

Die Begriffe der Geschichte. Der „Historische Thesaurus (HIT)“ als Normdatenressource für die historisch arbeitenden Geisteswissenschaften

Köther, Felix

felix.koether@herder-institut.de

Herder-Institut für historische Ostmitteleuropaforschung –
Institut der Leibniz-Gemeinschaft, Deutschland
ORCID: 0000-0002-6432-1431

Donig, Simon

simon.donig@herder-institut.de

Herder-Institut für historische Ostmitteleuropaforschung –
Institut der Leibniz-Gemeinschaft, Deutschland
ORCID: 0000-0002-1741-466X

Einleitung

Die historisch arbeitenden Geisteswissenschaften haben ein Normdaten-Problem. Es fehlen insbesondere übergreifende digitale Begriffs- und Klassifikationssysteme, die spezifisch auf den Bedarf historischer Disziplinen ausgerichtet sind. Auch die Debatte darüber, welche Qualitätskriterien historische kontrollierte Vokabulare erfüllen sollen, steht noch am Anfang und ist, vorsichtig gesagt, *under construction*. Verfahrensweisen, wie fachspezifische Prämissen – beispielsweise die Historizität von Begriffen und Bezeichnungen, die Abgrenzung von Quellen- und Forschungsbegriffen und alle weiteren Fragen und Herausforderungen der historischen Semantik – strukturiert, interoperabel und mehrsprachig abgebildet werden können, sind bislang kaum breiter diskutiert, geschweige denn abgestimmt worden. Einzelne Ansätze waren bislang domänenspezifisch ausgerichtet (stellvertretend: Gelati, 2019), weit entwickelt sind vor allem die Sozialwissenschaften und die Kunst- oder Sammlungswissenschaften (stellvertretend: Zapilko et al., 2016; Baca, 2003; Harpring, 2010; Roche/Damas/Roche, 2014).

Von Einführungen und Überblickstexten zu den *Digital Humanities* im Allgemeinen und der *Digitalen Geschichte* im Besonderen wurde das Thema in der jüngsten Ver-

gangenheit nicht intensiver aufgegriffen (vgl. u. a. Jannidis/Kohle/Rehbein, 2017; Döring et al., 2022). Im Gegenteil scheint die ressourcenintensive Grundlagenarbeit, die mit dem Aufbau solcher Systeme verbunden und auf absehbare Zeit auf intellektuelle Arbeitsschritte angewiesen ist, in der aktuellen Euphorie um automatisierte Erschließungsformen und (generative) künstliche Intelligenzen eher noch weiter ins Hintertreffen zu geraten.

Für das digitale Arbeiten, für Indexierung, Annotation, Retrieval, Matching und die maschinelle Erschließung von historischen Forschungsdaten sind und bleiben geeignete Normvokabulare dennoch von fundamentaler Bedeutung. Forschungsprojekte wie auch digitale Infrastrukturen sind vielfach allein auf die großen, semantisch flachen Normdateien, einige Spezialvokabulare oder die Nutzung fachexterner Ressourcen angewiesen. Die historischen Disziplinen benötigen zeitnah geeignete terminologische Werkzeuge, wollen sie ihre Forschungsgebiete mit der notwendigen Selbstbestimmung und fachlichen Autonomie ins Digitale überführen.

Der „Historische Thesaurus“

Der „Historische Thesaurus“ ist ein Information-Retrieval-Thesaurus für die historisch arbeitenden Geistes- und Kulturwissenschaften, der aktuell am *Herder-Institut für historische Ostmitteleuropaforschung* entwickelt und aufgebaut wird. Er vereint Begriffsgruppen, die bislang entweder nur unspezifisch oder über thematisch enggefasste Spezialvokabulare, vielfach jedoch noch gar nicht in Form kontrollierter Vokabulare verfügbar sind. Er umfasst beispielsweise Begriffe zu historischen wie aktuellen Ideologie- und Theoriebegriffen, Ereignissen und Akteuren, Epochen und Zeiträumen, Praktiken, Objekt- und Werktypen oder Disziplinen.

