Datenfelder

Beschreibung der Felder:

Feld "DaJapanisch"

Enthält die japanischen Wörterbuch-Einträge in ihren "üblichen" Schreibungen. Oftmals sind das zehn oder mehr Schreibungen. (Man könnte wahrscheinlich Schreibungen automatisch generieren. Dann bekäme man aber auch Schreibungen die ungebräuchlich und/oder weniger eindeutig sind. Unser arbeitsaufwändigerer Ansatz scheint daher die
vernünftigere Lösung).

Dieses Feld ist vor allem für Suchindex und Datenabgleich angelegt und nicht für die Anzeige. Für die Anzeige ist das Feld [DaMidashigo] optimiert, auch wenn es naturgemäß Überschneidungen gibt.

- Unterschiedliche Schreibungen sind durch Semikolon getrennt.
- Schreibungen, die Nicht-Toyo-Kanji enthalten oder die Nicht-Standard-Lesungen von Toyo-Kanji enthalten, stehen in Klammern.
- Identische Schreibungen (Homographen) sind - normalerweise - durch ein Indexsystem aus Nummern und Buchstaben in eckigen Klammern markiert (wenn Index fehlen sollte, ist das ein Fehler).
    - Homographen mit unterschiedlichen Bedeutungen sind durch Nummern markiert.
    - Homographen mit gleicher Bedeutung aber unterschiedlicher Lesung sind mit Buchstaben markiert.
- Jede Schreibung sollte durch die Indexmarkierung eineindeutig sein. (Dies müsste evtl. mal wieder überarbeitet werden). Das wäre das Feld, in dem nach japanischen Wörtern gesucht werden soll.

In einem nächsten Schritt könnte man jede Schreibung genauer beschreiben. Wichtig wäre das insbesondere bei häufigen falschen Schreibungen, die wir auch aufnehmen. Kana-Schreibungen wären leicht zu beschreiben. Ebenso wahrscheinlich Schreibungen mit Nicht-Toyo-Kanji. Nicht Standard-Schreibungen sind schwieriger. Für Varianten bei Okurigana wird es noch schwieriger. Wahrscheinlich könnte man beschreiben, was Kanji ist, was Okurigana, was ersetztes Kanji, was Kanji-Variante ist. Fehlen von best. Schreibungen z.B. Kana-Schreibung von wissenschaftl. Tier- und Pflanzennamen könnte man
leichter überprüfen; bestimmte Schreibungsvarianten ließen sich leichter automatisch errechnen.

Evtl. könnte man Indexierung der Homographen von diesem Feld abtrennen.

Feld "DaLesung"

Enthält die japanischen Lesungen in Hiragana.
Homophone sind durch Indexnummern in eckigen Klammern unterschieden.
Hiragana "" ist als zwei Zeichen geschrieben (Shift JIS enthält kein Hiragana)

Man könnte das Feld für eine Dopplung des Feldes
DaRomajiBearbLesung halten; wenn man Markup aus DaRomajiBearbLesung
entfernt sind die Felder gleich. Das Feld lässt sich so jedoch
leichter pflegen; Redundanz sollte auch verhindern, dass sich Fehler
einschleichen. Das wäre ein Kandidat für Daten, die man nur intern
verwaltet.

Feld "DaDeutsch"

Enthält die Übersetzungen des Eintrages DaJapanisch.
Einzelne Übersetzungen sind durch Semikolon unterschieden.
Bei größeren inhaltlichen Unterschieden der Übersetzungen steht doppelter Schrägstrich "//".
Wenn die Übersetzungen sehr unterschiedlich sind,
werden die Übersetzungen mit Indexnummern in eckigen Klammern gefolgt
von einem schmalen Nicht-umbrechenden-Leerzeichen unterschieden.

Bei unterschiedlicher grammatischer Verwendung werden Großbuchstaben
zur Unterscheidung verwendet.

Bei ähnlichen abgeleiteten Bedeutungen können weiter Kleinbuchstaben
stehen.

Sachgebiete stehen in geschweiften Klammern und sind mit "Dom."
markiert.

Deutsche Genera sind mit kurzen Abkürzungen und Tag "Gen" markiert:

- *f* = feminin
- *fm* = femin oder maskulin
- *fmNAr* = feminin oder maskulin ohne Artikel verwendet
- *fn* = feminin oder neutrum
- *fNAr* = feminin ohne Artikel verwendet
- *fpl* = feminin plural
- *mfpl* = maskulin oder feminin plural
- *m* = maskulin
- *mf* = maskulin oder feminin
- *mfn* = maskulin, feminin oder neutral
- *mn* = maskulin oder neutral
- *mpl* = maskulin plural

- *mNAr* = maskulin ohne Artikel verwendet
- *mnpl* = maskulin oder neutral plural
- *n* = neutral
- *nf* = neutral oder feminin
- *nm* = neutral oder maskulin
- *npl* = neutral plural
- *NAr* = keine Artikel verwendet
- *NArN* = kein Artikel notwendig
- *nmplNAr* = neutral oder maskulin, kein Artikel verwendet
- *nNAr* = neutral kein Artikel verwendet
- *plNAr* = plural kein Artikel verwendet
- *pl* = plural, Geschlecht nicht angebbar
- *fNAr* = feminin kein Artikel verwendet
- *fmpl* = feminin, maskulin plural

Genus-Marker steht nur, wenn Eintrag Meishi ist und nur bei dem Wort, um das es geht.

