eine korrekte Zusammenführung zu ermöglichen. Wie hier gezeigt wird, steigt durch das Hinzufügen des Geburtsdatums die Übereinstimmung für den Gold Standard bei den Testern von Kappa 0.0135 auf Kappa 0.816. Um die Übereinstimmung noch weiter zu verbessern, zusätzliche Information über den Patienten ist nötig.

**Stichwörter:** Phonetische Vergleichsalgorithmen, Vergleichsalgorithmen für Zeichenketten, Tippfehlerkorrekturen, Namenskorrektur, Datumskorrektur

## English

In medical systems, it is important to get complete patient records, also if the records have no unique identifier. This is especially the case in epidemiological studies and studies that need to analyze survival rates. Furthermore, completed patient records are also needed within the scope of international cooperation in the field of Biobanks. Phonetic similarity, plain string similarity and typing errors algorithms can help to find misspelled names. But with only the first and last name of the patient a trusted combination is not possible. More parameters are needed to assure correct merging. In this work I show that, with adding the date of birth to the combination, the agreement on the gold standard rises from Kappa 0.0135 to Kappa 0.816. To increase the agreement even further additional patient attributes are needed.

**Keywords:** phonetic similarity, plain string similarity, typing errors, typos, date correction