Projekt PETRUS, Parallele Ausgaben, Stufe 2

Im Jahr 2011 wurde als PETRUS Szenario 3 ein Match- und Mergeverfahren implementiert, das Datensätze von parallel erschienenen, monografischen Online- und Printpublikationen miteinander verknüpft und materialartunabhängige Informationen wie Inhaltserschließungsdaten und Normdatenverknüpfungen wechselseitig austauscht. In der mehrjährigen Praxis haben sich eine Reihe von Ansätzen zur Anpassung und Weiterentwicklung des Verfahrens ergeben.

Seit Dezember 2012 werden in monografischen Datensätzen Informationen zu sogenannten „related products“ beim Import von Fremddaten übernommen. Diese umfassen Identifier (ISBNs, DOIs etc.) von verwandten Ausgaben, beispielsweise Vorgängern/Nachfolgern oder parallel erscheinende Ausgaben auf verschiedenen Datenträgern. Am 26. Mai 2014 ist eine neue Version des PETRUS-Match- und Mergeverfahrens produktiv genommen worden, welche die Angaben zu "related products" erstmalig zur Identifikation von parallel erschienenen, monografischen Online- und Printpublikationen nutzt. Am Beginn des Abgleichsverfahrens werden die Identifier ausgewertet und gefundene Paare anschließend dem gängigen Match- und Mergeverfahren unterzogen. Gleichzeitig fand eine Fehlerbereinigung statt: Bisher fanden keine maschinellen Verknüpfungen zwischen parallelen Ausgaben statt, wenn in den Datensätzen "related products" enthalten waren. Dies stellt nun kein Hemmnis mehr dar und es wird erwartet, dass dadurch die Anzahl verknüpfter Datensätze deutlich steigt. Ein Retrolauf ist in den nächsten Wochen geplant.

Eine weitere Verbesserung ist im Bereich der Übernahme von Verknüpfungen zu GND-Personen- bzw. GND-Personennamenssätzen erzielt worden. Teilweise werden die Netzpublikationen ohne Angaben zu beteiligten Personen geliefert. Dies betrifft in großem Maße die Veröffentlichungen des Springer-Verlages und des Deutschen Universitätsverlages. In den parallelen Druckausgaben sind die Angaben i. d. R. enthalten. Bisher fand kein Transfer der GND-Verknüpfungen im Bereich Personen-/Personennamen statt. Seit dem 25. Juli 2014 werden die Datensätze der Netzpublikationen um diese GND-Verknüpfungen angereichert.

Antje Kubbernuß und Tanja Maisner