Kommentare oder Ergänzungen stehen in runden Klammern.

Herkunft eines Wortes ist mit "Etym." markiert.
Fremdwörter, die z.B. kursiv dargestellt werden sollten, sind mit "For." markiert. Bei manchen Einträgen wäre es eine Wiederholung, wenn man das Wort von dem abgeleitet wurde, wiederholt. In diesem Fall ist nur die Ursprungssprache mit der Markierung "impli." angegeben; es gibt dann noch eine explizite Version, bei der das Wort tatsächlich steht: "expli.". Bei Fremdwörtern aus dem Deutschen z.B. steht das Wort, von dem abgeleitet wurde häufig schon in der Übersetzung. Eine englische Version würde jedoch das deutsche Wort als Fremdwort brauchen. Das ist also bereits vorbereitet.

Allgemeine Erklärungen sind mit "Expl." markiert. Diese Erklärungen stehen normalerweise in Klammern.

Wissenschaftliche Name aus Biologie, Botanik, Zoologie, Medizin, Anatomie etc sind mit "Scientif." markiert: man könnte diese kursiv setzen. Wissenschaftliche Namen stehen häufig in Klammern.

Jahreszeitenwörter der japanischen Poesie sind mit "SeasonW." markiert; es gibt als Inhalt: Frühling, Sommer, Herbst, Winter und Neujahr.

Wortart ist jetzt mit dem Tag "POS" markiert.

Familiennamen sind mit "FamN." markiert (man müsste eigentlich auch den Rufnamen extra markieren, wenn es mehrere Vornamen gibt).

URL sind als "URL" markiert.

Verweis auf Wikipedia sind mit "WikiDE", "WikiEN" bzw. "WikiJA" markiert.

Japanischer Text im Deutsch-Feld ist mit Jap. markiert; hier muss nämlich ggf. Font oder Codierung besonders gehandhabt werden.

Transkription einer japanischen Schreibung ist mit "Transcr." markiert.

Übersetzungen sind als "Transl." markiert

Referenzen auf einen anderen Eintrag sind mit "Ref." markiert.

- Synonyme:
- Antonyme:
- Homographen mit anderer Lesung aber selber Bedeutung: , normalerweise wird von der selteneren Version auf die übliche verwiesen;
- andere Kana-Transkription eines Fremdwortes:

ID eines Eintrags auf den referenziert wird, ist mit "DaID" makiert; es flogt die ID des entsprechenden Eintrags. wenn auf eine best. Bedeutung eines Eintrags verwiesen wird,

Bilder sind markiert als "Pict."; Bildtitel ist "Capt."; entsprechender Dateiname ist "FileN"

Definition sind mit "Def." markiert. Es gibt Definitionen, die zusätzlich zu Übersetzungen in Klammern stehen. Es gibt Einträge, bei denen Übersetzung schwierig bzw. unmöglich ist. Da steht dann nur die Definition.

Zahlwort mit dem das Lemma gezählt wird, steht als "Count."

weitere Beschreibung der Grammatik sind als "Descr." markiert.

Veröffentlichungsjahr von Werken sind mit "Date" markiert.

Hervorhebungen, die z.B. fett gesetzt werden sollten, sind mit "Emph." markiert.

Lebensdaten von Persönlichkeiten sind mit "BirthDeath" markiert.

Sprachniveau ist mit "LangNiv" markiert: ugs., schriftspr., vulg., gaunerspr., kinderspr., poet.

Verwendungsweise ist als "Usage" markiert; hier z.B. Höflichkeit, onomatopöia

Anführungszeichen sind markiert wie sie gemeint sind: wörtlich "literal", ironisch "iron", Thema: "Topic"

Kursiv zu setzende Werktitel sind mit "Title" markiert; Originalsprache ist als "ORIG" markiert; Sprache ist z.B. als Attribut "L_FRA" angegeben.

Priorität von Übersetzungen ist zurzeit mit "<Prior_1>" markiert; das muss überarbeitet werden; wahrscheinlich werde ich für Eintragsbearbeitung mehrere Datenbanken anlegen.

