



Babylonische Keilschrifttafel,
dokumentiert in der Datenbank
„electronic Babylonian Literature“
(eBL) der LMU München,
einem zentralen Tool für das neue
Akademievorhaben.

Das ALTE RTUM mit dem Cyberspace verbinden

Von der Tontafel zum Datensatz: Das CAIC-Vorhaben nutzt die **neuesten Technologien** für die wissenschaftliche Arbeit. Ziel ist eine universelle, modulare Plattform zur Katalogisierung, Dokumentation und Edition von Keilschrifttexten, die auch bereits vorhandene Korpora zusammenführt.

Von **Enrique Jiménez** und **Fabian Simonjetz**

Das CAIC-Vorhaben verbindet das Altertum nahtlos mit dem Cyberspace. Es nutzt für die Bearbeitung einiger der ältesten Textzeugnisse der Menschheit nämlich neueste Technologien und arbeitet dabei mit führenden digitalen Keilschriftinitiativen in aller Welt zusammen. Gemeinsam haben wir ein ehrgeiziges Ziel: eine universelle, modulare Plattform zur Katalogisierung, Dokumentation und Edition von Keilschrifttexten zu etablieren, die die vorhandenen Tools und Korpora zusammenführt und um wesentliche neue Funktionen ergänzt.

Bestehendes verbinden und verbessern

Die Welt der digitalen Keilschriftstudien ist bisher ein Archipel von vielen Inseln

und Inselchen, die die CAIC-Plattform zu einer Landmasse zusammenführen soll. Unser Ziel ist es, einerseits mit allen bestehenden internationalen Großprojekten der digitalen Altorientalistik wie dem französischen Archibab, dem spanischen BDTNS und den internationalen Initiativen CDLI und Oracc zusammenzuarbeiten und andererseits zur Interoperabilität all dieser insularen Projekte beizutragen. Ganz wesentlich ist hierfür das Leibniz-Rechenzentrum (LRZ) der Bayerischen Akademie der Wissenschaften, das auf seiner enorm leistungsstarken Server-Infrastruktur unsere Plattform und alle Forschungsdaten beherbergt. Forscher weltweit werden die Möglichkeit erhalten, Unterprojekte in CAIC zu erstellen und von allen Funktionen des Systems zu profitieren. So wollen wir nicht nur



die Vernetzung innerhalb der Altorientalistik, sondern auch die interdisziplinäre Zusammenarbeit voranbringen.

Eine paläographische Revolution

Alle im Irak-Museum aufgenommenen High Dynamic Range-Fotos (HDR) und Reflectance Transformation Imaging-Dateien (RTI) stehen auf der Projektwebsite zum Download zur Verfügung. Ein bahnbrechendes Werkzeug wurde im Rahmen des von der Alexander von Humboldt-Stiftung finanzierten und von Enrique Jiménez geleiteten Projekts „electronic Babylonian Literature“ (eBL, ebl.lmu.de) entwickelt, das die direkte Verknüpfung der Fotos mit den Umschriften der dargestellten Tontafeln ermöglicht, und zwar Zeichen für Zeichen.

Dieses Tool ist zentral für die CAIC-Plattform. Neu ist nun, dass alle generierten Daten direkt in die dynamisch wachsende Zeichenliste des Projekts einfließen, die so die Formen jedes Zeichens quer durch die Jahrtausende der Verwendung der Keilschrift dokumentiert. Dass Tontafeln anhand der Zeichenform datiert werden können („Paläographie“), ist seit Langem bekannt, aber bisher fehlte ein umfassendes Repertoire dieser Formen, was die Forschung stark behinderte. CAIC wird das ändern. Bei der Abfrage der Zeichenformen erhält man immer auch direkten Zugang zu sämtlichen Kontextinformationen der Tafeln im Irak-Museum. Ohne Übertreibung kann man deshalb sagen, dass dies die Keilschriftforschung revolutionieren wird.

