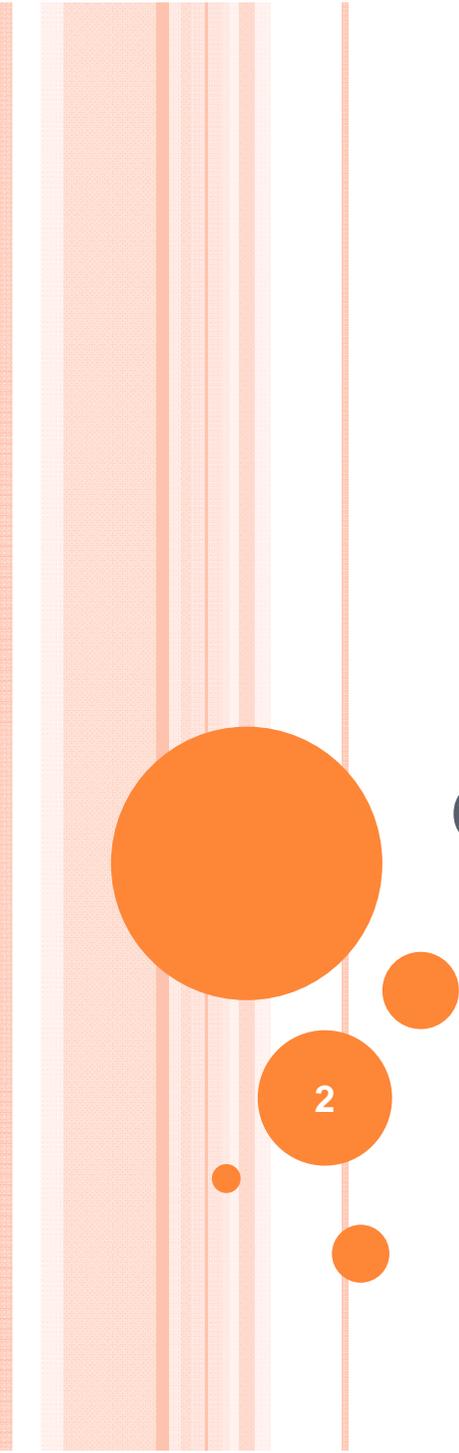


# CiNii Books紹介 NACSIS-CATの正規化处理

国立情報学研究所  
学術基盤推進部 学術コンテンツ課  
藤井 眞樹

平成24年3月1日



# CiNii BOOKS 紹介

2

# 1. 1 CiNii Books - 大学図書館の本を探す



<http://ci.nii.ac.jp/books/>

2011年11月9日リリース

- NACSIS-CAT書誌所蔵データの検索・表示サービス
  - 書誌1100万件・所蔵1億1千万件
  - 参加館1200館以上
  - 著者 約150万件
  - ウェブAPIの提供
    - 外部サイトからNACSIS-CATのデータを利用可能に
- NACSIS Webcatの後継
  - 平成24年度末(2013年3月)に終了予定

## 1.2 コンセプト, 設計方針

- NACSIS-CATシステムとの分離
  - モジュール化・軽量化
  - クラウド対応可→業務継続計画への対応
- ユーザーインターフェイス
  - CiNii Articlesと同等, 一般のユーザーのためのデザイン
- CATデータの活用
  - 著者名典拠
  - 書誌のNOTEなども検索可能に
- 図書館との連携
  - 図書館システム連携
  - ILLに有用な情報の提供

## 1.3 検索

The screenshot shows the CiNii Books search interface. At the top, there is a navigation bar with the CiNii logo, the text '日本の論文をさがす' and '大学図書館の本をさがす', and an 'English' link. Below this, a date notice reads 'CiNii Books, 全文検索の公開について (2011年11月9日)'. The main header features the CiNii Books logo and the text '大学図書館の本をさがす'. The search interface includes tabs for '図書・雑誌検索' and '著者検索', with a link to '日本の論文をさがす'. A search input field is present. Below it, a '▼ 詳細検索' section contains various search criteria: 'タイトル' (Title) with a checkbox for 'タイトル完全一致'; '著者名' (Author name), 'ISBN', '件名' (Subject), '図書館ID' (Library ID); '著者名典拠ID' (Author name authority ID), 'ISSN', '分類' (Classification), '機関ID' (Institution ID); '出版者' (Publisher), 'NCID', '注記' (Notes), '出版年' (Publication year) with a range selector; '言語種別' (Language type) with a dropdown for 'すべての言語' and a 'コード' (Code) field; and '地域' (Region) with a dropdown for 'すべての地域'. At the bottom of the form, there are radio buttons for 'すべての資料' (All materials), '図書' (Books), and '雑誌' (Magazines), and a checkbox for 'ローマ字をカナに変換' (Convert Roman letters to Kana). A '検索' (Search) button is located at the bottom center.

