







hbz

Wissen. Information. Innovation.





das neue hbz

Produkte und Dienstleistungen

-  Flexibilisierung des Angebotes und der Produkte
-  Fortführung und Ausbau der klassischen hbz-Dienstleistungen
-  Neue Produkt-Kombinationen und Entwicklung neuer Produkte
-  Stärkere Markt-Orientierung und flexiblere Reaktion auf Markt-Entwicklung



Personalressourcen

-  Straffung und Optimierung von Ressourcen und Personal-Einsatz
-  Verlagerung der Fortbildung an die FH Köln
-  Umwandlung von bisherigen Stellen in acht qualifizierte IT-Stellen
-  Verlagerung von einigen Stellen, z.B. aus dem Bereich des ehemaligen ZK, an andere Hochschulen



hbz

Wissen. Information. Innovation.

Erschließung von Dokumenten in nicht-lateinischen Schriften in Verbundsystemen








*CJK-Katalogisierung in Originalschrift und
Fremddatenübernahme aus MARC-Datenbanken*

Günter Hupfer

Vortrag auf dem 94. Deutschen Bibliothekartag in Düsseldorf, 16.3.2005

Inhalt

CJK=chinesisch, japanisch, koreanisch

-  Einführung: *die Situation in Deutschland*
-  das hbz-Projekt: *die Voraussetzungen*
-  Grundlagen CJK: *die Originalschrift*
-  Implementierung: *ALEPH 500*
-  Entscheidung: *Arbeitspakete*
-  Arbeitsstand: *Setup*
-  Weitere Arbeiten: *Perspektiven*



hbz-Verbund Datenbank

CJK: Einführung

Wie sieht die CJK-Welt in
Deutschland aus?

Situation in Deutschland

"CJK-Literatur ist normale Literatur"

- ✍ Bibliotheken mit Ostasiatischer Literatur:
 - ✍ große Bestände in zwei Bibl.: SBB, BSB
 - ✍ Mehrzahl der Bibliotheken = Seminarbibl.

- ✍ Erschließung: heterogene, lokale Lösungen
 - ✍ *"fehlende"* Mehrschrittlichkeit in Verbänden
 - ✍ keine gemeinsame Datenbasis
 - ✍ Realisierungen z.B. mit CJK-Allegro
 - ✍ keine Nutzung gemeinsamer Standards
 - ✍ Katalogisierung in "fremden" Systemen

CJK-Workshop, 2001

Das macht jetzt das HBZ!

- ✍ CJK-Workshop, Hamburg, 2001
 - ✍ Thema: *"CJK-Fachbibliotheken auf dem Weg in die Verbände"*
 - ✍ Organisation: Institut für Asienkunde, Hamburg
- ✍ Ziel: Integration in die Verbände
 - ✍ Verwendung gemeinsamer Standards
 - ✍ zentraler Nachweis + Fremddatenübernahme
 - ✍ Kooperation der beteiligten Einrichtungen

Voraussetzungen im hbz

- ✍ UNICODE in ALEPH 500
 - ✍ vollständige Unicode-Fähigkeit ab Version 14
 - ✍ Umstieg des hbz von Version 12.3 auf Version 16.02 im Juli 2004
- ✍ Grundvoraussetzungen gegeben
- ✍ hbz-Kundenanforderungen hinsichtlich Katalogisierung in Originalschrift



hbz-Verbund Datenbank

CJK: Projekt am hbz


Was hat das hbz mit CJK vor?