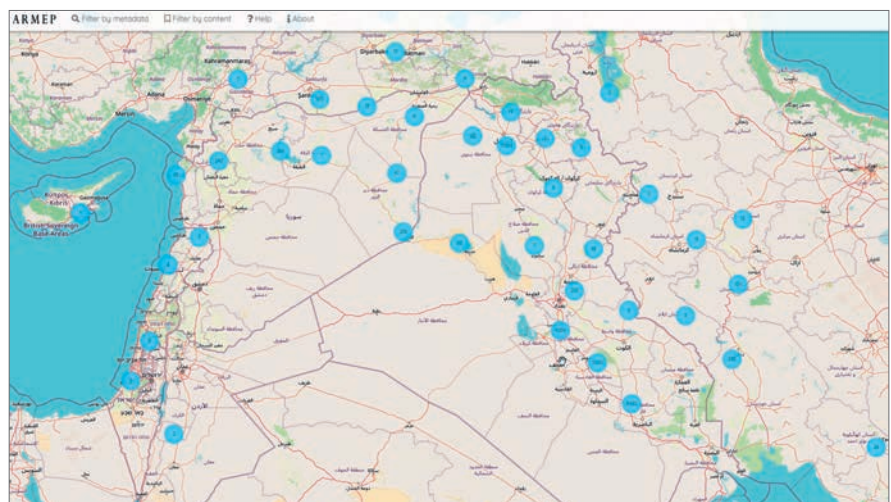
Keil für Keil: Wo beginnen, wo enden die Zeichen, und wie liest man sie? Hier: DINGIR-MEŠ ša₂ AN-e, was „die Götter im Himmel“ bedeutet.

Eines Tages wird die KI auch ganze Tontafeln lesen können.

Die Daten werden außerdem verwendet, um Künstliche Intelligenz darin zu trainieren, Zeichen auf Fotos zu erkennen. Eines Tages wird die KI auch ganze Tontafeln lesen können. Da es weltweit viel zu wenige Menschen gibt, die die Keilschrift entziffern können, wird das niemandem die Arbeit wegnehmen, sondern vielmehr ein ernstes strukturelles Problem lösen, das dazu geführt hat, dass viel zu viel Material über Jahrzehnte unbearbeitet geblieben ist.

Komplexe Daten, strukturiertes Format

Das CAIC-Team fertigt anhand verbindlicher Transliterationsregeln Umschriften der Keilschrifttexte an und erweitert sie um ergänzende Informationen, denn für



Übersicht über die Fundorte der mesopotamischen Keilschrifttafeln mit genauer Anzahl für jede Stadt.

die Dokumentation reicht es nicht aus, nur den Text selbst zu speichern. Weil die dreidimensionalen Tafeln auf mehreren Seiten beschriftet sind, sind vielmehr auch Angaben zu Layout und Zeichenverteilung wichtig. Wir kodieren ferner Daten wie Fundort, Epoche, Sprache, außerdem linguistische Annotationen und Verweise auf zugehörige Bruchstücke. Natürlich übersetzt das CAIC-Team die mühsam transliterierten Texte vollständig und kommentiert sie philologisch und historisch.

All dies wird in einem komplexen Datenformat gespeichert. Denn Daten sind nur dann sinnvoll für die digitale Forschung nutzbar, wenn sie in einer Form gespeichert sind, die es erlaubt, sie zu durchsuchen, zu sortieren und zu filtern. Wenn die Daten in die Datenbank überführt werden, erhält jedes Stück eine eindeutige ID und einen Eintrag, der die Bearbeitung, linguistische Informationen (z. B. Morphologie), Meta-Informationen (z. B. Fundort, Datierung, archivalischer Kontext), Fotos und statistische Daten enthält. Dafür wird Java Script Object Notation (JSON) verwendet, ein strukturiertes Datenformat, das mit Schlüssel-Wert-Paaren arbeitet. Die Schlüssel sind festgelegte Bezeichnungen wie „Periode“ oder „Transliteration“, und die Werte sind Zahlen für Datumsangaben, zeitliche Kategorien wie „altbabylonisch“ oder auch komplexere Einträge wie bibliographische Referenzen. Die Herausforderung besteht darin, Schlüssel und Werte so festzulegen, dass möglichst alle Eigenschaften darstellbar sind, und das abstrakte Datenmodell muss stetig angepasst und erweitert werden. Ein Vorteil von JSON ist die leichte Konvertierung in andere Formate, wie etwa TEI P5 XML. So kann jeder Nutzer das Material für eigene Zwecke nutzen. Backups der Datenbank werden mehrmals täglich auf dem LRZ-Server gespeichert, sodass Datenverlust vorgebeugt ist.

