

Die Deutsche Forschungsgemeinschaft hat mit Schreiben vom 12.02.2014 einen Projektantrag der Universitätsbibliothek Trier auf Überführung der Jahrgänge 1897-1944 der Bibliographie der deutschsprachigen Zeitschriftenliteratur in eine frei zugängliche Online-Datenbank bewilligt. Das auf insgesamt fünf Jahre ausgelegte Projekt wird „in vollem Umfang mit hoher Priorität“ gefördert.

## Worum geht es bei Dietrich online?

Mit *Dietrich online* sollen bibliographische Angaben zu ca. 5 Mio im deutschen Sprachraum von 1897-1944 erschienenen Zeitschriftenaufsätzen und Zeitungsartikeln in einer nach Autorennamen, Titelstichwörtern, deutschsprachigen Schlagwörtern, Zeitschriften- resp. Zeitungstiteln und Jahren durchsuchbaren Online-Datenbank nachgewiesen und weltweit verfügbar gemacht werden. Grundlage der digitalen Ausgabe bilden die urheberrechtlich frei gewordenen Bände der von dem Leipziger Verleger Felix Friedrich Dietrich (1874-1938) als Fortsetzungswerk hrsg. *Bibliographie der deutschsprachigen Zeitschriftenliteratur* dieses Zeitraums. Die dort unter Schlagwörtern verkürzt und formalisiert aufgelisteten bibliographischen Angaben werden mit automatisierten Verfahren aus den OCR-erkannten Seiten des Werks ausgelesen und in eine XML-konforme Datenbank überführt. Neben den bibliographischen Datenbankeinträgen werden den Nutzern der Online-Ausgabe die Originalseiten des Werks zur Verfügung gestellt. Titel von Zeitschriften und Zeitungen werden anhand der Zeitschriften-Datenbank (ZDB) ergänzt und normiert. Die Schlagwörter werden nach der Dewey Decimal Classification (DDC) klassifiziert und Personenschlagwörter anhand der Gemeinsamen Normdatei (GND) normiert. Der Nachweis bereits digitalisierter Aufsätze und deren Verlinkung soll über die ZDB-ID-Nummer und Crowdsourcing gewährleistet werden.

## Worin liegt der Nutzen für die Forschung im 21. Jahrhundert?

Zeitschriftenartikel waren im 20. Jahrhundert neben Monographien das wichtigste Medium der wissenschaftlichen Kommunikation. Da diese früher in Bibliothekskatalogen nicht erfasst wurden, bildeten gedruckte Bibliographien die einzigen Nachweisinstrumente für Zeitschrifteninhalte. Für den deutschsprachigen Bereich war bis zum Beginn der Online-Ära ‚der Dietrich‘ das zentrale Instrument, um Zeitschriftenaufsätze zu einem gegebenen Thema oder zu einer bestimmten Person zu finden. Für die zeitgenössische Forschung sind die darin verzeichneten ca. 5 Mio in der ersten Hälfte des 20. Jahrhunderts erschienenen deutschsprachigen Zeitschriften-

tenaufsätze nur in geringer Zahl mit Suchmaschinen wie Google, Yahoo oder Bing zu ermitteln. Die darin enthaltenen Forschungsergebnisse sind aber durchaus noch von (bisweilen herausragender) Bedeutung, und das nicht nur für die historisch arbeitenden Fächer. Mit *Dietrich online* werden diese sonst nur in langwieriger Sucharbeit erschließbaren Daten erstmals in einer heutigen Suchverfahren angemessenen Online-Präsentation erschlossen, die erlaubt, mit wenigen Klicks zu der gewünschten Information zu gelangen.

## Wie wird's gemacht?

Die Projektarbeit gliedert sich in mehrere, teilweise parallel verlaufende Arbeitsprozesse:

- Scannen der ca. 70.000 Seiten umfassenden Bibliographie bei einem externen Dienstleister
- OCR-Erkennung und Nachbearbeitung
- Segmentierung der bibliographischen Einheiten
- Transformierung in XML-konforme bibliographische Angaben nach den Vorgaben der Text Encoding Initiative
- Auflösung von Siglen und Abkürzungen durch Klartext
- Einpflegen der Daten in eine Online-Datenbank

Knifflig ist die Aufgabe wegen der aus Platzgründen verschachtelten und mit vielen Abkürzungen operierenden Struktur der gedruckten Vorlage: In Halbjahresbänden werden die im Berichtszeitraum zu einem bestimmten Thema publizierten Aufsätze unter Schlagwörtern nachgewiesen. Die Titel der Zeitschriften sind mit numerischen oder alphanumerischen Kürzeln (**331a** = *Blätter für Heilerziehung*), die der Zeitungen mit Buchstabenkürzeln aufgeführt (**MNN** = *Münchner Neueste Nachrichten*). Sind unter einem Schlagwort mehrere zu diesem gehörende Artikel aufgelistet, wird dieses ab dem zweiten Artikel nur noch durch ein Tilden-Symbol (~) wiedergegeben (bis 1903 findet auch die Doppeltilde (≈) Verwendung, die für die Pluralform des Schlagworts steht). Verfasserangaben erfolgen in der Regel in runden Klammern, in manchen Bänden in invertierter, in manchen Bänden in nicht invertierter Form, innerhalb eines Bandes jedoch immer identisch. Vornamen sind stets abgekürzt. Daran schließen sich Band- und Seitenangaben



an (das Erscheinungsjahr ergibt sich, abgesehen von Ergänzungsbänden, aus dem auf dem Titelblatt genannten Berichtszeitraum (z. Bsp. **Band LXXX: Alphabetisches, nach Schlagworten sachlich geordnetes Verzeichnis von Aufsätzen, die während der Monate Januar bis Juni 1937 in zumeist wissenschaftlichen Zeitschriften, Sammelwerken, Jahrbüchern, Sitzungsberichten etc. deutscher Sprache erscheinen sind. Mit Nachträgen aus den vorhergehenden Jahren u. alphabetischem Verfasserregister**). Ein Eintrag sieht im Original so aus:

