

目録情報に関する質問書/回答書データベース検索システム について

筑波大学図書館部

近藤 務

1 はじめに

学術情報センターが提供している「目録情報に関する質問書/回答書ベース¹」(以下「質問書データベース」と書く)は、コーディングマニュアルやニュースレター抜刷集と並んで大変強力な目録用のツールである。特に質問書データベースは、コーディングマニュアルとは違い具体例が豊富であるため、自分が悩んでいるのと同じような質問がデータベースに見つければ非常に役に立つ。問題はこの「質問がデータベースに見つければ」という点である。

自分の周囲でも、「質問書データベースは使いにくい」「欲しい情報を見つけるのが面倒」との声を聞く。そこで、このレポートではまず、現在の質問書データベースの問題点について考える。質問書データベースには、システムとしての作りと、そして内容自体にも一部問題があると思われたので、それぞれについて別々にコメントする。

レポート後半では、フリーの全文検索システムを使用して、質問書データベースのプロトタイプを作成したので、それについて簡単に説明する。結論としては、質問数が1,000~10,000のオーダーであれば、充分実用に耐えるものであると言える。と考える。

2 現在の質問書データベースの使いにくい点 - システム面

それではまず、質問書データベースのシステム的な問題点(主にユーザインターフェース)から調べていきたい。

2.1 検索結果が多数ヒットしたときに質問書を順番に見ていくのが苦痛

たとえば、「標題関連情報」をキーワードとして、質問書データベースを検索すると、86件のレコードがヒットして、質問書のタイトルが簡略表示される。

最初のレコードを表示してみよう。使用するブラウザで異なると思うが、Netscape やインターネットエクスプローラのようなGUIベースのブラウザの場合は、まずタイトルをクリックして、その後でマウスを「詳細表示」ボタンへ移動させてもう一度クリックしないと行かない。普通のWEBページのリンクをたどるように、項目の上でクリックしただけで詳細表示をしてほしい。

さらに、簡略表示のリストが10項目しか表示されず、リストの下の方を見ようと思うとスクロールさせないといけないのも煩わしい。そして、決定的に使いにくいのは、一部のブラウザでは、簡略表示リストの下の方にある質問書を表示して戻ってくると、簡略表示リストの先頭が再び表示されてしまっている。

¹ <http://136.187.232.163/cgi-bin/websql/%7Ewebsql/release/kensaku.hts>

検索された結果を通読しようとするとき、非常にストレスになる。なお、この問題はマイクロソフト社のインターネットエクスプローラではないことが確認されている。

2.2 キーワードとして1つの単語しか指定できない

複合検索ができない。例えば「タイトル関連情報」と「シリーズ名」の両方の語を含む質問書を検索できない。これは現在の質問書データベースの非常に大きな欠点である。

2.3 大文字・小文字・2バイト文字のアルファベットを区別して検索してしまう

例えば、RECON ファイルについての質問を見ようと思って小文字の“recon”をキーワードとしてデータベースを検索しても、ヒットしない。大文字で“RECON”と入力しないと行けない。

大文字小文字を区別して検索できるようにするかどうかはポリシーの問題だが、

多くのデータベースは大文字小文字どちらでも同じ結果が返ってくる。

質問書に書かれたコマンド名等は、大文字で書かれていたり小文字で書かれていたりして一定していない。

ので、質問書データベースもどちらの入力も効くのが良いと思われる。

また、現在のシステムは英文字を2バイト文字(いわゆる全角文字)で入力してもヒットしない。かな漢字変換システムの設定によっては利用者の気づかない間に2バイト文字のアルファベットを送っている可能性もあるので、検索文字中のこれらの違いはデータベース側で吸収して貰えると有難い。

2.4 検索が遅い

ただだか2,000件程度のデータベースを検索している割には、反応が遅い。検索結果が複数件あって、それを順番に見ていく時にストレスがたまる。検索のためにSQLサーバを使用しているようだが、検索・レコード表示のたびにサーバと接続しているためのオーバーヘッドがあると思われる。