Das hbz-Projekt

- ✍ Verbundkatalogisierung in Originalschrift
- ✍ Universität Bonn: Planung
Fachbereichsbibliothek Asienliteratur
- ✍ Pilotpartner ULB Bonn
- ✍ Pilotbibliothek:
Japanologisches Seminar, Univ. Bonn
- ✍ CJK-Katalogisierung mit Kostenbeteiligung

Fremddaten: NACSIS-CAT

Fremddatenübernahme

-  gegenseitige Vereinbarung hbz und National Institute of Informatics (NII), Tokyo

National Institute of Informatics

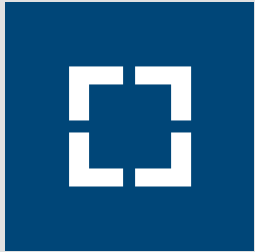
-  Betreiber größter Bibliotheksverbund Japans
-  Katalog NACSIS-CAT

Datenbestand: ca. 8 Mio Titelsätze

-  darunter hoher Anteil an Originalschrift
-  Stichproben Fremddatennutzung positiv

Kooperation ALEPH-Verbünde

- ✍ Österreichischer Bibliothekenverbund
 - ✍ homogene ALEPH-Lokalsysteme im OBV
 - ✍ Erfahrungen mit originalschriftl. Erfassung in zentraler Datenbank ACC01
 - ✍ "regionale" Lösung
 - ✍ ohne Fremddatenübernahme
- ✍ Kooperation ALEPH-Verbundsysteme
 - ✍ OBV-Lösung: ein "Baustein" für hbz-Projekt
 - ✍ hbz-Projekt:
Schritt einer erweiterten CJK-Lösung



hbz-Verbund Datenbank

Grundlagen CJK

Was sind das für Zeichen?

Wie werden die Zeichen
eingegeben/dargestellt?

Wie erfolgt der Datenaustausch?

CJK: Schriftsystem

- ✍ CJK ist ein "Satz" von Schriftzeichen
- ✍ Schriftliche Darstellung CJK-Sprachen
 - ✍ durch Ideogramme bzw. Logogramme
 - ✍ jedes Zeichen repräsentiert ein Wort oder eine bedeutungstragende Worteinheit
- ✍ Umfang der Schriftzeichen
 - ✍ gemeinsam: 20–70 tsd. Han-Ideogramme
 - ✍ jede Sprache nutzt weitere spezifische Schriftzeichen


CJK: Zeicheneingabe

- ✍ Betriebssystem in "CJK"
- ✍ Microsoft Windows: Input Method Editor
 - ✍ Eingabe der Zeichen in Transliteration
 - ✍ "übersetzt" vom IME nach CJK
 - ✍ IMEs für die verschiedenen Sprachen
- ✍ Nachinstallation IME problemlos
 - ✍ Systemsteuerung
 - Regions-Sprachoptionen (Ländereinstellg.)
 - Eingabe

Regelwerk

- ✍ RAK: lateinische Schrift vorgesehen
 - ✍ Umschrift für nicht-lateinische Schriften über Transliterationstabellen
- ✍ AACR: "Arbeitssprache"
 - ✍ Angaben in Sprache und Schrift der Vorlage (wenn möglich, sonst: Transliteration)
- ✍ Transliteration = Informationsverlust
- ✍ Ziel: Original-/Mehrschriftlichkeit
 - ✍ unabhängig vom Status des Regelwerks

Beispiel IME-Eingabe

<u>100</u>	<u> </u>	<u>a</u>	Murakami, Haruki
		<u>9</u>	119037092
<u>A00</u>	<u> </u>	<u>6</u>	Jpan
		<u>a</u>	村上春樹
<u>331</u>	<u> </u>	<u>a</u>	Kokkyō no minami, taiyō no nishi
<u>C31</u>	<u> </u>	<u>6</u>	Jpan
		<u>a</u>	国境の南 太陽の西
<u>359</u>	<u> </u>	<u>a</u>	Mur 
<u>C59</u>	<u> </u>	<u>6</u>	Jpan
		<u>a</u>	村上春樹
<u>410</u>	<u> </u>	<u>a</u>	Tōkyō
<u>D10</u>	<u> </u>	<u>6</u>	Jpan
		<u>a</u>	東京

Datenformat

- ✍ Austauschformate: MARC21 und MAB2
- ✍ "Datenmodell" für Mehrschrittlichkeit
 - ✍ ein Datensatz, nicht mehrere Datensätze
- ✍ primäre Schrift, sekundäre Schrift
- ✍ Felder für sekundäre Schrift
 - ✍ Wiederholung der Standard-Felder
 - ✍ MARC21: Feld 880
 - ✍ MAB2: Feld 671



hbz-Verbund Datenbank

CJK: Implementierung

Was ist alles zu bedenken?