### ○ 詳細検索画面

- 地域, 機関ID, 図書館ID (参加組織番号) での検索可能
- 件名・分類・注記・資料種別・言語
- 所蔵からのOPACリンク
- ILL関連 (所蔵館の地域での絞り込み, ILL参加, 相殺参加, 日米ILL, 日韓ILL参加での絞り込み)
- 部分一致検索。タイトルの完全一致で検索したい場合は, 「タイトル完全一致」をチェック。(2012年2月改修)

# 1.3 検索

## ○ 図書館ID (参加組織番号) での検索

Cinii Books, 全文検索の公開について |

図書・雑誌検索      著者検索      [日本の論文をさがす >](#)

▼ 詳細検索

タイトル   タイトル完全一致

著者名       著者名典拠ID       出版者

ISBN       ISSN       NCID

件名       分類       注記

資料種別       言語種別        出版年  年 ~  年

**図書館ID**       機関ID       地域

すべての資料     図書     雑誌       ローマ字をカナに変換

検索

検索結果: 12,162件中 1-200を表示

1 2 3 4 5 6 ... 61 [次へ](#)

操作を選択...   全件選択

200件ずつ表示

- بحوث في تاريخ السنة المشرفة** 1  
تأليف اكرم ضياء العمري  
1967 مطبعة الارشاد  
所蔵館1館
- الاستعداد ليوم المعاد** 2  
لاحمد بن علي بن محمد بن احمد ابن حجر العسقلاني ; ضبطه واشرف على طباعته عمر الديراوي ابو حجلة

- 関連著者
- Egypt. al-Hay' 'Āmmah lil-Kitāb
  - Kuwait. al-Majli Thaqāfah wa-Ādāb
  - Suyūṭī
  - 'Imārah, Muḥa
  - Mu'assasat al-I
  - Markaz al-Imārā Buhūth al-Istīrā

# 1.3 検索

## ○ OPACリンク

 日本におけるアラビア文字資料の所蔵及び整理状況の調査

 柳谷, あゆみ

書誌事項

日本におけるアラビア文字資料の所蔵及び整理状況の調査  
柳谷あゆみ編

NIHUプログラム「イスラーム地域研究」東洋文庫拠点東洋文庫研究部イ

タイトル別名 日本におけるアラビア文字資料の所蔵及び整理状況

タイトル読み ニホン ニオケル アラビア モジ シリョウ ノ ショゾウ

詳細情報

NII書誌ID(NCID):  
BA89709099

ISBN:

国立情報学研究所図書室 蔵書検索 OPAC

[フリーワード検索](#) | [詳細検索](#) | [新着図書](#) | [新着雑誌](#) | [雑誌一覧](#) | [貸出ランキング](#) | [MyLibrary](#)  
[電子ジャーナル検索](#) | [Webcat検索](#)

大学図書館所蔵 142件

東京  すべての図書館   OPACリンク有  絞り込み

 青山学院大学 図書館 OPAC  
000902289

 桜美林大学 図書館 OPAC  
829.76/Y53 20313727

 お茶の水女子大学 附属図書館 OPAC   
014TT5//2006-2008 009010007509

 国学院大学 図書館   
1090016872

 国士舘大学 附属図書館 川図書館 OPAC   
829.76|N 77 85659

 国立情報学研究所 OPAC  
2009|655 110079824

 首都大学東京 図書情報センター OPAC  
014.7/Y53n/2009 10001643420

 昭和女子大学 図書館 OPAC 

図書詳細情報    

### 日本におけるアラビア文字資料の所蔵及び整理状況の調査 /

タイトルのヨミ: ニホン ニオケル アラビア モジ シリョウ ノ ショゾウ オヨビ セイリ ジョウキョウ ノ チョウ

出版事項: 東京: NIHUプログラム「イスラーム地域研究」東洋文庫拠点東洋文庫研究部イスラ  
2009.3

形態: ii, 66p: 挿図; 27cm

ISBN: 9784904039069 

注記: アラビア文字資料所蔵機関の所蔵状況と利用案内: p21-46

NIHU program islamic area studies TBIASの表示が表紙にあり。NIHU:人間

著者標目形: 柳谷, あゆみ || ヤナギヤ, アユミ <DA16665501>

件名: NDLSH: 特殊コレクション

NDLSH: アラビア文字

BSH: 資料整理法

## 1.4 著者名典拠

- 著者URIとしての公開
  - 生年・没年等は削除
- 書誌検索との統合

### ↓ NACSIS-CAT

<DA00151899> CRTDT:19860602 CRTFA:FA0000

HDNG: 夏目, 漱石(1867-1916) || ナツメ, ソウセキ

TYPE: p

PLACE: 牛込(東京)

