

KOMMISSION FÜR ZENTRAL- UND OSTASIATISCHE STUDIEN

Neues Datenbanksystem für das Wörterbuch der tibetischen Schriftsprache

DAS ZETTELARCHIV MIT ÜBER 200.000 EINTRÄGEN WIRD KÜNFTIG AUSSCHLIESSLICH IN EINER DATENBANK WEITERBEARBEITET.

VON PETRA MAURER UND JOHANNES SCHNEIDER

Die Kommission für zentral- und ostasiatische Studien sammelt seit 1954 Materialien für ein Wörterbuch der tibetischen Schriftsprache: Nachweise aus allen Bereichen der tibetischen Literatur sind alphabetisch unter den jeweils relevanten Wörtern abgelegt. Die Sammlung umfasst über 200.000 Zettel, auf deren Grundlage ein neuartiges Belegstellenwörterbuch erarbeitet wird. Seit 2005 sind drei Faszikel erschienen, der Abschluss des Projekts ist für das Jahr 2020 ins Auge gefasst.

Bereits Ende der 1990er Jahre war ein erster Versuch unternommen worden, die Bearbeitung mit Hilfe der elektronischen Datenverarbeitung zu erleichtern. Der Hamburger Iranist Ronald E. Emmerick hatte damals die ersten Elemente einer Datenbank (BOD) bereitgestellt, doch erwies es sich nach seinem Tod im Jahre 2001 als unmöglich, diese Bausteine weiterzuentwickeln. Das Wörterbuch wird daher bis heute mit Hilfe eines Desktop-Publishing-Programms geschrieben. Da jedoch die Arbeit mit Verweisungen und Verschiebungen sehr zeitraubend und fehleranfällig ist, wurde immer deutlicher, wie notwendig eine Datenbank für die Eingabe und Verwaltung der Belegstellen sowie für die automatische Ausgabe der

Datensätze im Format der einzelnen Artikel wäre. Dieses Problem haben 2006 auch die Gutachter bei der letzten Evaluation im Rahmen des Akademienprogramms aufgegriffen und dringend empfohlen, die Arbeiten auf Basis eines zeitgemäßen Datenbanksystems fortzuführen, damit die Sammlung kontinuierlich erweitert und letztlich auch online verfügbar gemacht werden kann.

Finanzierung gesichert

In diesem Jahr ermöglichten nun die großzügige Spende einer Privatperson, die nicht namentlich genannt werden will, sowie eine Sonderzuweisung von Sachmitteln die Umsetzung dieser Forderung. Helmut Gassner von der Züricher Rabten-Stiftung, ein erfahrener Programmierer, kann dabei auf Module zurückgreifen, die er für sein bewährtes Clearlook-Textverarbeitungs- und Datenbank-System entwickelt hat: Spezielle Sortieralgorithmen ermöglichen die Anordnung der Einträge nach dem komplexen tibetischen Alphabet; auch andere asiatische Schriftzeichen können ohne großen Aufwand eingebunden werden.

Bisher aufwändige Prozeduren wie die Verwaltung der Daten, die Handhabung von Verweisungen, die Durchführung von Sortierungen und die Einstellung von Schriftattributen werden nunmehr wesentlich vereinfacht. Insbe-

sondere können dieselben Sätze unter verschiedenen Stichworten in einem Arbeitsgang abgelegt werden. Das System besteht aus drei interaktiven Datenbanken, in denen die vollständigen Quellbelege, deren Kurzfassungen für einzelne Einträge sowie die Stichwörter selbst gespeichert sind. Ein spezieller Ausgabemechanismus bündelt die zu einem Lemma gehörenden Belegstellen nach einem chronologischen Schlüssel, ordnet die Artikel nach dem tibetischen Alphabet an und generiert daraus das Wörterbuch mit den nötigen Formatierungen. Satzspiegel und Schrifttypen der bisher veröffentlichten Faszikel werden beibehalten. Die Datensätze lagern auf dem Server des LRZ, so dass mehrere Benutzer gleichzeitig daran arbeiten können; das lästige und fehleranfällige Hin- und Herkopieren von Textblöcken zwischen verschiedenen Dateien gehört damit der Vergangenheit an.

Die grundlegenden Elemente des Datenbank-Systems werden bis Jahresende fertig gestellt sein; die Übernahme der alten BOD-Datensätze, die Einbindung der bisher verwendeten Schriften und die Anpassung an das jetzige Layout sollen in der ersten Hälfte des Jahres 2008 erfolgen. Die Kommission geht davon aus, dass die Arbeiten dann ausschließlich auf Grundlage des neuen EDV-Systems fortgeführt werden.

Die Autoren sind wissenschaftliche Mitarbeiter der Kommission für zentral- und ostasiatische Studien der Bayerischen Akademie der Wissenschaften.

