

Historische Begriffe der Erziehungswissenschaft - Erzeugung einer Ontologie

,
l.mueller@dipf.de
DIPF, Deutschland

Historische Begriffe der entstehenden Erziehungswissenschaft sollen als maschinenlesbare Terminologie für digitale historische Bildungsforschung bereitgestellt werden. Bibliografische Titelaufnahmen von Lemmata aus 24 historischen erziehungswissenschaftlichen Nachschlagewerken (1774 – 1942) werden hierfür in eine Ontologie transformiert (Abbildung 1). Die Lexikonbeiträge dokumentieren Herausbildung und Wandel erziehungswissenschaftlicher Fachsprache und Gegenstände. Christian Ritzi (2003: 114) zeigte anhand der „Prügelstrafe“, wie diese "in pädagogischen Nachschlagewerken als pädagogisches Problem im Sinne einer 'Kodifizierung des Wissens' in das System des ausgebreiteten pädagogischen Wissens integriert" wurde. Das Beispiel illustriert das Potential, das für die Forschung in einem integrierten, semantisch modellierten, historischen Vokabular liegt. Im Fachportal Scripta Paedagogica Online (SPO) (vgl. BBF) sind die digitalisierten und detailliert mit Metadaten beschriebenen Lexika (Ritzi 2003: 103ff.) bereits frei zugänglich.

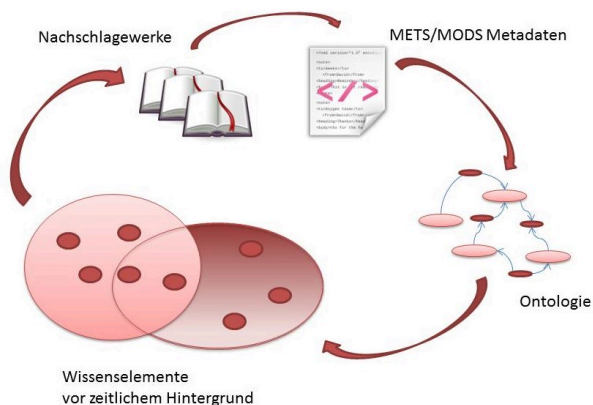


Abb. 1: Rekonstruktion einer Wissensdomäne aus bibliografischen Daten

Ell et al. (2013) haben gezeigt, wie Bibliografische Daten zur Bildungsgeschichte in die virtuelle Forschungsumgebung Semantic CorA (vgl. German Institute for International Educational Research et al.) zur semantischen Repräsentation und Analyse integriert werden können. Die digitalen Dienste der Bibliothek für

Bildungshistorische Forschung (BBF) können künftig über die Bereitstellung von Daten hinausgehen und gegenüber den bestehenden Funktionen um eine Wissens-Ebene erweitert werden (vgl. Oramas et al. 2014; Feng et al. 2005), um die Bildungsgeschichte bei ihrer Hinwendung zu digitalen Methoden optimal zu unterstützen.

Im hier dargestellten Ansatz werden bibliografische Titeldaten als Begriffe bzw. Entitäten konzeptualisiert, neu modelliert und semantisch angereichert. So können sie in bildungshistorischen Analysen ausgewertet oder als Datensets für Textanalysen verwendet werden. Diese zu Forschungsdaten transformierten Katalogdaten erweitern somit die Informationsinfrastruktur für die historische Bildungsforschung auf der Wissens-Ebene und stellen exemplarisch einen neuen bibliothekarischen Infrastrukturservice für Digital Humanities dar, der in der Erzeugung und Bereitstellung semantischer Vokabulare besteht.

Als Ausgangsmaterial liegen in SPO frei zugängliche METS / MODS-Metadaten digitalisierter Nachschlagewerke und Lemmata im RDF / XML-Format vor. Mittels XSLT können sie auf die für das Vorhaben relevanten Elemente reduziert und umgeformt werden. Die als Wissens-elemente betrachteten Titeldaten werden dabei aus der Wissensdomäne Buch- und Bibliothekswesen in die Wissensdomäne historische Bildungsforschung überführt. Die Ontologie wird als Linked Open Data (LOD) in RDF aufgebaut.

Jedes Lemma wird im ersten Verarbeitungsschritt zum Deskriptor eines Konzepts, das durch den zugehörigen Lexikonbeitrag definiert wird. Das Vokabular wird in SKOS (vgl. W3C 2012) modelliert, indem die Instanzen umklassifiziert werden. In den Metadaten befindliche Identifier für Katalogeinträge und digitale Dokumente werden mitgenommen und erhalten die Beziehung zu den Quellen. Im folgenden Schritt werden die erzeugten Vokabulare so überarbeitet, dass explizite Begriffsbeziehungen innerhalb jedes einzelnen Nachschlagewerks abgebildet werden. Zunächst wird jedes Nachschlagewerk einzeln transformiert. Anschließend werden die so entstandenen historischen Vokabulare zur Pädagogik durch Terminologiemapping (vgl. Keil 2012) aufeinander bezogen.

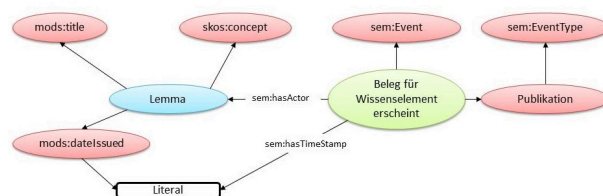


Abb. 2: Schematische Darstellung des Transformationsmodells

Eine besondere Herausforderung stellen Erhalt und Darstellung des zeitlichen Wandels von Sprachgebrauch

und Kategorien dar. Für die Abbildung des Zeitbezugs wird jedes Nachschlagewerk als Snapshot-Ontologie modelliert (vgl. Ide / Woolner 2007: 142; Kauppinen / Hyvönen 2007). Jeder Begriff wird mit dem Erscheinungsdatum des ihn beinhaltenden Nachschlagewerks assoziiert, welches als „Publikationsereignis“ im Simple Event Model (SEM) (van Hage et al. 2011) modelliert wird (Abbildung 2). Da die aus den Lexika extrahierten Lemmata viele Überschneidungen aufweisen werden, sollen sie abschließend auch in einer einzigen Ontologie zusammengeführt werden (vgl. Kalfoglou / Schorlemmer 2003: 4).