Lösungen für ein stetig wachsendes Textkorpus

Das Korpus der Keilschriftsprachen (in der Hauptsache Sumerisch und Akkadisch) wächst jedes Jahr um viele tausend Wörter an. Wir müssen Werkzeuge schaffen, die diesem Wachstum Rechnung tragen.

Die Nutzer werden in der Lage sein, virtuell in jeden irakischen Fundort zu reisen.

Schon jetzt ermöglicht die CAIC-Plattform für akkadische Texte eine Lemmatisierung, bei der jedes Wort eines Textes mit seiner Wörterbuchform verknüpft wird. Die so mögliche Erstellung einer dynamischen Konkordanz ist für ein noch immer rasch wachsendes Sprachkorpus am besten geeignet. Ein zentrales Ziel der ersten Projektphase ist die Erarbeitung der Lemmatisierung für sumerische Texte, die besonders Walther Sallaberger momentan beschäftigt; er kann hierfür auf seine langjährige Erfahrung mit einem sumerischen Glossar aufbauen.

Lemmatisierung erlaubt die automatische Generierung von Glossaren, diese können etwa mit Material zur Etymologie oder auch thematischen Sammlungen verlinkt werden. Besonders wichtig wird uns die Verknüpfung zu den Online-Einträgen des direkten Vorgängerprojekts an der BADW sein, dem Reallexikon der Assyriologie und Vorderasiatischen Archäologie (rla.badw.de).

Von der Landkarte zur Tontafel zur Edition

Unsere Plattform wird für jede Tontafel nicht nur die von CAIC generierten Fotografien und Editionen bieten, sondern sie

auch in ihrem Fundkontext visualisieren. Das trägt der Vielfalt an gut dokumentierten Provenienzen Rechnung, die die Sammlungen des Irak-Museums auszeichnet. Dafür bauen wir auf das von Karen Radner und Jamie Novotny mit der IT-Gruppe Geisteswissenschaften an der LMU München entwickelte Visualisierungstool (www.armep.gwi.uni-muenchen.de) auf, das es ermöglicht, Keilschrifttexte von einer Kartenoberfläche aus abzufragen. Die Nutzer werden in der Lage sein, virtuell in jeden irakischen Fundort zu reisen und die dort gefundenen Dokumente in ihrer Zeitstellung und ihrem Archivzusammenhang zu konsultieren. So wird man in der Stadt Uruk sehen können, in welchen Bibliotheken Manuskripte der Abenteuer ihres legendären Königs Gilgamesch aufbewahrt wurden. Ein Mausklick führt dann zur Übersetzung, und wer will, kann sofort auch auf die gesamte Edition zugreifen.

Tim Berners-Lee, der Erfinder des World Wide Web, hat gesagt: „Data is a precious thing and will last longer than the systems themselves.“ Das trifft auch auf unsere Plattform zu, die irgendwann veraltet sein wird. Aber dank ihrer überlegten und offenen Datenstruktur werden die Editionen und die sonstigen Informationen, die wir mit den beschriebenen Werkzeugen generieren, die Wissenschaft und die allgemeine Kenntnis der reichen Textschätze des alten Iraks dauerhaft prägen. Denn CAIC wendet sich mit seiner Plattform und seinen Daten bewusst nicht nur an Fachleute, sondern auch an die breite Öffentlichkeit.

Prof. Dr. Enrique Jiménez

ist Professor für altorientalische Literaturen an der LMU München und Co-Leiter des Projekts „Cuneiform Artefacts of Iraq in Context“ (CAIC) an der Bayerischen Akademie der Wissenschaften.

Fabian Simonjetz M. A.

ist Computerlinguist und im CAIC-Vorhaben zuständig für die digitale Aufbereitung der Keilschriftdaten.