2.5 機種依存文字が多用されている

これは、内容に関する項目かもしれない。丸付数字や、‘No.’、‘IV’のような文字をJIS漢字コードの空き領域に割り当てた機種依存文字がデータベース中に非常に沢山使われている。下は極端な例(センター管理番号9600168000)

書誌 <BN03496212>、 <BA04577863>及び <BA10452534>についてお尋ねいたします。

の集合書誌は重複していると考えられます。には責任表示として、“Department of International Economic and Social Affairs”が記述されていますが、の子書誌にも同じ責任表示があるものや、“At head of title”で記述されていたり、古い組織名である“Department of Economic and Social Affairs”が責任表示のものが有ります。従って、の書誌が適当だと考えられますが、NCIDがよりも大きく、また作成館も不明です。(の作成館も不明です。)また、11月1日現在には15件、には33件、には10件の子書誌がリンクしています。

MS-DOS パソコンで見ると、初めて意味が通るようになる。

書誌(1)<BN03496212>、(2)<BA04577863>及び(3)<BA10452534>についてお尋ねいたします。
(1)(2)(3)の集合書誌は重複していると考えられます。(1)には責任表示として、“Department of International Economic and Social Affairs”が記述されていますが、(2)の子書誌にも同じ責任表示があるものや、“At head of title”で記述されていたり、古い組織名である“Department of Economic and Social Affairs”が責任表示のものがああります。従って、(2)の書誌が適当だと考えられますが、NCIDが(1)よりも大きく、また作成館も不明です。(1)の作成館も不明です。)また、11月1日現在(1)には15件、(2)には33件、(3)には10件の子書誌がリンクしています。

これらの文字は、特定の機種で動くブラウザでないで見られないので使うべきではない。

2.6 関連資料として自分自身が参照されている

回答書詳細ページの下部に、「関連資料」という欄が用意されている。「表示中の質問/回答に関連する、データベース中の別の質問」へのリンクが用意される欄だと思うが、なぜか自分自身へのリンクが一つだけ書かれているものが多い。

2.7 アルファベットの語を検索したときのノイズが多い

例えば図書ファイルのVOL フィールドに‘: est’という知らない文字列があったので、これについて質問書を調べたとする。「ファイル」から「図書」を選択し、キーワードに‘est’を入力すると²、19件のレコードがヒットする。しかし、これらの中には「なぜこのレコードか?」と疑問に思うものが17件含まれている。これらは、‘best’、‘gestaltete’、‘question’等の単語中の一部分に含まれ‘est’に反応したと思われる。

アルファベットで構成された単語については、特に指定しない限り「キーワードを文字列として含む文書」ではなく「単語として分解した結果の中にキーワードが含まれる文書」を検索してくれる仕様でない、短い文字数の単語を検索する時不便だと思う。

2.8 関連するフィールドのリンクが足りない

WEB ベースの検索システムなのだから、回答に「コーディングマニュアル 2.1.14D1 にあるとおり」とあれば、WEB 版のコーディングマニュアルの該当ページへのリンクが用意されていてしかるべきである。

2.9 詳細ページから、質問書DB 検索ページへ一気に戻るリンクがない

ページ下部のリンクをたどると、いきなり目録情報課のページまで戻ってしまう。

2.10 質問書の特定のページへリンクを張る方法が提供されていない

今や多くの図書館が WEB 上に業務用のマニュアルを作成している。そのマニュアルから質問書データベースの特定の回答にリンクを張る方法がない³。

² ちなみに、キーワードとして‘: est’を指定するとこれまたヒット件数ゼロになってしまう。

³ 質問書の管理番号を与えられて、質問書データベースを検索して結果を返す CGI プログラムを書けられないこともないが、どこの図書館でもできるという訳でもない。

例えばコーディングマニュアルの場合は、

```
... VT:0H フィールドを作った場合は、コーディングマニュアルの  
<A HREF="http://www.cat.op.nacsis.ac.jp/MAN/code.vt.html">VT  
の種類</A>  
にあるように...
```

のようにHTMLで書くことができる。しかし、質問書データベースでは、「センター管理番号:9600068100“ VT フィールドのコードについて”」の質問/回答書にリンクを張る方法がない。質問書データベースに直接リンクを張る方法が提供されれば、業務用マニュアルを作成する手間が少なくなることが期待できる。

