

IAB-DiskAB

Einladung

Automatische Kodierung von Berufen in die KldB 2010 und Ideen für neue Datenerhebungsinstrumente in IAB Surveys

Malte Schierholz (KEM)

Die Erfassung des Berufs in Umfragen ist aufwändig: Die berufliche Tätigkeit wird üblicherweise mit zwei bis drei offenen Fragen erfasst und erst nach der Umfrage in eine Klassifikation (KldB 2010 oder ISCO-08) eingeordnet. Dies ist fehleranfällig, wenn Freitextangaben allgemein oder mehrdeutig sind, und zeitintensiv, da ein Großteil der Antworten nur manuell kodiert werden kann. Zur automatischen Kodierung werden bisher oft umfangreiche Diktionäre verwendet, in denen die zu einer Antwort passende Berufskategorie nachgeschlagen wird.

Eine Alternative zur automatischen Kodierung mittels Diktionären bieten Machine-Learning-Methoden. Dabei werden bereits kodierte Berufsangaben verwendet, mit deren Hilfe für neue Antworten mögliche Berufskategorien automatisch vorgeschlagen werden. Im Vortrag werden verschiedene derartige Methoden diskutiert. Eine Weiterentwicklung ist geplant und soll eine vollautomatisierte Kodierung für eindeutige Antworten und eine teilautomatisierte Kodierung ermöglichen. Um auch das Problem der allgemeinen und mehrdeutigen Berufsangaben aus Interviews in den Griff zu bekommen, ist auch ein Test zur Kodierung während des Interviews geplant.

Dienstag, 29. April, 2014

12.00 Uhr

Raum 126a