Zu seinen Zielen gehört einerseits die sukzessive Bereitstellung von möglichst vielfältigen Deskriptoren für das historische Arbeiten; andererseits überhaupt die inkrementelle Entwicklung und Erprobung von Verfahrensweisen zur terminologischen Kontrolle, Auszeichnung und Relationierung historischer Begriffe. Dazu zählen Strategien der semasiologisch wie onomasiologisch diachronen Disambiguierung ebenso, wie zur Bewältigung der erheblichen translatorischen und Matching-bezogenen Herausforderungen, die insbesondere von historischen Individualbegriffen (*Befreiungskriege*), kulturell-ideologisch aufgeladenen Begriffen (z. B. des kolonialen Erbes) (vgl. Davenport/Cronin, 2000), populären Geschichtsbegriffen (*Zeitalter der ...*) oder etwa auch zeitbedingten Objekt-, Gattungs- und Genrebegriffen (*Belles Lettres*, *Almanach*) ausgehen.

Herausfordernd und zugleich von besonderer Relevanz für die fachliche Praxis sind zudem Theorie- und analytische Forschungsbegriffe, die in langfristige und teils noch laufende Fachdebatten eingebettet sind. Dazu gehören im Thesaurus beispielsweise begriffliche Prägungen und Anzeichnungen einzelner Forschungsrichtungen, etwa die spezifischen Verständnisse von *Okzidentalismus* und *Orientalis-*

mus in der postkolonialen Theorie oder die Einführung des Begriffs des *Eigensinns* in den geschichtswissenschaftlichen Diskurs als historiographisches Konzept seit den späten 1980er Jahren. Ebenso betrifft dies noch junge und räumlich-zeitlich stark gebundene begriffliche Entwürfe wie das Konzept der *Bloodlands* in der osteuropäischen Geschichte, das sogar auf nur eine einflussreiche Einzelarbeit zurückgeht. Selbst im Fall eigentlich traditionsreicher und etablierter Theoriefiguren ist die Unterscheidung disziplinär spezifischer oder auch nur engerer und weiterer Begriffsverständnisse von Bedeutung (*Bricolage*). Ein erheblicher Teil des Begriffskorpus ist damit überhaupt erstmals als (mehrsprachiges) Normdatum verfügbar.

Der Thesaurus wird in mindestens vier Sprachen (Deutsch, Englisch, Französisch, Polnisch) aufgebaut und hält SKOS-konforme Mappings zu den wichtigsten internationalen Fach- und Referenzvokabularen vor. Die Aufnahme und Bearbeitung neuer Begriffsdatensätze erfolgt primär bedarfsorientiert, die Anmeldung neuer Kandidaten ist über ein frei zugängliches Onlineformular möglich. Gleichzeitig werden die bereits bestehenden Begriffsgruppen zielgerichtet kontextorientiert ergänzt, um sowohl den Strukturaufbau wie nachfolgende Such- und Retrievalszenarien zu unterstützen. Entwickelt und kuratiert wird der Thesaurus in der Vokabularverwaltung digiCULT.xTree. Der Korpus von aktuell bereits knapp 3.000 Deskriptoren ist als *Linked Open Data* und hier sowohl im Format JSON-LD und als RDF-XML zugänglich. Spezifische Queries können zudem über eine RESTful-Schnittstelle ausgeführt werden. Ebenso gibt es bereits ein offen zugängliches Anmelde- und Kandidatenverfahren.