2.11 「どんな検索がされたか」のログが取られていない

一般の利用者に直接は関係ないが、大切である。なぜなら、利用者がどんな検索をしてどのファイルを読んだかという記録は、データベースを使い易くするためのアイディアの宝庫であるからである。

以下の項目は、質問書データベースの使いやすさ自体とは無関係であるが、システム側の問題であると思われるのでこのセクションに書いておく。

2.12 内部用の情報が見えてしまう

今回の研修で質問書データベースに外部用と内部用の2バージョンが存在することを初めて教えて貰ったが、内部用のデータベースがあることは一般の利用者にも知られる可能性がある。

質問書の詳細ページを表示している時に、そのソースを見ると先頭には次のコメントが入っていることがわかる。

機能：

```
質問書 / 回答書のレコードを検索し表示する。  
検索開始画面(kensaku.hts or kensaku_naibu.hts により、  
外部用、内部用表示を切り分ける。
```

特にアクセス制限をしていないようで、WEB ブラウザの URL を書換えるとそのままアクセスできてしまう。

2.13 出力されるHTML が正しい文法に従っていない

この項目もまた、質問書データベースの使い勝手自体とは関係ないが、機会があればぜひ修正することを考えて欲しい。

質問書データベースが出力するHTML は、文法的に正しいHTML であるとは言い難い。HTML の文法に従っていないWEB ページの何がいけないかと言うと、たまたまあるWEB ブラウザでそのページが正しく表示できなかった時に、「WEB ページを提供する側(つまり学術情報センター)は悪くない」と主張できない点にある。

「Netscape 社とマイクロソフトのブラウザでは一見正しく表示できるから、文法的に正しくないWEB ページでも問題はない」と主張するのも一つの意見である。しかし、この理屈は「検索さえできれば、記述文法に従わない書誌レコードを作っても構わない」というのと同じである。書誌の書き方についての2

分冊の立派なコーディングマニュアルを出しているのであれば、WEB ページも文法に従った正しいものを外に出すべきであろう。「外面よりも正しい文法」をぜひ実践してほしい。

2.14 <TITLE>や<H1>のタグが正しく使われていない

<TITLE>タグや<H1>等のタグが正しく使われていない。ここにはそのページの内容を簡潔に表現した内容が入るべきものだが、「ブラウザで見ると、大きい文字で表示されるから」というだけの理由でこれらのタグが使われているとしか思えない。

質問書の詳細表示ページでは、

<TITLE>	質問書・回答書詳細
<H1>	学術情報センター
<H2>	目録情報に関する質問書/回答書詳細

というタグの使われ方がしている。「質問書・回答書詳細」や「学術情報センター」が、ページの内容を簡潔に表現しているとは言い難い。

<TITLE>タグや<H1>等のタグに限らないが、本来の目的とは異なった使い方をすると

WEBのページがデータベースとして機能しなくなる。

質問書に関していえば、質問内容(例えば「“ 童話 ” 等は固有の標題となるか」)が<TITLE>タグに入るのが望ましいと考える。

3 現在の質問書データベースの使いにくい点 - 内容面

次に、質問書の内容自体について問題と思われる点を挙げていきたい。レポートを作成するに当たって、1996 年分の質問書データを提供していただいた。従って質問書データベース 1,700 件あまりの内容すべてを吟味したわけではないが、おおよその傾向は掴めたと思う。

答から先に書くと、質問書データベースの内容についての不満は、現在のデータベースが「紙の質問書を何も加工せずにファイルに落したまま使って、メンテナンスもされていない」ことにあると言えると思われる。

3.1 質問のタイトルの付け方が不備

この項目に該当する質問が非常に多い。そしてこれがデータベースを使いにくくしている要因の一つになっている。

まず、タイトルが質問書の内容を具体的に表していなくて、「について」という非常に大まかなものになっているものが多い。そのため、検索結果一覧ページ(タイトルしか表示されない)では、個々のページが自分の探したい情報かどうか知るためにいちいち開いてみないといけない。