DATE: 1867-1916

SF: 夏目, 金之助 || ナツメ, キンノスケ

SF: \*Natsume, Sōseki, 1867-1916

### ↓ CiNii Books

The screenshot shows the CiNii Books search results page. At the top, there are tabs for '図書・雑誌検索' and '著者検索', with the latter selected. A search bar contains '夏目 漱石' and a '検索' button. Below the search bar, the author's name '夏目, 漱石' and 'ナツメ, ソウセキ' is displayed, along with the ID 'ID:DA00151899'. A list of '別名' (aliases) follows, including '夏目, 金之助(ナツメ, キンノスケ)', 'Natsume, Sōseki', 'Natsume, Sōseki', 'Natsume, Kinnosuke', 'Нацумэ, Соэки', 'Natsumé, soseki', 'Sōseki', 'Nātsūmah, Sūsahki', 'Nacume, Sōseki', '나쓰메, 소세키', 'Nacume, Sōseki', 'Nacume, Szószeki', and 'Nacume, Sooseki'. A link for '同名同名の著者を検索' is provided. Below this, a search results summary shows '検索結果: 1,532件中 1-200を表示'. A pagination bar shows '1 2 3 4 5 6 ... 8 次へ'. At the bottom, there are two search results listed: '詩と真実' (Poetry and Reality) and '行人' (Pedestrian).

夏目, 漱石  
ナツメ, ソウセキ  
ID:DA00151899

別名 夏目, 金之助(ナツメ, キンノスケ)  
Natsume, Sōseki  
Natsume, Sōseki  
Natsume, Kinnosuke  
Нацумэ, Соэки  
Natsumé, soseki  
Sōseki  
Nātsūmah, Sūsahki  
Nacume, Sōseki  
나쓰메, 소세키  
Nacume, Sōseki  
Nacume, Szószeki  
Nacume, Sooseki

同名同名の著者を検索

検索結果: 1,532件中 1-200を表示

1 2 3 4 5 6 ... 8 次へ

操作を選択... 実行  全件選択 200件ずつ表示 出版年:新しい順 表示

詩と真実  
鶴見俊輔 [ほか] 編  
筑摩書房 2012.1 ちくま哲学の森 / 鶴見俊輔 [ほか] 編 5  
所蔵館7館

行人  
夏目漱石著

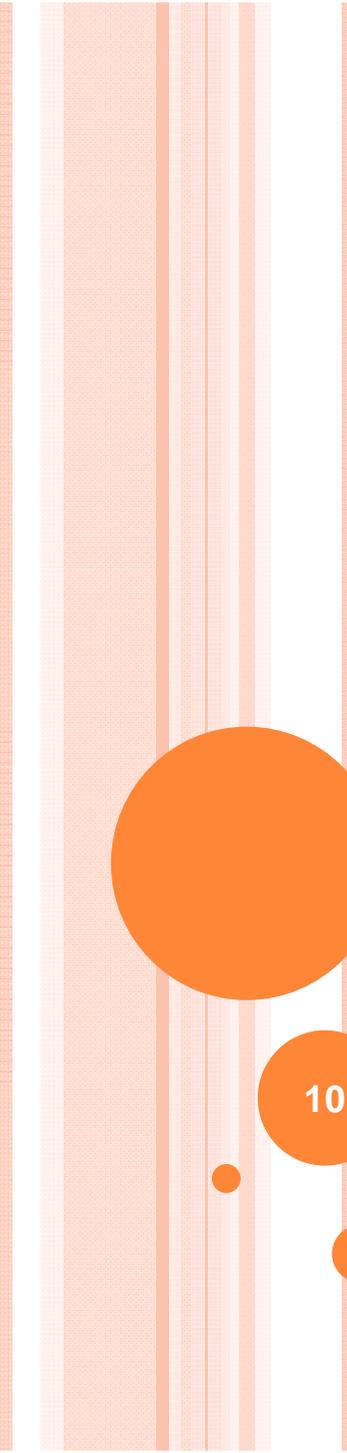
## 1.5 今後の開発

### ○ 小回りのきく開発

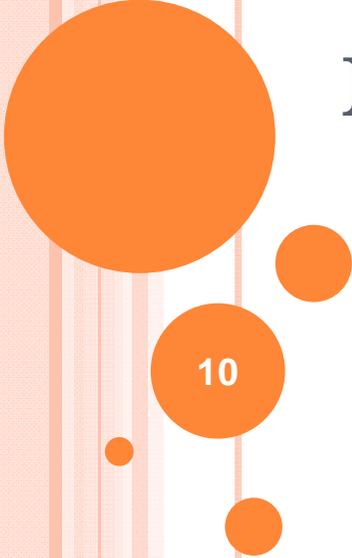
- 公開以降(約4か月間)の改修
  - 詳細検索による必須項目をなくし, 分類, 件名, 注記のみでも検索可能に
  - 図書詳細画面において, 件名, 分類からの再検索のリンクを追加
  - 所蔵館一覧を図書館名称の五十音順ソートに
  - 図書館略称と機関IDの項目追加
  - 詳細検索での分類指定で, 前方一致検索が可能に
  - IDENTのリンク表示
  - タイトルの完全一致, 前方一致検索(サブタイトルは除く)
  - VOLが複数ある場合の所蔵の表示の変更
  - ISBD形式での書き出し