Es gibt einige Attribute wie BezKomm (Bezug auf Kommentar); BezÜbers (Bezug Übersetzung, nicht Eintrag/Lemma); BezTeilBed (Bezug auf Teilbedeutung); BezEintr. (Bezug auf Eintrag und nicht nur letzte Übersetzung); Einträge, bei denen japanischer Eintrag und deutsche Übersetzung problemlos umkehrbar sind, sind mit rev markiert.

Feld "DaWortart"

Enthält die Beschreibungen der Wortart und der Konjugationsart auf Japanisch. Feld ist durch die Angabe POS im Deutsch-Feld redundant geworden. Häufig ist Angabe Wortart nicht eindeutig oder nicht sinnvoll.

DaID

Eindeutiger Schlüssel der Einträge.

Feld "DaJapKommentare"

Hier wird auf Synonyme und Antonyme und alternative Lesungen verwiesen:

- Weißer Pfeil "" verweist auf Synonym;
- weißer Doppelpfeil " " verweist auf Antonym;
- schwarzer Pfeil "" verweist auf eine alternative Lesung des japanischen Eintrages.

Feld ist inzwischen durch Referenzen "Ref." ersetzt und könnte eigentlich gelöscht werden.

Feld "DaDokumentation"

Enthält Kommentare zu Bearbeitung und Verweise oder auch Kommentare zu anderen Wörterbüchern, insbes., wenn diese Fehler enthalten; es besteht die Gefahr, dass sich diese Fehler wieder bei uns einschleichen; eignet sich nicht unbedingt zur Veröffentlichung.

Feld "DaRomajiBearbLesung"

Enthält die Lesung in Hiragana mit Markierungen zur Segmentierung der Einträge durch Markierungen, Leerzeichen, Unterstriche, Tilden, Nakaten, Pipes etc.. Mit diesem Mark-up kann man verschiedene Geschmacksrichtungen Latein-Umschrift und Furigana generieren.

Bei Untereinträgen stehen die Kana des entsprechenden Haupteintrags in Spitzklammern und könnten für verkürzte Schreibweise durch Tilde ersetzt werden. Wenn der Haupteintrag aufgrund von z.B. Konjugation nicht wirklich mit dem Haupteintrag identisch ist, ist der öffnenden Spitzklammer ein Schrägstrich "/" als Escape-Zeichen vorangestellt. Wenn sich ein Untereintrag auf mehrere Haupteinträge bezieht, sind die Spitzklammern nummeriert.

Das Feld dient dazu, die Hiragana z.B. per Formel in Romaji umzurechnen.

Markiert werden Joshi, bei denen Umschrift problematisch ist (ist wa nicht ha und ist e nicht he), mit "Jo".
Bei Großschreibung wird "Gr" voran gestellt.
Devokalisierung wird mit "Dev" vor der entsprechenden Silbe angezeigt. Bei diesen Silben könnte bei Lateinumschrift ein kleiner Kringel unter i, u oder o gestellt werden.
G-Laut im Wortinneren, der nicht nasalisiert wird, ist mit "NN" vor der Silbe markiert; in der Lateinumschrift könnte man z.B. einen Apostroph vor diese Silbe setzen. Das betrifft Fremdwörter, bei denen der velare Verschluss nicht nasaliert ist.

Leerzeichen stehen vor Worten.

Tilden "~" stehen vor dem flektierten Teil eines Verbs oder Verbaladjektives.

Nakaten steht, um Worte weiter zu segmentieren und lesbarer zu machen.

Pipe steht bei Segementierung, die für Furigana eine Rolle spielt aber nicht für Lateinumschrift.

Unterstriche stehen, wenn eine Silbenkombination zu einer Längung führen könnte, obwohl zwei Vokale stehen sollen. Möglicherweise überschneidet sich das mit Verwendung der Apostrophen und fällt irgendwann weg.

Apostroph trennt Silben, die sich auf unterschiedliche Kanji beziehen; das wäre wichtig für Furigana.

Um DIN-Umschrift zu erreichen, müssten einige dieser Nakaten in Leerzeichen umgewandelt werden und einige müssten ignoriert werden. An manchen Stellen würde ich Leerzeichen setzen, wo DIN keines möchte (z.B. Nomen mit suru). Hierfür müssen wir uns noch weiteren Markup ausdenken.

Feld "DaMidashigo"

Enthält Kurzfassungen für die Schreibung der japanischen Einträge. Wenn das Feld leer ist, wird stattdessen der Inhalt des Feldes "DaJapanisch" verwendet.

- Nicht-Toyo-Kanji sind mit Kreuz "×" markiert.
- Toyo-Kanji mit Nicht-Standard-Lesung sind mit Dreieck "" markiert.
- Hiragana, die wegfallen können, stehen in runden Klammern.
- Einträge mit unregelmäßigen Lesungen, die trotzdem als Standard verwendet werden, stehen in geschweiften Klammern.
- Andere Einträge mit unregelmäßiger Lesung stehen in flachen Spitzklammern.
- Sind die Markierungen selbst Teil des Eintrages z.B. das Kreuz als batsu, dann geht ihnen ein Schrägstrich als Escape-Zeichen voran.

