Wadoku Projekt

Überblick über das WaDoku-Projekt und statistische Daten dazu

Das WaDoku-Wörterbuch ist das mit großem Abstand umfassendste und aktuellste japanisch-deutsche Wörterbuch. Es enthält zurzeit (Mai 2008) über 110.000 Stichwörter und insgesamt knapp 260.000 Datensätze.

Ein Großteil der Daten sind also Ableitungen, Komposita, Verwendungsbeispiele, Beispielsätze, Sprichwörter und so weiter. Solche Datensätze sind per Datenbankrelation fest bestimmten Stichwörtern zugeordnet.

Eine besondere Schwierigkeit des Japanischen für Wörterbücher ist die große orthografische Varianz - also die Tatsache, dass dasselbe „Wort" auf bis zu zehn oder mehr unterschiedliche Weise geschrieben werden kann, die alle korrekt sind. Nur ein Teil dieser Varianten lässt sich durch klare Regeln erfassen. Das Projekt legt daher großen Wert darauf, die übliche Varianten in die Datei aufzunehmen, weil sonst Abfragen zu keinem Ergebnis führen, obwohl der gesuchte Eintrag eigentlich da ist bzw. Abgleich mit anderen Wörterbüchern erschwert wird. Für die 260.000 Datensätze gibt es knapp 350.000 unterschiedliche Schreibungen.

Es werden etwa 300.000 unterschiedliche Bedeutungen der japanischen Einträge erläutert und über 460.000 Übersetzungsvorschläge dafür angeboten.

Viele Online-Wörterbücher stellen einer Schreibung der Quellsprache genau eine Übersetzung der Zielsprache gegenüber. Wenn man einen solchen Ansatz auf die Daten des WaDoku-Wörterbuches übertragen würde, käme man auf über 730.000 solcher Paare.

Bei mehr als 60.000 Datensätzen ist das Fachgebiet der Einträge angegeben. Knapp 20.000 Datensätze verweisen auf andere Datensätze, z.B. weil diese Synonyme oder Antonyme sind, eine andere Aussprache der angegebenen Schreibung aufführen oder eine andere japanische Transkription eines Fremdwortes angeben. Je nach Datenformat sind diese Verweise als Sprungverweis anzuklicken.

Etwa 200 Datensätze verweisen auf Bilddaten. Unsere Bilddaten liegen im SVG-Format (Scalable Vector Graphics) vor und sind bei entsprechender Ausgabe ohne Qualitätsverlust beliebig vergrößerbar.

Für alle Datensätze gibt es Umschrift der Aussprache in japanischer Silbenschrift und in lateinischer Schrift. Lateinische Transkription ist in verschiedenen Stilen möglich. Die Kana-Umschrift ließe sich relativ leicht auch als Furigana (Ruby) darstellen.

Die Wörterbuch-Daten sind in verschiedenen Formaten und Versionen zugänglich - online z.B. unter <http://wadoku.de/>, als Download in verschiedenen Formaten für die Verwendung in bestimmten Wörterbuch-Programmen oder als Datenbank-Laufzeitprogramm. Es gibt vielfältige Kooperationen mit anderen Projekten. Die WaDoku-Daten werden etwa in JMdict <http://www.csse.monash.edu.au/~jwb/j_jmdict.html>, Papillon-Projekt <http://www.papillon-dictionary.org/>, Reading-Tutor <http://language.tiu.ac.jp/index_g.html>, w3dict <http://wadokujt.w3dict.com/> und POPjisyo <http://www.popjisyo.com/WebHint/Portal_e.aspx> verwendet

Da die Daten elektronisch vorliegen, ist es möglich, auch die deutsche Seite zu durchsuchen. Dadurch wird das Wörterbuch allerdings noch kein richtiges deutsch-japanisches Wörterbuch. Es war bislang zum Beispiel nicht die Absicht, den gesamten deutschen Wortschatz abzudecken. Japanische Nutzer werden weiter Informationen zur Grammatik, zur Aussprache oder zu Verwendungsweise deutscher Wörter vermissen.

Wir arbeiten daran, auf der deutschen Seite z.B. Ergänzungen vorzunehmen, die besonders für japanische Muttersprachler interessant sind, und wir versuchen, eine bessere deutsch Suche zu erreichen.

Beginn des Wörterbuch-Projektes im Frühjahr 1998 und Motivation dazu

Bei der Recherche zu seiner Dissertation über japanische Zukunftsforschung und Zukunftsplanung stand Ulrich Apel vor dem Problem, dass die existierenden japanisch-deutschen Wörterbücher die damit zusammenhängenden Themengebiete wie aktuelle Soziologie oder Zukunftstechnologien nicht abdeckten.