Was kann ALEPH 500?

ALEPH 500: Erfassung

- ✍ Standard-Client: Unicode
 - ✍ Erfassungsformat: MAB2 oder MARC21
 - ✍ Austauschformat nicht für Erfassung und Indizierung in ALEPH 500 geeignet
- ✍ Varianten für mehrschritt. Erfassung
 - ✍ identische Felder: Spezifizierung durch Indikatoren/Unterfelder
 - ✍ abweichende Feldbezeichnungen
- ✍ Konverter für Austausch-/Internformat

ALEPH 500: Fremddaten

- ✍ Fremddaten mit CJK-Schriftzeichen liegen in MARC21 und nicht in MAB2 vor
- ✍ Z39.50-Setup für MARC21-Quellen
- ✍ Datenkonvertierung
 - ✍ MARC21-Datenbank:
Nutzung MARC21 ohne Konvertierung
 - ✍ MAB2-Datenbank:
ALEPH 500-Konverter MARC21 – MAB2
- ✍ Zeichensatz: UTF 8

ALEPH 500: Indizierung

- ✍ Die Herausforderung Indizierung
 - ✍ CJK-Schriftzeichen sind/werden i.d.R. ohne Blank erfasst sind
- ✍ Die Lösung: spezielle Routinen
 - ✍ spezielle ALEPH 500 Indizierungsrountinen je Sprache ("Worterschlagung")
- ✍ Das Problem: gemeinsame Indizes
 - ✍ Bibliotheken mit überwiegend CJK
 - ✍ Bibliotheken mit gemischten Katalogen

ALEPH 500: filing-routinen

cjk_pinyin	Adds ! before each cjk "character", translates the characters to pinyin, using the Z114 table, and adds the unicode value in decimal notation. The ! causes the pinyin filing-text to be sequenced separately from regular Latin characters.
cjk_stroke	Same as cjk_pinyin, except that each character is translated to stroke value, using the Z114 table.
chi_pinyin	Translates each character to pinyin, using the Z114 table, and adds the unicode value (in decimal notation) for each character. The Unicode value is added in order to differentiate between different characters that have the same pinyin value. Because the pinyin filing-text will be sequenced together with regular Latin characters, this routine should be used for browse lists that use the language code from 008 to create separate browse lists.
chi_stroke	Same as chi_pinyin, except that each character is translated to stroke value, using the Z114 table.
jpn	Translates each character to the decimal value of the unicode character. This causes the filing-text to be sequenced together with regular characters. Therefore, this routine should be used for browse lists that use the language code from 008 to separate by language, and are separate for "jpn".
kor	Translates each character to the decimal value of the unicode character. This causes the filing-text to be sequenced together with regular characters. Therefore, this routine should be used for browse lists that use the language code from 008 to separate by language, and are separate for "kor".

ALEPH 500: CJK-Specials

- ✍ Weitere Mechanismen für Bearbeitung von CJK-Daten in ALEPH 500
- ✍ chinesisches Wörterbuch für spezielle Indizierung chinesischer Feldinhalte
- ✍ fix-doc-Programme für automatische Transliteration
- ✍ weitere Programme auf Basis von MARC21 für Erfassung

ALEPH 500: Exportschnittstellen

- ✍ KAT-Download MAB2/MARC21
 - ✍ bisher noch nicht in UTF8 möglich
- ✍ Versorgung von Nicht-ALEPH-Lokalsyst.
 - ✍ bisherige Datendienste: MAB2-Zeichensatz
 - ✍ Entwicklung UTF8-Export MAB2/MARC21
- ✍ Versorgung ALEPH-Lokalsysteme
 - ✍ zentrale Erfassung MAB2 bzw. MARC21 hat Auswirkungen auf Replikation
 - a) Information + Setup
 - b) Konverter wenn MARC21



hbz-Verbund Datenbank

Entscheidung u. Arbeitspakete

So machen wir es!