Im Ergebnis entsteht eine Ontologie zu historischen Begriffen der Erziehungswissenschaft, die zur Nachnutzung in einem geeigneten Forschungsdatenrepositorium publiziert wird. Damit die Zuverlässigkeit der Daten gewährleistet und überprüft werden kann, werden auch Datensätze der Teilergebnisse zusammen mit den Metadaten des Transformationsprozesses dokumentiert. Damit bleibt jeder Verarbeitungsschritt replizierbar (Abbildung 3). Aus der neuen Datenbasis lassen sich Versionen ableiten, die bspw. als Gazetteer für automatische Entitätenerkennung, Entity Linking, Netzwerkanalysen oder visuelle Datenexploration eingesetzt werden können.

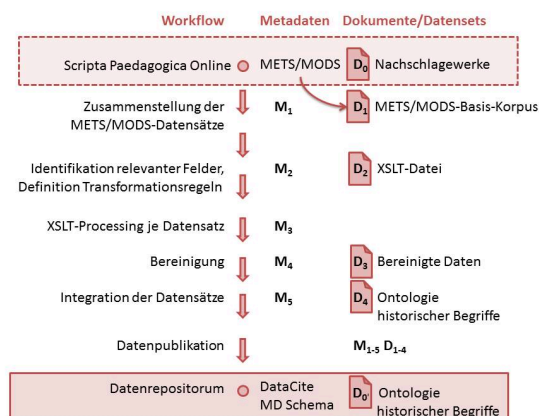


Abb. 3: Workflow zur Transformation und Erzeugung von Daten- und Metadatenansätzen

Bibliographie

Bibliothek für Bildungsgeschichtliche Forschung (BBF) (o.J.): SPO. Scripta Paedagogica Online <http://goobiweb.bbf.dipf.de/viewer> [letzter Zugriff 12. Februar 2016].

Ell, Basil / Schindler, Christoph / Rittberger, Marc (2013): "Semantically Enhanced Interactions between Heterogeneous Data Life-Cycles", in: Garoufallou, Emmanouel / Greenberg, Jane (eds.): *Metadata and Semantics Research*. Cham: Springer International Publishing 277–288.

Feng, Ling / Jeusfeld, Manfred A. / Hoppenbrouwers, Jeroen (2005): "Beyond information searching and browsing: acquiring knowledge from digital libraries", in: *Information Processing & Management* 41, 1: 97–120 <http://conceptbase.sourceforge.net/mjf/itrs008.pdf> [letzter Zugriff 12. Februar 2016].

German Institute for International Educational Research / Karlsruher Institute for Technology / Georg-August-Universität Göttingen (o. J.): *Semantic CorA* [letzter Zugriff 12. Februar 2016].

Ide, Nancy / Woolner, David (2007): "Historical Ontologies", in: Ahmad, Khurshid / Brewster, Christopher / Stevenson, Mark (eds.): *Words and Intelligence II*. Dordrecht: Springer Netherlands 137–152.

Kalfoglou, Yannis / Schorlemmer, Marco (2003): "Ontology mapping: the state of the art", in: *Knowledge Engineering Review* 18, 1: 1–31 <http://drops.dagstuhl.de/volltexte/2005/40/pdf/04391.KalfoglouYannis.Paper.40.pdf> [letzter Zugriff 12. Februar 2016].

Kauppinen, Tomi / Hyvönen, Eero (2007): "Modeling and Reasoning About Changes in Ontology Time Series", in: Sharman, Raj / Kishore, Rajiv / Ramesh, Ram (eds.): *Ontologies. A handbook of Principles, Concepts and Applications in Information Systems*. Boston, MA: Springer 319–338.

Keil, Stefan (2012): "Terminologie Mapping: Grundlagen und aktuelle Normungsvorhaben", in: *Information - Wissenschaft & Praxis* 63, 1: 45–55 <http://eprints.rclis.org/16716/1/Stefan%20Keil%20TM%20Grundlagen%20und%20Normungsvorhaben-Repository.pdf> [letzter Zugriff 12. Februar 2016].

Oramas, Sergio / Sordo, Mohamed / Serra, Xavier (2014): "Automatic creation of knowledge graphs from digital musical document libraries", in: *9th Conference on interdisciplinary musicology - CIM14*. Proceedings http://mtg.upf.edu/system/files/publications/CIM14_MAIN.pdf [letzter Zugriff 12. Februar 2016].

Ritzi, Christian (2003): "Scripta Paedagogica Online : Digitales Textarchiv zur Bildungsgeschichte des deutschsprachigen Raums", in: Thaller, Manfred (ed.): *Digitale Bausteine für die geisteswissenschaftliche Forschung*. Göttingen: Dührkohp & Radicke 103–135.

van Hage, Willem Robert / Malaisé, Véronique / Segers, Roxane / Hollink, Laura / Schreiber, Guus (2011): "Design and use of the Simple Event Model (SEM)", in: *Web Semantics: Science, Services and Agents on the World Wide Web* 9, 2: 128–136.

W3C = World Wide Web Consortium (2012): *SKOS. Simple Knowledge Organization System* [letzter Zugriff 12. Februar 2016].