Das nach wie vor neueste und mit über 50.000 Stichwörtern halbwegs umfassende gedruckte japanisch-deutsche Sprachwörterbuch - das "Wörterbuch der deutschen und japanischen Sprache - Japanisch-Deutsch" von Robert Schinzinger et al., verlegt bei Sanshusha - ist bereits 1980 erschienen. Das bedeutet, dass Computertechnologie und andere moderne Technikfelder weitgehend fehlen - nicht unbedingt die besten Voraussetzungen für die Forschung über Japans Zukunft.

Die Originalfassung von Kinji Kimuras "Großes Japanisch-Deutsches Wörterbuch", erschien 1937. Das Buch wird seither ohne inhaltliche Veränderung gedruckt. Auch wenn es mit knapp 60.000 Lemmata etwas umfassender als das Schinzinger-Wörterbuch ist, ist es für Forschung über ein aktuelles Thema keine Alternative.

Bei japanisch-englischen Wörterbüchern ist die Situation nicht wesentlich besser. Einige mittelgroße Wörterbücher werden regelmäßig aktualisiert. Aber ein großes Wörterbücher wie das Shinwaei Daijiten von Kenkyusha erfuhr auch erst 2004 nach 30 Jahren eine komplette Aktualisierung. Zu diesem Zeitpunkt war das oben erwähnte Dissertationsprojekt schon längst erfolgreich abgeschlossen.

Ulrich Apel begann also zunächst ein Glossar für Terminologie der Zukunftsforschung anzulegen. Weil eine Textdatei nach sehr kurzer Zeit nicht mehr sonderlich befriedigend war, war der Umstieg auf das Datenbank-Programm FileMaker einer der wichtigsten Schritte zum heutigen Wörterbuch. Die legendär einfache und trotzdem mächtige Programmierung von Datenbanken mit dieser Software erleichterte Arbeit an den Daten und ermöglichte eine sehr einfache Veröffentlichung der Daten zum Download und als Online-Version.

1999 konnten Nutzer erstmals Laufzeit-Datenbanken von einem Server der Universität Osaka herunterladen, und es gab wenig später eine erste Online-Version der Datenbank. Im selben Jahr hat Ulrich Apel das Projekt auch auf dem deutschsprachigen Japanologentag vorgestellt.

Steigender Umfang und Popularität von Wadoku seit 2000

Im Jahre 2000 richtete Thomas Latka die Webadresse "www.wadoku.de" ein, die zunächst auf den von Ulrich Apel genutzten Server in Osaka verwies. Im Jahre 2001 wurde daraus eine eigene Wörterbuch-Seite mit SQL-Datenbank. Die Suche und Indizierung wurde optimiert, sowie die Möglichkeit geschaffen, bestehende Wörterbucheinträge zu kommentieren oder auch neue Einträge vorzunehmen. In den darauffolgenden Jahren gewann die Seite immer mehr an Popularität und hatte täglich mehrere tausende Suchanfragen sowie immer mehr Vorschläge für Neueinträge erhalten, die von einem Editorenteam überprüft werden. Im Jahre 2006 kam ein Forum und 2007 auch ein Wiki dazu, das die immer größer werdende Community dazu nutzt, um sich auszutauschen und die gemeinsame Arbeit zu koordinieren.

Bei den Japanologentagen 2002 und 2006 wurden neue Ergebnisse der Arbeit am Projekt vorgestellt.

Seit 2003 ist die Umschrift mit Markierungen versehen, um korrekte Lateinumschrift nach dem Hepburn-System zu ermöglichen. Seit 2005 konnten wir für einen großen Teil der Einträge den japanischen Tonakzent markieren. 2006 und 2007 wurde die gesamte Datei strukturell überarbeitet um bessere Suche und bessere Darstellung zu ermöglichen.

Vorbilder

Natürlich gibt es für das Projekt auch Vorbilder, insbesondere das von Prof. em. Jim Breen, Monash University, 1991 initiierte edict-Projekt, ein sehr umfassendes freies japanisch-englisches Wörterbuch. Im gleichen Daten-Format und unter dem Namen "jddict" gab es auch eine japanisch-deutsche Datei, basierend auf Wolfgang Hadamitzkys "Kanji und Kana", erschienen bei Langenscheidt. Unter dem Namen "jddict" findet man heute mitunter eine Untermenge des WaDoku-Wörterbuches, nämlich die Einträge, die sowohl in edict bzw. der multilingualen Version JMdict und WaDokuJT vorhanden sind (JMdict umfasst jedoch wesentlich weniger Datensätze als WaDokuJT). Weiter sind die Daten von Wolfgang Hadamitzkys Zeichenwörterbuch "Langenscheidt Großwörterbuch Japanisch-Deutsch" 1993 auch in einer elektronischen Version namens "MacSUNRISE KanjiBank" fürs Macintosh-Betriebssystem erschienen.