- Oestrin**‘wirkgn. auf d. Corpus luteum-Funktion (A. Westman, D. Jacobsohn) **30cf**. XVII. 1-22 | Vor- kommen d. ~s in d. Hoden u. seine Wirkg. auf d. Knochenwachstum (H.Seemann) **479ee**. XVIII. 225-34
- Oestrogen**: Therapeut. Anwendg. v. ~en Substanzen (E.Novak) **2406**. 160-75
- Oestrus**: Nomenclatur d. ~ erzeug. Verbindgn. (C.A. Rothenheim) **73b**. 205-7
- Oetinger** in Württembg. (G.Nebinger) **323b**. 73./74. H. Beil., 1-8 | Stammliste ~. [Forts.] Beil., 2-37
- Oetker**, Frdr., † (A.Schoetensack) 1603. 310 | Prof.Dr. ~ (Dietz) **1852b**. II. 75
- Oetztal**: Ein Vierteljahrhundert ~er Gletschermessgn. (R.R.v.Srbik) **1694b**. 24. B. 188-91
- Oeynhausen**, Karl v., Frh. **1564**. 79

Textelemente, die für die automatisierte Weiterverarbeitung des OCR-erkannten texts eine Rolle spielen, sind rot hervorgehoben. Der Eintrag verweist auf zehn Zeitschriftenartikel, die im ersten Halbjahr 1937 in unterschiedlichen Zeitschriften erschienen sind, die aus Gründen der Platzersparnis in Form von Siglen zitiert werden: **30cf**, **479ee**, **2406**, **73b**, **323b**, **1603**, **1852b**, **1694b**, **1564**. Die Schlagwörter, unter denen die Artikel rubrifiziert sind (**Oestrin**, **Oestrogen**, **Oestrus**, **Oetker**, **Oetztal**, **Oeynhausen**) sind gefettet wiedergegeben, Verfasserangaben stehen in runden Klammern: (). Der senkrechte Strich (Pipe) | zeigt den Übergang zu einem weiteren Artikel zum gleichen Schlagwort an, das von da an nur noch durch eine Tilde ~ wiedergegeben wird.

Die Umsetzung dieser Angaben mit informatorischen Mitteln führt zu einer Liste bibliographischer Angaben, die im Idealfall so aussieht:

Westman, A. ; Jacobsohn, D. : Oestrin‘wirkgn. auf d. Corpus luteum-Funktion. In: Acta Obstetrica et Gynecologica Scandinavica 17 (1937), S. 1-22.

Seemann, H. : Vorkommen d. Oestrins in d. Hoden u. seine Wirkg. auf d. Knochenwachstum. In: Endokrinologie : Zentral-

blatt für d. Gebiet d. inneren Sekretion u. Konstitutionsforschung 18 (1937), S. 225-234.

Novak, E. : Therapeut. Anwendg. v. Oestrogenen Substanzen. In: Die Drüsen mit innerer Sekretion : ihre physiologische und therapeutische Bedeutung. Wien [u.a.] : Aesculap-Verlag, 1937, S. 160-75.

Rothenheim, C. A. : Nomenclatur d. Oestruserzeug. Verbindgn. In: Schweizer Apothekerzeitung [73b : ZDB: Jg. 00] 75 (1937), S. 205-7.

Nebinger, G. : Oetinger in Württembg. In: Blätter für württembergische Familienkunde 73./74. Heft (1937), Beil., 1-8.

ohne Autor: Stammliste Oetinger. In: Blätter für württembergische Familienkunde 73./74. Heft (1937), Beil., S. 2-37.

Schoetensack, A. : Oetker, Frdr., †. In: Zeitschrift der Akademie für Deutsches Recht 5 (1937), S. 310 ff.

Dietz : Prof. Dr. Oetker, Frdr., †. In: Zeitschrift für Wehrrecht 2 (1937/1938), S. 75 ff.

Srbik, R. R. von: Ein Vierteljahrhundert Oetztaler Gletschermessgn. In: Zeitschrift für Gletscherkunde, für Eiszeitforschung und Geschichte des Klimas 24 (1937), S. 188-91.

ohne Autor: Oeynhausen, Karl v., Frh. In: Das Werk : Monatsschr. der „Vereinigte Stahlwerke Aktiengesellschaft“, 17 (1937), S. 79 ff

Diese Daten dienen dem Aufbau einer Datenbank, in der die entsprechenden Angaben mit weiteren, der linguistischen Optimierung der Suchoptionen dienenden Informationen angereichert werden.

## Wann kann man es benutzen?

Ab 2015 soll eine Beta-Version des Suchmoduls, das natürlich zunächst nur einen geringen Teil des Datenbestandes beinhalten wird, online unter der Adresse <http://dietrich.uni-trier.de/> zur Verfügung gestellt werden. Derzeit sind Mitarbeiterinnen und Mitarbeiter der Bibliothek noch mit vorbereitenden Aufgaben (OCR-Tests, Optimierung der automatisierten Segmentierung bibliographischer Einheiten u.a.m.) befasst. Die Ausschreibungen für die von der DFG bewilligten Stellen für zwei Informatiker(innen) und eine Diplom-Bibliothekarin resp. einen Diplom-Bibliothekar sind angelaufen. Da die Gesamtprojektdauer auf 60 Monate veranschlagt ist, wird das vollständige Material ab 2019/20 online verfügbar sein.