タイトルは、質問/回答書の内容を簡潔かつ具体的に表現したものでないといけない。次の表は、質問書のオリジナルのタイトルと、それを修正してみたものの例である。

＝	(論外。下駄記号が2つ並んでいるだけ)
NC レコードの修正について	最終修正館がレコード修正の責任を負うことがあるか
著者名典拠のヨミの修正依頼	「望月」のヨミは「モチズキ」か「モチヅキ」か
沖縄県資料の子書誌について	沖縄県資料の子書誌(市販品と非売品)について

他にタイトルについての問題点をあげると、次のようなものがある。

「～について」で終わっているタイトルが多いが、不要なものが多い。

タイトルが途中で切れてしまっていると思われるレコードがある。

3.2 質問/回答の内容が古くなってしまっているものがある

レコードのメンテナンスがされていない。これも問題である。主なものとして、

VOL フィールドの個数を 30 個に制限していた時代の記述。

書誌ファイルの和洋統合前の記述。

親書誌の txt1 に関する記述。

新サーバテスト時のみに発生した障害の報告。

東西ドイツ統合以前の、CNTRY に関する説明。

等がある。大部分はシステムの仕様変更に伴うものである。目録の経験が短い初心者がこれらの質問書を見た場合、正しくないレコードを作成してしまうおそれがあるので、できる限り早くデータベースから削除してしまうべきである。

しかし、学術情報センターの職員だけでは、すべてのレコードを最新のものにアップデートするのは大変である。4.3.3 節(9 ページ)では、それぞれの質問ページにセンターへの連絡用のリンクを埋め込むことを提案している。

3.3 質問が出された時点での総合目録のデータを確認できない

データベースを検索した結果が「<BA03628131>は<BA00481616>と重複していないでしょうか」「ご指摘の通りです。<BA03628131>を削除します」と書いてあるだけでは、図書館 学情間の単なる事務連絡の記録でしかない。質問書データベースを事例集として使えるものにするためには、質問書が出された時点での BA03628131 や BA00481616 の書誌データが記録されていて、データベースから参照できないといけない。

3.4 一つの質問/回答書の中に、複数の質問が書かれている

図書館が1枚のファクスに2つの質問を書いたら、データベースの1レコードに2つの質問がそのまま入れられているように見受けられる。

一つのレコードに複数の質問/回答があっても、レコードのタイトル自体はどちらか片方の質問内容を要約したものになってしまう(両方を要約したものを使おうとすると、長くなり過ぎて現在のシステムでは結局役に立たない)。また、ページ自体も見にくくなる。

3.5 不要なフレーズが多い

図書館から寄せられた質問をそのままファイルにただけであるため、余計なフレーズがそのままデータベースに残っている。データベースに見られる必要のない語句とは例えば

「いつもお世話になっております」(冒頭の)さて、この度、「学情に登録した書誌(「書誌」
とだけ書けば総合目録データベースの書誌。「学情に登録した」は余分)」、「(末尾の)お忙しい所
申し訳ありませんが、よろしく願います」

のような語句である。これらは質問文に何が書いてあるのかすばやく理解する妨げとなる。テスト用のデータベースは、元ファイルからこれらの語句を取り除いて作成した。

3.6 質問書データベースに蓄積する価値のない質問

その場限りの質問で、後から見ても参考になることはないと思われるもの。例えば

質問:<BA00481616>は<BA03628131>に重複するのではないのでしょうか？

回答:確認しましたところ、ご指摘の通り、重複しておりました。<BA00481616> 統合<BA03628131>
(IDの大小の原則とは異なりますが)という調整をする予定です。

このような質問は、質問当時の書誌データの確認もできない現状では、読んでも何の参考にもならない。本当に重要な質問/回答書がこれらの無意味なレコードに埋もれて見つけ難くなるので、公開用のデータベースからは削除すべきである。

3.7 更新が遅い

現在のデータベースはかなりのブランクの後に 1998 年 5 月に最終更新されたが、更新の頻度をもっと頻繁にしていただけたらと思う。

4 質問書データベースのテストシステム

この節では、フリーの全文検索システムである `namazu` を使った質問書データベースのテストシステムを作成してみたので、それについて説明する。

全文検索システムを使用した質問書データベースは、学術情報センターでも実験中という話を聞いたので、テストシステムは個々の図書館の WEB サーバ上の FAQ(よく聞かれる質問)の管理システムとしても使えるように少しだけ考慮した。

