

Kontrollierte Vokabularien, Taxonomien, Thesauri

Zeppezauer-Wachauer, Katharina; katharina.wachauer@plus.ac.at

Kontrollierte Vokabularien, Taxonomien und Thesauri sind Wissensorganisationssysteme und etablierte Werkzeuge in den Digital Humanities. Bei kontrollierten Vokabularien handelt es sich um spezielle Terminologielisten oder Wortsammlungen, die in einem bestimmten Kontext oder für eine bestimmte Zwecksetzung standardisiert wurden. Sie werden beispielsweise in Bibliotheken, Archiven, Museen und wissenschaftlichen Projekten eingesetzt, um eine einheitliche Verwendung von Begriffen sicherzustellen, und spielen eine entscheidende Rolle als stabile Referenz für Ressourcen sowie zur Gewährleistung der Interoperabilität zwischen verschiedenen Projekten und ihren Daten. Mittels kontrollierter Vokabulare können unterschiedlichste Quellen abgefragt, analysiert und zum Zwecke der Kontextualisierung mit weiteren Informationsobjekten verknüpft werden. Sie können verschiedene Formen annehmen, von einfachen Listen von Schlüsselwörtern bis hin zu komplexen Taxonomien oder Thesauri. Standardisierung findet auf verschiedene Arten statt, etwa durch die Verwendung von Normen, Regelwerken oder durch Expertinnen und Experten in dem jeweiligen Fachgebiet.

Ganz simpel erklärt, handelt es sich bei kontrollierten Vokabularien um Listen von Begriffen, auf die sich eine Gemeinschaft geeinigt hat. Zum Beispiel: Die Finger einer Hand heißen im Deutschen “Daumen, Zeigefinger, Mittelfinger, Ringfinger und kleiner Finger”.

Eine Taxonomie wiederum ist ein kontrolliertes Vokabular, das in einer bestimmten Hierarchie organisiert ist. Zum Beispiel lassen sich die Begriffskonzepte “Hand”, “Daumen” und “Zeigefinger” in einer genauen Relation darstellen: “Daumen” und “Zeigefinger” sind Teile von “Hand”. Schließlich ist ein Thesaurus eine Taxonomie mit zusätzlichen Informationen zu jedem Konzept, etwa präferierter oder alternativer Bezeichnungen/Synonyme (“Kleinfinger”) sowie mehrerer Sprachen (“Pinky“ auf Englisch, “Digitus minimus“ auf Lateinisch, “kleinez vingerlîn” auf Mittelhochdeutsch). Zusätzlich kann ein Thesaurus Beziehungen zu verwandten Konzepten enthalten. Beispielsweise haben die Konzepte “Kleiner Finger“ und “Kleiner Zeh“ auch eine Art Beziehung.

All’ diese Begriffe sind eindeutig besetzt, so gibt es keinen Zweifel, welcher Finger einer Hand mit “Daumen” gemeint ist. Diese Eindeutigkeit kann auch bei der Erforschung komplexerer Sachverhalte hilfreich sein, insbesondere bei digitalen Vorhaben. Ein wichtiges Merkmal von kontrollierten Vokabularien ist, dass sie eine präzise und konsistente Terminologie bereitstellen, die es ermöglicht, Informationen effektiver zu suchen, zu finden und zu teilen. Dadurch können Fehler, Missverständnisse und Fehlinterpretationen reduziert werden, was insgesamt zu einer höheren Qualität der Daten und Informationen führt.

Eindeutigkeit bei digitalen Daten ermöglicht stabile Referenzen und gewährleistet digitale Nachhaltigkeit im Sinne der FAIR-Prinzipien.

Kontrollierte Vokabularien, Taxonomien und Thesauri bieten also eine Reihe von Vorteilen für Digitale Editionen:

- **Konsistenz:** Sie helfen bei der Aufrechterhaltung von Konsistenz und Genauigkeit in der Terminologie, die in einer Digitalen Edition verwendet wird. Dies ist besonders wichtig bei der Arbeit mit großen Korpora oder bei der Zusammenarbeit mit anderen Forschenden.
- **Interoperabilität:** Durch die Verwendung von standardisierten Vokabularien lassen sich Digitale Editionen besser mit anderen digitalen Ressourcen verknüpfen. Beispielsweise können Digitale Editionen, die eine gemeinsame Taxonomie verwenden, leichter mit Bibliothekskatalogen oder digitalen Archiven vernetzt werden.
- **Wiederverwendbarkeit:** Durch die Verwendung von standardisierten Vokabularien können Digitale Editionen nachhaltiger wiederverwendet werden. Eine gemeinsame Taxonomie erhöht beispielsweise die Nachnutzung in anderen Projekten als Referenzmaterial oder als Grundlage für weitere Forschung.
- **Suchfunktionen:** Durch die Verwendung von standardisierten Vokabularien können Suchfunktionen effektiver gestaltet werden. Dies ist besonders wichtig, wenn Digitale Editionen in großen Datenbanken gespeichert werden, da eine standardisierte Terminologie die Suche nach bestimmten Begriffen erleichtert.