Entscheidung



Kernfrage

-  Integration in HBZ01 (MAB2)
oder eigene Datenbank (MARC21)?

Entscheidung: Integration in HBZ01

-  Erfassung UND Fremddatenübernahme

Arbeitspakete mit dem Ziel:

-  Erweiterung HBZ-ALEPH 500 Datenformat
-  Erweiterung Z39.50-Schnittstelle für MARC21-Daten mit CJK-Feldern (Konverter zu MAB2)

CJK-Erfassung in HBZ01

- ✍ Arbeitspaket "CJK-Erfassung"
- ✍ spezielle, ausgewählte Felder für CJK-Schriften, Umschrift weiterhin Pflicht
- ✍ Selektionskennzeichen für CJK
- ✍ Standardindizierung der CJK-Inhalte
- ✍ Anzeige der Felder im GUI-Client und Verbundkatalog (Web)
- ✍ "lokale" Sichten für Recherche

Fremddatenübernahme

- ✍ "Arbeitspaket" Fremddatennutzung
- ✍ Z39.50-Anbindung NACSIS-CAT
- ✍ Zeichensatz UTF8
- ✍ Analyse der Daten
- ✍ Evaluierung des Konverters von ALEPH 500 MARC21 – MAB2
- ✍ Anpassung des Konverters an die HBZ-CJK-Verhältnisse

Produktionsbetrieb

- ✍ CJK-Projekt ohne Auswirkungen auf bestehenden Produktionsbetrieb
- ✍ noch keine CJK-Sonderindizierung
- ✍ noch kein Export der CJK-Felder
- ✍ keine Auswirkungen auf bestehende MAB2-Exportschnittstellen
 - ✍ KAT-Download, MAB-Dienst
 - ✍ Z39.50, Replikation zu ALEPH-Lokalsystemen

Realisierter Arbeitsstand

- ✍ Konfiguration in ALEPH 500
 - ✍ sehr flexibel
- ✍ Arbeitspaket "Erfassung" erledigt
 - ✍ Felddefinitionen, Plausibilitäten
 - ✍ Anzeige (Client, Web)
 - ✍ Standardindizierung der CJK-Felder
 - ✍ Export CJK-Felder "unterdrückt"
 - ✍ noch keine Übernahme in Produktion

Erfassungsformat

- ✍ Standard-MAB2-Felder ohne Änderungen
- ✍ neue Felder für Originalschrift
 - ✍ alphanummerische Feldnamen
 - ✍ 1. Stelle MAB2 = Buchstabe
ansonsten identisch
1 = A, 2 = B, 3 = C usw., 0 = Z
 - ✍ Beispiele: Feld 100 -> A00
Feld 331 -> C31
 - ✍ Indikatoren wie MAB2
 - ✍ UF 6 für Schriftcode

ALEPH-Lokalsysteme

- ✍️ Auswirkungen des CJK-Setups im Verbundsystem auf ALEPH-Lokalsysteme
 - ✍️ Replikation von HBZ01-Datensätzen mit den CJK-Feldern in Originalschriften
- ✍️ Maßnahmen Setup ALEPH-Lokalsysteme
 - ✍️ "Unterdrückung" der Felder
 - ✍️ "Setup" der CJK-Felder



hbz-Verbund Datenbank

Beispiele

So sieht es aus?