Zielgruppen. Der „Historische Thesaurus“ als Service

Der „Historische Thesaurus“ ist sowohl für den Einsatz in historischen Sammlungen als auch in historischen Forschungsumgebungen und Projekten vorgesehen. Historischen Sammlungen soll er ein kontrolliertes Vokabular bieten, das die bisherige Objektzentriertheit konventioneller Kulturerbevokabulare aufbricht und erweitert zugunsten einer stärker kontextorientierten, forschungsnahen und semantisch offeneren Ressource. Als selbst nicht an eine Projektförderung gebundenes Vorhaben soll der Thesaurus künftig zudem einen Hafen und ein Unterstützungs- und Harmonisierungsangebot für insbesondere kleine und mit begrenzter Laufzeit ausgestattete Projekte (etwa Forschungsprojekte, digitale Editionen, Spezialdatenbanken) darstellen, deren hochspezialisierter Begriffsbedarf ansonsten nicht abgedeckt oder in adäquate Begriffssysteme integriert und dauerhaft gepflegt werden könnte.

Die externe Nutzung des Thesaurus ist dabei nicht gebunden an eine einzige, autoritativ vorgegebene Begriffsordnung: Zwar werden sämtliche Begriffsdatensätze in der für Thesauri üblichen, logischen polyhierarchischen Ordnung verwaltet, parallel können die Begriffe jedoch nach Bedarf

zu eigenen Auswahlvokabularen zusammengestellt und mit eigenem Einsprungspunkt zugänglich gemacht werden.

Aktuell ist geplant, den Thesaurus als Service-Angebot in das Dienstportfolio des Konsortiums *NFDI4Memory* einzubringen. Die Posterpräsentation auf der DHd 2025 soll auch vor diesem Hintergrund das Konzept und den Entwicklungsstand des „Historischen Thesaurus“ vorstellen, und an seinem Beispiel zentrale Arbeitsschritte, Herausforderungen und Grenzen des Aufbaus spezifisch geisteswissenschaftlicher Normvokabulare benennen und diskutieren.

Bibliographie

Baca, Murtha. 2003. „Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information“. In *Cataloging & Classification Quarterly* 36,3–4: 47–55. https://doi.org/10.1300/J104v36n03_05 (zugegriffen: 23. Juli 2024).

Davenport, Elisabeth und Blaise Cronin. 2000. „Knowledge Management: Semantic Drift or Conceptual Shift?“. In *Journal of Education for Library and Information Science* 41,4: 294–306. <https://doi.org/10.2307/40324047> (zugegriffen: 23. Juli 2024).

Döring, Karoline, Stefan Haas, Mareike König und Jörg Wettlaufer. 2022. *Digital History. Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft*. Berlin: De Gruyter Oldenbourg.

Gelati, Francesco. 2019. „Implementing an Archival, Multi-Lingual and SemanticWeb-Compliant Taxonomy by Means of SKOS (Simple Knowledge Organization System)“. In *Proceedings of the Workshop on Language Technology for Digital Historical Archives - with a Special Focus on Central-, (South-)Eastern Europe, Middle East and North Africa*: 24–27. https://doi.org/10.26615/978-954-452-059-5_005 (zugegriffen: 23. Juli 2024).

Harpring, Patricia. 2010. „Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works“. Los Angeles: Getty Research Institute.

Jannidis, Fotis, Hubertus Kohle und Malte Rehbein. 2017. „Digital Humanities. Eine Einführung“. Stuttgart: J. B. Metzler.

Roche, Christophe, Luc Damas und Julien Roche. 2014. „Multilingual Thesaurus. The Ontoterminology Approach“. In *CIDOC 2014 (International Committee for Documentation) – Access and Understanding – Networking in the Digital Era, ICOM - International Council of Museums*. <https://hal.science/hal-01356631> (zugegriffen: 23. Juli 2024).

Zapilko, Benjamin, Johann Schaible, Timo Wandhöfer und Peter Mutschke. 2016. „Applying Linked Data Technologies in the Social Sciences“. In *KI - Künstliche Intelligenz* 30: 159–162. <https://doi.org/10.1007/s13218-015-0416-6> (zugegriffen: 23. Juli 2024).