4.1 `namazu` について

`namazu` は、愛知大学の高木哲氏が開発した全文検索システムである⁴。特徴としては次の点が挙げられる。

GPL(GNU General Public Licence = GNU 一般公有使用許諾書)のもとで、ソースプログラムが公開されている。GPL に従う限り、誰もがそのソースを利用し、改造し、再配布することができる。

日本国内では多く使われていて、安定している。動作実績のあるプラットフォームも多い。

⁴ <http://saturn.aichi-u.ac.jp/%7Eeccsatoru/Namazu/>

非常に軽快に動く。

索引の作成は perl スクリプトで行なわれる。改造が容易。

最近のバージョンでは、前方一致だけでなく中間一致・後方一致、そして正規表現を使った検索も可能になっている。

4.2 留意した点

テストシステムを作る際に、次の点に注意した。

4.2.1 質問/回答書のHTML ファイル

テストシステムでは、個々の質問/回答書をあらかじめ一つの HTML ファイルにしておいて、それに対して namazu で索引を作る形にした。

HTML ファイルは、文法的に正しいものであるように努力した。<TITLE>や<H1>タグには、「学術情報センター質問/回答書データベース」の代りにそれぞれの質問を埋め込んだ。

質問書の原稿ファイルの中にある 'sh:固有の標題' のような件名や、'field:PTBL' のような関連フィールド名はヘッダ中の '<META NAME="keywords" CONTENT="キーワードの並び">' タグに埋め込むようにした。ここに書かれたキーワードを含むファイルは、検索時に高いスコア(評価値)が得られる。HTML ファイル作成時に、機種依存文字は、どの端末でも表示できる文字列に変換する。また、機種依存文字ではないが、2バイト文字の '[' ']' は1バイト文字の '[' と ']' に機械的に置き換えた⁵。質問文の前後を '<!-- QUESTION -->' と '<!-- /QUESTION -->' で、回答文の前後を '<!-- ANSWER -->' と '<!-- /ANSWER -->' で囲んだ。これらは HTML のコメントであるため、ブラウザで見る限りはタグを入れても入れなくても何の変化もない。これらのタグは、4.2.3 で後述される「質問/回答書の内容にコメントを付ける」で使用される。

回答文の後には同じくコメントで '<!-- NOINDEX -->' を挿入した。namazu に手を入れて、このコメント以降に書かれている参照リンク等の文字列からは索引を作らないようにしてある。

4.2.2 ブラウザで表示して目で見える変更

<TABLE>タグは、使用しない。表形式にすることが必須ではないし、使用するブラウザによっては見にくくなるため。

関連項目へのリンクの充実。各ページに、他のマニュアル類へのリンクを追加した。本来なら文脈に合うページへのリンクを張るべきであるが、量が膨大であるので機械的に各マニュアルのトップページへリンクを張ることしかできなかった。

- 文中に「抜刷集」または「目録情報の基準」の語があれば、それらの目次ページへのリンクを張る。
- 文中に「コーディングマニュアル」の語があれば、まずコーディングマニュアルの目次ページへのリンクを張る。さらに、「コーディングマニュアル」の直後にセクション番号と思われる文字列が続いた場合は、そのセクションへのリンクを張る。実際にそのページが存在するかどうかのチェックはしていないので、存在しないページへのリンクが張られてしまう可能性ももち

⁵ 実は機械的に置き換えるのは危険である。「データ中に '[' ']' という括弧が使われていますが、これは '[' ']' と書くべきではないでしょうか?」「お説ごもっともです」という質問書があった場合、読んで意味の分からないものになってしまう。

ろんある。

- ソースファイルに ' field:OTHN' のようなフィールドがあった場合は、その値をキーとしてデータベースを検索するための検索式が埋め込まれる。
- ソースファイルに参照用画像ファイルフィールドがあった場合は、その画像へのリンクが埋め込まれる。
- ソースファイルに ref:(関連情報)フィールドがあれば、その質問へのリンクを張る。ただし、フィールドが自分自身を指していたら何もしない。