<u>100</u>		<u>a</u>	Murakami, Haruki
		<u>9</u>	119037092
<u>A00</u>		<u>6</u>	Jpan
		<u>a</u>	村上春樹
<u>331</u>		<u>a</u>	Kokkyō no minami, taiyō no nishi
<u>C31</u>		<u>6</u>	Jpan
		<u>a</u>	国境の南、太陽の西
<u>359</u>		<u>a</u>	Murakami, Haruki
<u>C59</u>		<u>6</u>	Jpan
		<u>a</u>	村上春樹
<u>410</u>		<u>a</u>	Tōkyō
<u>D10</u>		<u>6</u>	Jpan
		<u>a</u>	東京
<u>412</u>		<u>a</u>	Kōdansha
<u>D12</u>		<u>6</u>	Jpan
		<u>a</u>	講談社
<u>425</u>	<u>a</u>	<u>a</u>	1995
<u>433</u>		<u>a</u>	299 S.
<u>451</u>		<u>a</u>	Kōdansha bunko
<u>D51</u>		<u>6</u>	Jpan
		<u>a</u>	講談社文庫
<u>540</u>	<u>a</u>	<u>a</u>	4062630869

Katalogisierung: Editor

ID	[HT014298306]
1. Person	Murakami, Haruki [119037092]
1. Person	Jpan 村上春樹
Titel	Kokkyō no minami, taiyō no nishi
Titel	Jpan 国境の南、太陽の西
Verfasserang.	Murakami, Haruki
Verfasserang.	Jpan 村上春樹
Ort	Tōkyō
Ort	Jpan 東京
Verlag	Kōdansha
Verlag	Jpan 講談社
Jahr	1995
Umfangsang.	299 S.
1. Gesamttitel	Kōdansha bunko
1. Gesamttitel	Jpan 講談社文庫
BESTAND	
Ansicht	Alle Exemplare
Bestand	5 Bonn: ULB

Katalogisierung: Anzeige


Gesamtkatalog - Suche über mehrere Felder - Mozilla

File Bearbeiten Ansicht Gehe Lesezeichen Tools Fenster Hilfe

Zurück Vor Neu laden Stopp <http://193.30.112.129/F/DD413UM7PSHU2RTQVFT4K5PMVU16MDHQ778GN1k> Suchen Drucken

Startseite Lesezeichen ALEPH FAST-Links DDB Ex Libris Support RAK/AACR-MAB/MARC KSTA 94. DBT -> Startseite

HBZ-Verbundkatalog - Gesamtkatalog

 **HBZ**

[Standardsuche](#) [Expertensuche](#) [Indexsuche](#) [Ergebnisliste](#) [Hilfe](#)
[Einstellungen](#) [Suchverlauf](#) [Korb](#) [Kataloge](#) [Feedback](#) [Neustart](#)

Standardsuche

Alle Felder* ?

und Titelstichwort 国境の南、太陽の西

und Person

* Alle Felder = Stichwörter aus den Feldern Titel, Person, Körperschaft, Schlagwort.

Suche beschränken auf:

Jahr: Sprache:

Dokumenttyp: [Suche nach Bibliothek einblenden](#)

Verbundkatalog: Suche

6	<input checked="" type="checkbox"/>	Murakami, Haruki: Kokkyō no minami, taiyō no nishi / Murakami, Haruki Tōkyō : Kōdansha, 1995 (Kōdansha bunko)	5-Bonn.-ULB
7	<input type="checkbox"/>	Murakami, Haruki: Kokkyō no minami, taiyō no nishi / Murakami Haruki. - 6-satsu Tōkyō : Kōdansha, 1997 (Kōdansha bunko [; Mu-6-14])	385-TR.-UB
8	<input type="checkbox"/>	Murakami, Haruki: South of the border, west of the sun / Haruki Murakami. Transl. from the Japanese by Philip Gabriel New York : Knopf, 1999	61-D.-ULB
9	<input type="checkbox"/>	Murakami, Haruki: South of the border, west of the sun / Haruki Murakami London : Harvill Press, 1999	385-TR.-UB

Verbundkatalog: Ergebnisliste

Gesamtkatalog - Vollanzeige - Mozilla

http://193.50.112.120/F/DH/13UN7P5H.DRTQ/FT4K5FNULI6NDHQ7782CU/702CRVT0/SAG-014567func=find=afanf_code=WRD&ac=Suchen

Vollanzeige

Ergebnistafel Speichern/Senden In den Korb

Satz 1 von 1
Format: [Standard](#) [Feldnamen](#) [Feldnummern](#)