等々

- フィードバック
- Twitter
- 多言語



# NACISIS-CATの正規化処理



10

## 2. 1 特殊文字・特殊言語資料に関する取扱い及び解説, コーディングマニュアル抜粋集の公開

- 2011年12月「特殊文字・特殊言語資料に関する取扱い及び解説, コーディングマニュアル抜粋集」の公開
- これまでに規則が定められた文字・言語「以外」の特殊文字・言語を対象(日本語および英語等のラテン文字以外)とする。
  - これまでに個別の規則が定められたもの＝中国語, 韓国・朝鮮語, アラビア文字, タイ文字, デーヴァナーガリー文字
  - キリル文字, ギリシア語を含む
- 内容
  - 転記の原則により, 記述対象資料に表示されているままの字体を使用
  - 英米目録規則第2版(1988年改訂, 1993年修正)を適用
  - 分かち書きはヨミフィールド(TRR, VTR, CWR)に記録
  - 翻字は, ALA-LC翻字表により, その他のヨミ(TRVR, VTVR, CWVR)フィールドに入力

## 2. 2 NACISIS-CATにおける検索

- (アラビア数字) 半角でも全角でも可
- (ラテン文字, ギリシア文字, キリル文字) 大文字でも 小文字でも可, 音標記号等の付加はなくても可

A À Á Â Ã Ä Å a à á â ã ä

→インデクスではすべてAなので, どれで検索してもよい

- (かな) ひらがなでもカタカナでも可
- (漢字) 新字体でも旧字体でも可

漢字統合インデクス

齋 齋 齋 → 齋

- (記号) 半角でも全角でも可, 長音・ダッシュ・ハイフン・マイナス記号は入力してもしなくてもよい

## 2.3 NACISIS-CATにおける正規化

### ○ これまで公開された資料

- コーディングマニュアル付録2.2 正規化処理
- NACISIS-CAT/ILLニュースレター(追加が生じた際に掲載)  
→アラビア文字については13号(2003.12.19)に掲載
- 漢字統合インデクス

★全体を一覧できる資料は公開されていなかった★

### ○ 「NACISIS-CATにおける正規化処理について」

- 特殊言語・特殊文字資料の取扱いで言及
- 平成23年度末公開予定
- 正規化の原則と具体的な一覧を公開
- 実際のシステム改修は、2012年6月～7月頃の予定

## 2.3 NACISIS-CATにおける正規化処理について

- ① 大文字と小文字のそれぞれに個別のコードが割り当てられているものは正規化する。

例: 10A0 Ⴑ Georgian Capital Letter AN

2D00 Ⴕ Georgian Small Letter AN

➡ 10D0 Ⴖ Georgian Letter AN

今回追加の対象となるのはグルジア文字, アルメニア文字, ギリシア文字(一部), キリル文字(一部), コプト文字(一部), グラゴル文字, デザート文字

※音標記号等の付加されたものは、付加されないものに揃え、  
インデクスでは大文字化する。

A À Á Â Ã Ä å à á â ã ä ➡ A

## 2.3 NACISIS-CATにおける正規化処理について

- ② 同じ形の文字に複数の文字コードが割り当てられているものは正規化する。

例:CJK Compatibility Ideographs

- ③ 複数の要素によって構成されている文字であって、個別の要素と文字全体についてそれぞれ文字コードが割り当てられているものは正規化する。

例:0BCA ᱚ Tamil Vowel Sign O

← 0BC6 ᱚ + 0BBE ᱚ

0A59 ਖ Gurmukhi Letter Khha ← 0A16 ਖ + 0A3C ੜ

- ④ 数字については、アラビア数字での検索を可能とする。

アラビア文字の数字「٠١٢٣٤.....」は表示も「01234...」

その他(タイ文字, デーヴァナーガリ文字)の数字は表示はそのまま、検索(インデクス作成)は「01234...」

## 2.3 NACISIS-CATにおける正規化処理について

### 【その他関連の問題】

- アラビア文字「アリフ(ハムザ)」および「アイン」の翻字形
- アラビア文字のカーフ「U+0643」と「U+06A9」
- チベット文字の翻字
- チベット語における分かち書き, ツェクの取扱い
- 引用符「'」「"」
- 記号「～」
- アラビア語と他言語とが混在する場合、キーワードが正しく作成されない？

等々