各ページから、質問書データベース・コーディングマニュアル等を検索するための入力フィールドを追加。

さらにページ最下部に、利用者がコメントを簡単に付けられるようにするためのリンクを追加。これについては次項で詳しく説明する。

詳細表示画面の例を付録A(12ページ)に示す。

4.2.3 質問/回答書の内容にコメントを付ける

1,000以上のすべての項目を、少ない人数で常に最新のものに維持するのはほとんど不可能である。データベースの利用者自身に内容が古くなったレコード・記述があいまいなレコードを見つけてもらうの手伝ってもらうのが一番である。

簡単な方法は、それぞれのページに連絡先メールアドレスを書いておく方法がある。しかし、この方法だと、メールに利用者自身が次の情報を入力しないとイケない。

質問書の管理番号。

質問書のどこに問題があるのか。

もしあれば、修正案。

質問書の管理番号は最初から分かっているし、質問書の問題の典型的なパターンは、このレポートの3節で挙げられている。ならばあらかじめこれらの項目を、用意してやれば良い。

このような考えで、質問/回答書データベースの内容にコメントをつける仕組を用意した。「レコードの内容に問題あり」と考えた利用者は、ページの一番下の「この質問書の内容についてのコメントを送る」のリンクを選択すれば良い。このリンクは、質問書の管理番号を引数とする CGI スクリプトが呼ばれるようになっている。スクリプトは、質問書の HTML ファイルを読んで、' <!-- QUESTION -->' と ' <!-- /QUESTION -->', ' <!-- ANSWER -->' と ' <!-- /ANSWER -->' で囲まれた部分を抜き出し、適当な幅で改行した上で⁶、引用記号 '>' を付けて入力フィールドに初期値としてセットする。

付録B(13ページ)のようなページが表示される。利用者はこのページの内容を適当に編集する。送信ボタンを選択すると、ページの内容が設定しておいたメールアドレスに送信されると共に、ログファイルに追加書き込みされる。

4.3 索引の作成時間

質問書データベースの索引を作成する時間を計測した。

質問書とは無関係だが、目録情報の基準(第3版)、コーディングマニュアル、そしてオンラインシステム・ニュースレター抜刷集についても namazu で索引を作ってみた。

⁶ 正確には出力される文字幅でなく単に文字の個数で改行しているので、英単語が多く含まれる質問/回答だと引用されたものも見苦しくなる。

	ファイル数	HTML ファイル 合計	namazu の 索引ファイル	インデックス 作成時間
質問書データベース	1,720	4.6 メガバイト	5.9 メガバイト	770 秒
目録情報の基準	96	420 キロバイト	1.0 メガバイト	50 秒
コーディングマニュアル	338	1.3 メガバイト	2.0 メガバイト	180 秒
ニュースレター抜刷集	200	880 キロバイト	1.7 メガバイト	120 秒

インデックス作成時間は、K6-266MHz の CPU を持つ FreeBSD マシン(メモリ 64 メガバイト)でインデックスを作成した時の、1 秒の桁を切り上げたものである。

4.4 テスト用データベースの評価

索引の作成時間は十分短いので、毎晩データを WWW サーバから取得 索引作成しても何ら問題ないと思われる。

検索時間は、検索式を入力して 2 秒以内に結果が返ってくる。現在の質問書データベースと比べて非常に高速である。

ちなみに、検索された結果の最初の 25 件を詳細表示するのに、テスト用データベースでは約 2 分半、現在の質問書データベースでは約 4 分かかった。

抜刷集については、<TITLE>タグの内容がすべて同じなので、検索結果の要約表示をオフにすると

検索式にマッチする 8 個の文書が見つかりました。

1. オンライン・システムニュースレター抜粋集 (score: 5)
2. オンライン・システムニュースレター抜粋集 (score: 4)
3. オンライン・システムニュースレター抜粋集 (score: 2)