Verfasser: [Murakami, Haruki \[119037092\]](#)
 Titel: [Kokkyō no minami, taiyō no nishi](#)
 ID: HTD14298306

1. Person: [Murakami, Haruki \[119037092\]](#)
 1. Person: [Jpan 村上春樹](#)
 Titel: [Kokkyō no minami, taiyō no nishi](#)
 Titel: [Jpan 国境の南、太陽の西](#)
 Verfasser/Urheber: Murakami, Haruki
 Verfasser/Urheber: Jpan 村上春樹
 Ort: Tōkyō
 Ort: Jpan 東京
 Verlag: Kōdansha
 Verlag: Jpan 講談社
 Jahr: 1995
 Umfangsang.: 299 S.
 1. Gesamttitel: [Kōdansha bunko](#)
 1. Gesamttitel: [Jpan 講談社文庫](#)
 ISBN: 4062630869

BEZEHUNGEN

BESTAND

Bestand: [Alle Exemplare](#)
 Bestand: [5 Bonn, ULE](#)

Verfasser	● Murakami, Haruki [119037092]
Titel	● Kokkyō no minami, taiyō no nishi
ID	HTD14298306
1. Person	● Murakami, Haruki [119037092]
1. Person	● Jpan 村上春樹
Titel	● Kokkyō no minami, taiyō no nishi
Titel	● Jpan 国境の南、太陽の西
Verfasser/Urheber	Murakami, Haruki
Verfasser/Urheber	Jpan 村上春樹
Ort	Tōkyō
Ort	Jpan 東京
Verlag	Kōdansha
Verlag	Jpan 講談社
Jahr	1995
Umfangsang.	299 S.
1. Gesamttitel	● Kōdansha bunko
1. Gesamttitel	● Jpan 講談社文庫
ISBN	4062630869

Verbundkatalog: Anzeige



hbz-Verbund Datenbank

Arbeiten / Perspektiven

Was ist noch zu tun?

Welche weiteren Ziele haben wir?

Weitere Arbeiten

Setup-Arbeiten

 Fremddatenübernahme

 Konsolidierung des Setups

 Lokale Sichten für Suche im hbz-Katalog

 Indizierung mit CJK-Routinen

Test mit Bibliotheken

 Pilotbibl. Japanolog. Seminar, Bonn

 NEU: UB Duisburg/Essen, Sto. Duisburg als weitere Referenzbibliothek

Produktionsbetrieb ab Sommer 2005

Perspektiven

- ✍ Weitere hbz-Bibliotheken
- ✍ Kundenanforderungen
 - ✍ Exportschnittstellen
 - ✍ weitere Fremddatenquellen
 - ✍ CJK-Normdaten-Anbindung
- ✍ Originalschriftlichkeit außerhalb CJK
 - ✍ CJK ist der "Schlüssel"
 - ✍ z.B. kyrillische Schrift, griechische Schrift

Perspektiven

- ✍ Interessante Lösung auch für bislang Nicht-Verbundbibliotheken
 - ✍ neue hbz-Kunden
- ✍ Kooperation mit ...
 - ✍ ... Bibliotheken aus Nicht-hbz-Bereich
 - ✍ ... anderen ALEPH-Verbänden
- ✍ Vision "*CJK-Verbund*"

Zusammenfassung

- ✍ CJK-Situation in Deutschland
- ✍ hbz-Projekt
 - ✍ Japanol. Seminar, Universität Bonn
 - ✍ NACSIS-CAT, Tokyo
- ✍ Ziel: HBZ01 + Fremddatennutzung
- ✍ Implementierung ALEPH 500
- ✍ Ausweitung CJK-Teilnehmerkreis
 - ✍ *"CJK-Verbund"*

Weitere Informationen

 im hbz

 Brigitte Block: block@hbz-nrw.de

 Dr. Ronald Schmidt: schmidt@hbz-nrw.de

 Günter Hupfer: hupfer@hbz-nrw.de