

Tutorium „Einführung in Theorie und Anwendung von Record-Linkage-Verfahren“

Kursbeschreibung

Durch die Verknüpfung bereits vorhandener Forschungsdaten aus verschiedenen Quellen (Record-Linkage) kann deren Analysepotential bei relativ geringem Ressourceneinsatz beträchtlich gesteigert werden. Allerdings erfordert eine Erschließung von Daten in diesem Sinne in der Regel den Einsatz technisch anspruchsvoller Record-Linkage-Verfahren, da die meisten Datenbestände in Deutschland nicht mit eindeutigen Identifikationsnummern versehen sind, sondern nur über Identifikatoren wie Namen und Adressangaben verknüpft werden können. Eine zusätzliche Schwierigkeit erwächst, wenn die Verwendung solcher Identifikatoren aufgrund von Datenschutzbestimmungen nur in verschlüsselter Form möglich ist.

Die erfolgreiche Anwendung von Record-Linkage-Verfahren erfordert den Einsatz geeigneter Software und setzt Kenntnisse ihrer theoretischen Grundlagen voraus. Daher besteht das Tutorial aus zwei Teilen. Zunächst soll in die Theorie von Record-Linkage-Verfahren eingeführt werden. Die Auswahl der Themen erfolgt dabei in Hinblick auf die Ausführung typischer Record-Linkage-Anwendungen. Besprochen werden u.a. deterministisches, distanzbasiertes und probabilistisches Record-Linkage, Stringähnlichkeitsfunktionen, wertespezifische Gewichte, Blocking-Verfahren, Parameterschätzung und Schwellenwertbestimmung. Neben Hinweisen zu vertiefender Literatur und einer Übersicht gebührenfreier Softwarelösungen werden zusätzlich Möglichkeiten besprochen, wie auch verschlüsselte Identifikatoren fehlertolerant abgeglichen werden können. Im zweiten Teil des Tutorials wird dann die praktische Anwendung der besprochenen Verfahren mit der Merge Toolbox (MTB) demonstriert.

Dozent

Tobias Bachteler
Zentrum für Record Linkage

Dauer: 4 x 45 Minuten

Teilnehmerzahl: Die Teilnehmerzahl ist auf 20 beschränkt.