と文書の中を覗かないと何が書いてあるか分からない結果になってしまう⁷。

4.5 問題点と今後の課題

テスト用データベースでは、現在の質問書データベースにある「件名」「フィールド」「ファイル」の選択ができなくなっている。「件名」と「フィールド」については、個々の質問書の HTML ファイルに '`<META NAME="keywords" CONTENT="キーワードの並び">`' の形で埋め込んであるので、検索式入力欄に件名またはフィールド名を入れれば検索結果の先頭部分に来るようにはしてある。

シソーラスの導入。

5 まとめ

現在の質問書の規模であれば、フリーの全文検索システムを使って充分使用に耐えるものができる。問題はどの検索システムを使うにしても、データ自体のメンテナンスである。

⁷ ところで、「抜刷集」が「抜粋集」と書かれている。多分直した方が良いと思う。

センタへの要望

文法的に正しいHTML 文書。

すべての情報は、HTML でリンクが張れるように(特定のプログラムを通さないとアクセスできないページを作らない)。

コーディングマニュアル・抜刷集についても全文検索機能の提供を。

質問書データベースの内容はスリムに。

データのメンテナンスは、日本中の利用者に手伝ってもらう。

謝 辞

最後になりましたが、3 週間の長い期間に渡って我々の研修の面倒を見ていただいた学術情報センターの皆様へ、心からお礼を申し上げます。特に研修課の皆様には非常にお世話になりました。本当にありがとうございました。

A 質問書詳細表示

The screenshot shows a web page with a navigation bar at the top containing links: 'What's New?', 'Who's Cool?', 'Desktops', 'Net Search', 'People', and 'Software'. The main content area has a title '標題関連情報の相違は別書誌作成の根拠となるか' and a sub-header '《図書》'. The text explains that differences in edition numbers lead to separate book entries and asks for guidance. Below this is a '【回答】' section stating that according to the 'コーディングマニュアル0.4.1', differences in titles alone are not sufficient for separate entries; content, page count, and size are also considered. It notes that if these differ, a note should be added to the existing entry, and titles not used should be recorded as 'VT'. A '【参照】' section lists 'コーディングマニュアル' (with sub-link '0.4.1') and 'フィールド' (with sub-links 'TR' and 'VT'). At the bottom, there is a search bar with a dropdown menu set to '質問書DB' and a 'Search' button. A link 'この質問書の内容についてのコメントを送る' is also present. Footer information includes 'センター管理番号: 9600088900', 'センター回答日: 1996/07/15', and links for 'Q&A検索ページへ' and '目録担当者のページへ'.

What's New? Who's Cool? Desktops Net Search People Software

標題関連情報の相違は別書誌作成の根拠となるか

《図書》

刷の違いにより、副書名の記述が異なっている場合、別書誌を作成してよろしいのかご教示願います。

【回答】

コーディングマニュアル0.4.1のとおり、標題関連情報の相違のみでは別書誌作成の根拠とはなりません。内容、ページ数、大きさ等の要素に違いがないのであれば、既存の書誌に8刷に関する注記を付し、採用しなかった標題はVTに記録するのが適当であると思います。

【参照】

- [コーディングマニュアル](#)
 - [0.4.1](#)
(個別セクションへのリンクは正しくないことがあります)
- フィールド
 - [TR](#)
 - [VT](#)

を含むページを [質問書DB](#) から検索

[この質問書の内容についてのコメントを送る](#)

センター管理番号: 9600088900 センター回答日: 1996/07/15
[Q&A検索ページへ](#) [目録担当者のページへ](#)

B データベースの内容についてのコメントを送るページ

What's New? What's Cool? Destinations Hot Search People Software

管理番号 '9400121600' の質問へのコメント

タイトルが内容を具体的に表していない
 内容が古く現状に即していない
 一つの質問書に2つ以上の質問があり紛らわしい
 質問が意味不明
 回答が意味不明
 回答が間違っている

タイトル・質問・回答を「このように直したらわかりやすくなる」という案があったらお願いします

タイトル:

質問文:

回答:

