

## Zur Herstellung und Benutzung rückläufiger Wörterbücher als PC-Textdateien

SHIROOKA Keiji (jjksiro@hss.shizuoka.ac.jp)

Seit einigen Jahren gibt es Data Discs, die mit Data Discman von Sony benutzt werden. Data Discs sind nicht nur in Japan, sondern auch im Ausland (einschließlich in Deutschland) erschienen. Einige Beispiele für solche Data Discs sind das Duden Universalwörterbuch, Mackensen Deutsche Rechtschreibung, Langenscheidts zweisprachige Wörterbücher, und in Japan ist das Wörterbuch Deutsch-Japanisch Crown auf den Markt gekommen.

Mit der freien Software DDwin, die von KUSAMOTO Kazuma geschrieben wurde und z.B. im Netzdienst NIFTY-Serve jedem frei zur Verfügung steht, kann man aus Data Discs Daten als Textdatei auf einen PC mit Windows 3.1 oder 95 holen. DDwin ermöglicht es, alle Stichwörter aus einem elektronischen Wörterbuch auf eine Textdatei aufzunehmen. Mit diesem Werkzeug ist es sogar möglich, nicht nur die Stichwörter, sondern auch die vollständigen Einträge eines Wörterbuchs zu dekodieren und als Textdatei zu extrahieren. Allerdings können z.B. phonetische Zeichen nicht als Textdatei ausgedrückt werden.

Ich habe in der vorliegenden Arbeit gezeigt, wie man mit DDwin die gewünschten Daten dekodiert und extrahiert, die Daten bearbeitet und daraus ein rückläufiges Wörterbuch zusammenstellen kann. Bei der Beschreibung der Arbeitsgänge habe ich möglichst konkret arbeitspraktische Probleme beschrieben, damit jeder, der daran interessiert ist, diese neue Möglichkeit nutzen kann. Die Kenntnis von Programmiersprachen wie BASIC ist nicht mehr nötig. Zur Herstellung und Benutzung der maschinenlesbaren rückläufigen Wörterbücher braucht man nur sogenannte UNIX-like Tools wie REV, SORT, UNIQ, CUT, FOLD, PASTE, NL, WC, WORD, GREP, TR, SED, AWK :