Auswahl kontrollierter Vokabularien für den Themenkomplex der Digitalen Edition:

- CIDOC-CRM
- Getty Arts and Architecture Thesaurus
- Getty Thesaurus of Geographic Names (TGN)
- Getty Union List of Artist Names (ULAN)
- Cultural Objects Name Authority
- Ontologie historischer, deutschsprachiger Berufs- und Amtsbezeichnungen
- Union List of Artist Names
- GeoNames

- Gemeinsame Normdatei (GND)
- Iconclass
- T-PRO Thesaurus der Provenienzbegriffe
- Virtual International Authority File (VIAF)
- UNESCO Thesaurus
- Wikidata

Literatur:

- Vocabs services Vocabs. URL: <https://vocabs.dariah.eu/>
- Assfalg, Rolf: Metadaten. In: Grundlagen der Informationswissenschaft. Berlin, Boston: 2022, S. B 9, 245-256.
- vocab.org - A URI space for vocabularies. URL: <https://vocab.org/>
- The Shape of Data in Digital Humanities - Modeling Texts and Text-based Resources. Hrsg. von Julia Flanders und Fotis Jannidis. London: 2018, URL: <https://doi.org/10.4324/9781315552941>.
- Getty Vocabularies (Getty Research Institute). URL: <https://www.getty.edu/research/tools/vocabularies/>
- Gindele, Johannes: *<i xmlns:t="http://www.tei-c.org/ns/1.0"> Der Thesaurus der Provenienzbegriffe und seine Praxisregeln als Mittel der Provenienzerschließung von Autorenbibliotheken </i>*. Stuttgart: 2021. URL: https://hdms.bsz-bw.de/frontdoor/deliver/index/docId/6647/file/BA_Gindele_Provenienzbegriffe_2021.pdf.
- Hedden, Heather: Taxonomies and controlled vocabularies best practices for metadata. In: Journal of Digital Asset Management 6: 2010, S. 279-284.
- CIDOC Conceptual Reference Model. URL: <http://www.cidoc-crm.org/html/5.0.4/cidoc-crm.html>
- Kempf, Andreas Oskar: Thesauri. In: Grundlagen der Informationswissenschaft. Berlin, Boston: 2022, S. B 4, 183-194.
- Klee, Carsten: Vokabulare für bibliographische Daten Zwischen Dublin Core und bibliothekarischem Anspruch. In: (Open) Linked Data in Bibliotheken. Berlin, Boston: 2013, S. 45-63.
- Mayr, Philipp: Thesauri, Klassifikationen amp; Co. Die Renaissance der kontrollierten Vokabulare?: 2006, S. 151-170.

- Nasarek, Robert; Moeller, Katrin: Die Ontologie historischer deutschsprachiger Berufs- und Amtsbezeichnungen. Interoperationalität und Berufsklassifizierung durch semantisches Topic Modeling. In: DHd 2018. Kritik der Digitalen Vernunft. Konferenzabstracts DHd 2018. Köln: 2018, S. S. 173-177.
- Nasarek, Robert; Moeller, Katrin: Die Ontologie historischer deutschsprachiger Berufs- und Amtsbezeichnungen. Interoperationalität und Berufsklassifizierung durch semantisches Topic Modeling. In: DHd 2018. Kritik der Digitalen Vernunft. Konferenzabstracts DHd 2018. Köln: 2018, S. S. 173-177.
- , : Guidelines for the construction, format, and management of monolingual controlled vocabulary. Bethesda, Md.: 2010, URL: <https://www.niso.org/publications/ansiniso-z3919-2005-r2010>.
- Reimer, Ulrich: Einführung in die Wissensorganisation. In: Grundlagen der Informationswissenschaft. Berlin, Boston: 2022, S. B 1, 145-158.

Software:

protegé, SKOS Play! Sparna, VocBench, SKOS Simple Knowledge Organization System, Skosmos, TEITOK WissKI (Wissenschaftliche Kommunikationsinfrastruktur - Scientific Communication Infrastructure), Praat, Erlangen CRM / OWL

Verweise:

Digitale Nachhaltigkeit, FAIR-Prinzipien, Geonames, Getty, GND, VIAF, Wikidata, CIDOC-CRM, Ontologie

Projekte:

Getty Vocabularies, T-PRO Thesaurus der Provenienzbegriffe, Vokabularien und Ontologien in der GAMS, Vocabs - ACDH-ÖAW, UNESCO Thesaurus, Iconclass, Union List of Artist Names (ULAN) , Art amp; Architecture Thesaurus® Online , WissKI (Wissenschaftliche Kommunikationsinfrastruktur - Scientific Communication Infrastructure), Erlangen CRM / OWL, Ontologie historischer, deutschsprachiger Berufs- und Amtsbezeichnungen

Themen:

Metadaten, Natural Language Processing

Zitervorschlag:

Zeppezauer-Wachauer, Katharina. 2021. Kontrollierte Vokabularien, Taxonomien, Thesauri. In: KONDE Weißbuch. Hrsg. v. Helmut W. Klug unter Mitarbeit von Selina Galka und Elisabeth Steiner im HRSM Projekt "Kompetenznetzwerk Digitale Edition". URL: <https://gams.uni-graz.at/o:konde.238>