C 今回使用したツール

テスト用データベースを作るにあたり、幾つかのツールを利用した。

namazu	全文検索システム。バージョン 1.3.0.0beta12 を使用。
kakasi	日本語の文章を単語に分割するために namazu から使用される。フリーソフト。
perl	日本語化された perl5.004 を使用した。質問書の HTML ファイルを作るスクリプトや、CGI スクリプトを実行するために必要。
Kconv	perl5 のモジュール。多くのスクリプト中で、日本語コード変換をするために使用。
libwww-perl	perl5 のモジュール。95 年度以前の質問書をローカルに持ってくるスクリプト (getqa) で使用。
GNU wget	コーディングマニュアル・抜刷集・目録情報の基準をローカルディスクに持つるために使用。
makehtml	perl スクリプト。学情が持っている質問書/回答書の元ファイルから、個々の回答書の HTML ファイルを作る。
getqa	perl スクリプト。質問書データベースの簡略表示ページの HTML ファイルから、学情管理番号を抜き出して、その管理番号に対応する質問/回答書をサーバから貰ってローカルにその管理番号をファイル名とするファイルを作る。
qa-makecomment.cgi	perl スクリプト。パラメータとして質問書の管理番号を受けとり、その管理番号の HTML ファイルを読み込んでコメントを送るための HTML を出力する CGI プログラム。
qa-sendcomment.cgi	perl スクリプト。qa-makecomment.cgi で作ったページから呼ばれる CGI スクリプト。コメントをメールで指定したメールアドレスに送信する。
db2src	perl スクリプト。getqa で持ってきた個々の回答書ファイル(HTML ファイル)を makehtml で処理できる形式に変換する。

makehtml 以下のスクリプトは、自作した。

C.1 mknmz の変更

namazu の索引作成に使用する mknmz だが、配布されているままだと要約表示をしたときの表示が気に入らないので、次の修正を加えた。

要約文の先頭に、タイトルと同じ文字があれば表示しない。

要約を表示すると URL も表示する指定になっているが、これを表示しないようにする。

索引を作る時に、<!-- NOINDEX -->と<!-- /NOINDEX -->で囲まれた部分からは索引を作らない。

```

--- /usr/local/bin/mknmz Thu Oct 29 09:09:15 1998
+++ mknmz Wed Nov 4 13:22:58 1998
@@ -686,6 +686,7 @@
sub html_filter ($$$$$) {
    my ($contents, $weighted_str, $title, $fields, $headings) = @_;

+   erase_noindex_part($contents);
+   escape_lt_gt($contents);
+   get_html_title($contents, $weighted_str, $title);
+   get_author($contents, $fields);
@@ -797,6 +798,12 @@
    $$contents =~ s/^[¥s¥S]*<BODY>///i;
}

+# <!-- NOINDEX --> と <!-- /NOINDEX -->で括られた部分を削除する
+sub erase_noindex_part ($) {
+ my ($contents) = @_;
+ $$contents =~ s/<!-- NOINDEX -->[¥S¥s]*<!-- ¥/NOINDEX -->///g;
+# $$contents =~ s/<!-- NOINDEX -->[^(<!-- ¥/NOINDEX -->)]*<!-- ¥/NOINDEX -->///g;
+}

# %TAGW に設定されている数値に応じて ¥tXX¥t, ¥t/XX¥t という架空のタグ
# を作り、これで挟んでおく (¥t は予めすべて空白に変換してある)
@@ -2064,16 +2071,18 @@
    my ($url, $title, $file_size, $contents, $headings, $file, $fields) = @_;
    my ($summary);

-   $summary = make_summary($contents, $headings, $file, $fields);
+   $title =~ s/¥s+//g;
+   $title =~ s/¥s+//;
+   $title =~ s/¥s+$///;
+   if ($$contents =~ /^¥s*$title/){
+     $$contents = $';
+   }
+   $summary = make_summary($contents, $headings, $file, $fields);
+   $file_size = commas($file_size);

# FLIST へ書き出し <DT> の後に改行が欲しいのです
print FLIST "<DT>¥n<STRONG><A HREF=¥"$url¥">$title</A></STRONG>¥n";
- print FLIST "<DD>$summary¥n";
- print FLIST "<DD><A HREF=¥"$url¥">$url</A> size ($file_size bytes)<BR><BR>¥n";
+ print FLIST "<DD>$summary<BR><BR>¥n";
# 最後に空行を入れる(これは重要な仕様)
print FLIST "¥n";
}

```