Feld "UntereintragMidashigo"

Hier werden die Haupteinträge, wie sie im Untereintrag vorkommen markiert. Wenn der Haupteintrag aufgrund von z.B. Konjugation nicht wirklich mit dem Haupteintrag identisch ist, ist der öffnenden Spitzklammer ein Schrägstrich "/" als Escape-Zeichen vorangestellt. Wenn sich ein Untereintrag auf mehrere Haupteinträge bezieht, sind die Spitzklammern nummeriert. Ziel dieser Markierungen ist es, in Untereinträgen den entsprechenden Haupteintrag z.B. durch Tilde zu ersetzen, um im Drucksatz Platz zu sparen oder um den Haupteintrag hervorzuheben.

Feld "DaEintragsebene"

Enthält Markierung, ob ein Eintrag Haupteintrag ist: HE. Bei Untereinträgen enthält das Feld den Inhalt des Feldes "DaJapanisch" des entsprechenden Haupteintrages. Über diesen Schlüssel werden die Felder per Relation verbunden und man kann über die Relation alle Untereinträge eines Haupteintrages zusammensuchen lassen. Das Feld kann mehrere Schlüssel enthalten. Sie müssten jeweils durch vertikalen Tabulator markiert sein.

DaEintrEbeneID

Wie "DaEintragsebene", nur bei Untereinträgen steht die ID (DaID) der Haupteinträge.

Feld "DaUE_Art"

Gibt die Art des Untereintrags an:

- Abl. mit <Umschr.: da>
- Abl. mit <Umschr.: de>

- Abl. mit <Umschr.: e>
- Abl. mit <Umschr.: kara>
- Abl., Adverb mit <Umschr.: ku>
- Abl. mit <Umschr.: na>
- Abl. mit <Umschr.: ni>
- Abl. mit <Umschr.: no>
- Abl. mit <Umschr.: shita>
- Abl. mit <Umschr.: shite>
- Abl. mit <Umschr.: teki>
- Abl. mit <Umschr.: to>
- Abl. mit <Umschr.: taru>
- Abl., Nominalisierung mit <Umschr.: sa>
- Abl. mit <Umschr.: suru>
- Komp. Anf.
- Komp. Hint.
- Verwendungsbeispiel
- W_Idiom
- XSatz
- Z_Sprichwort

Untereinträge sollten bei Anzeige mit ihrem Lemma entsprechend der Untereintragsart sortiert sein.

Feld "DaUnterbedBezug"

gibt die Unterbedeutung des Haupteintrages an, auf den sich der Untereintrag bezieht.

Feld "DaUnterbed"

auch ein Untereintrag kann verschiedene Unterbedeutungen haben. Wenn hier nicht eindeutig ist, welche Unterbedeutung des Untereintrages jeweils gemeint ist, wird hier per Nummer markiert.

Feld "DaBild"

ist gefüllt, wenn eine entsprechende Bilddatei vorhanden ist. Bildtitel steht in geraden Anführungszeichen. Dateiname folgt dann. Wenn es mehrere Bilder für einen Eintrag gibt, gibt es ein Semikolon plus Leerzeichen als Trennzeichen

Feld "DaInhaltlRelation"

hier können inhaltliche Relationen vermerkt werden, z.B. alle Großlandschaften

Feld "DaNormAkzent"

enthält eine Zahl; die den Standard-Tonakzent des Eintrags angibt. Die Zahl gibt die letzte Silbe mit hohem Ton an. In der Umschrift könnte nach der so bezeichneten Silbe das Zeichen für den Tonwechsel nach unten kommen. (Es fehlt noch ein Ansatz für Einträge, bei denen es Akzentmuster für mehr als ein Wort gibt)

Feld "DaAltAkzent"

enthält ein oder durch Semikolon getrennte Zahlen, die alternativ mögliche Tonakzente des Eintrags angeben.

Mapping

Feld	SQL-Spaltenname
DaJapanisch	japanese
DaLesung	reading
DaDeutsch	german
DaID	uid
DaWortart	wordtype
DaJapKommentare	japcomment
DaDokumentation	comment
DaRomajiBearbLesung	romajireading
DaMidashigo	midashigo
UntereintragMidashigo	submidashigo

DaEintragsebene	reference_word
DaEintrEbeneID	reference
DaUE_Art	subentrytype
DaUnterbedBezug	sub_entry_ref
DaUnterbed	sub_entry_ref2
DaBild	image
DaInhaltlRelation	content_relation
DaNormAkzent	accent
DaAltAkzent	altaccent