

OPAC 評価の実際（若干のまとめ）

渡邊隆弘（帝塚山学院大学）

OPACに求められる機能

検索の目的と目録の機能

- ・ファインディングリストとしての目録
 - 特定の資料がすばやく探せる
 - あいまいな情報からでも探せる
- ・「未知の資料」を検索するための目録
 - 特定キーワード（主題や著者）に合致する資料を「集中」
 - ブラウジング機能も重要

データベースの基本的な機能

かつては難しかったことが、Web-OPAC では当たり前を実現

- ・空間的制約の打破（どこからでもアクセス）
- ・特殊な機器・環境は不要（PC さえあれば）
- ・時間的制約の打破（24 時間稼働）

OPAC で全蔵書をカバー

- ・都道府県立レベルでも遡及入力は急速に進展
（大学図書館ではまだ大きな問題）
- ・今後の課題は多文化サービスか

データの更新頻度

- ・即時更新（「貸出中」等のステータス情報も含めて）が当たり前

検索の応答速度

- ・データベース技術の進歩
 - ヒット件数の多寡に関わらず、高速に検索集合作成
（ただし、一覧表示の応答速度は表示件数に依存）
 - ヒット件数と「最初の 10 件」なら高速に提供可能のはず
- ・インターネット検索エンジンの応答速度
 - 遅いシステムは、従来以上に「目立つ」
 - ヒット件数に比例して遅くなるシステムは問題

基本機能はかなりの割合でクリア

検索・表示そのものが問題に

検索語入力画面のインターフェース

検索対象の限定

- ・ 図書、雑誌、AV、その他
- ・ 図書検索と雑誌検索
 - 分けて扱う傾向が強かった 最近はそうでもない
 - 統合的に検索できて当然
 - (年鑑、白書、統計... 自明に分割できるものではない)
 - デフォルトは「全資料対象」であるべき
- ・ AV、電子資料等をどこまで細かく媒体区分するか？
- ・ 地域資料等の取扱い
 - 特別な資料群はそれのみに限定して検索可能とするべき

検索語入力フィールドの設計

A. 対象項目ごとに枠設定

「何から検索できるか」がマウス操作なしにわかる
画面の肥大化、利用者への圧迫感

B. 対象項目をプルダウンメニューに(ふつうは複数枠を設定)

入力枠を減らしてコンパクトに
内容の同じプルダウンが複数あるわかりにくさ

C. 簡易検索では全項目対象とし、枠は一つ

ひたすらシンプル

- ・ 大学図書館ではC.が主流
 - 空白で区切ると論理積(AND)
 - 検索エンジンとの親和性
 - 図書館トップページに置くのも一般的
- ・ 公共と大学の違い？
 - 「何から検索できるか」を利用者に伝える？
 - ヘルプの役割では？
- ・ 詳細検索画面
 - 項目ごとの枠設定かプルダウンか？
 - すべての対象項目を画面に示す必要はない
 - 限定して意味のある項目 + 「全項目」
 - 例えば「注記」・・・注記だけに限定して意味があるか？

キーワード検索の諸相

かつては、書誌記述とは別に「検索キーワード」を別途入力して検索に供していた
数に制限、長さに制限、使える文字種に制限...

その後、書誌記述中の検索対象フィールドからキーワード自動切り出し
空白等を区切りとみなして切り出し

分かち書きされたヨミからは単語ごとに切り出されるが、

分かちのない表記形からは全体形しか切り出されない

今では、検索対象フィールドに対して、「全文検索」手法

「タイトル」： タイトル中のどこかにその語があればひっかかる

全文検索の2方式

・形態素解析法

辞書と照らし合わせて単語に分割し、キーワードとして格納

例)「図書館の学校」 「図書館」「の」「学校」

(「の」は無意味単語として無視することも)

長所：「意味」に沿った手法

短所：分かち書きを利用者に強制

最近は複合語でも探せるよう処理することが多い

例)「画像情報処理」 「画像」「情報」「処理」

「画像」「情報」「処理」「画像情報」「情報処理」「画像情報処理」

システムの辞書に依存

(固有名詞や古い当て字などは間違える可能性)

・N - グラム法

決まった字数に分割し、順次キーワードとして格納

例)「図書館の学校」 「図書」「書館」「館の」「の学」「学校」

利用者の入力した検索語にも同様の分割を行う

単純かつ完全な「中間一致検索」が実現

長所：機械的(間違いの発生する余地がない)

短所：機械的(「京都」で検索すると「東京都」もヒット)

・一長一短があるが、N-グラム方式が主流になりつつある。

キーワードの「正規化」

・意味のない違いを意識せず検索できるように

カタカナとひらがな、大文字と小文字、記号類の無視、旧漢字と新漢字

・検索時に(のみ)対処するのは非効率

Cat or CAT or cat or cAT or CaT or ...

索引時(データベース格納時)に、どちらかに強制変換

- ・カタカナ・ひらがな、大文字・小文字、記号類は、できて当たり前
清音・濁音 やらないほうがよいのでは（ノイズが多すぎる）
長音記号 難しい問題
異体字 やるべきだが、文字種も多く関係複雑
例）NACSIS WebCat
日本語・中国語を視野に入れて「漢字統合インデクス」
- ・表示用データと検索用データの分離が前提
- ・インターネット検索エンジン（正規化はほとんどされていない）

高度な検索機能

主題検索のサポート

- ・同義語辞書
辞書の信頼性が鍵 多くは企業秘密でブラックボックス
更新できるべき（地域特有の語など）
- ・分類表の利用
広まっているが、十分使えるものになっているかは疑問
視覚的アクセス、分類表の名辞と索引語、...
- ・件名標目表の利用
あまり広まっていない

内容情報の増強

- ・内容細目、内容紹介、帯情報等
急速に広まり、検索対象フィールドにもなっている
検索精度への影響はどうか

表示機能とナビゲーション

一覧表示とソート

- ・一覧表示の限られたスペースに何を出すのか
表形式か連結式か
「最初の著者だけ」仕様がけっこうある 「ほか」と出したほうが...
- ・ソート機能は案外充実（大学図書館よりも？）
検索エンジンのような「ランキング表示」の可能性
大規模蔵書の場合、内容情報が充実してきた場合

書誌情報の詳細表示

- ・必要最小限の項目 vs. 全部出す
絞り込みすぎの感のあるシステムも
ブラックボックス（なぜヒットしたのかわからない） 無用の不安感

- ・ハイライト表示
広まっていないが、有用性は非常に大きい
インターネット検索エンジンと対比して

ナビゲーションのための配慮

- ・画面の統一性
- ・そこに至る経緯を表示すべき
何で検索してそこへ到達したのか
- ・検索履歴の再利用（どれだけ有効性があるか）
- ・ハイパーリンクの有効活用
著者や件名をクリックすると当該の一覧表示に
だいぶ広まってきた
典拠コントロールとの関係
著者名典拠がどこまで管理されているか

ヘルプ機能

絶対に、司書が書くべき

メーカーに「わかりやすいヘルプ」を要求しても無理
随時書き換えられる仕様を要求

新しいシステムに対して、動く前に完璧なヘルプは書けない

用語の使いかた・説明のしかた

本当に難しい・・・

多少長くなっても、例示を入れるべきでは

ヘルプの単位

ピンポイントで示せ、かつ通読もできるのが望ましい