

情報検索演習 第3回

2007年10月10日

後期 水曜5限

江草由佳

国立教育政策研究所

yuka@nier.go.jp

<http://momiji.mimoza.jp/lecture/2007/ir/>

講義Webサイトをお気に入りに入れておくこと！

前回のまとめ

- 講義内容
 - レコードと検索フィールド
 - 情報検索の理論
 - 論理演算子(論理積、論理和、論理差)
 - トランケーション(前方一致、後方一致、中間任意、中間一致)
 - 情報検索結果の評価
 - 検索漏れ、ノイズ
 - 再現率
 - 精度
 - 第1回演習課題: 論理演算とトランケーション

本日のお品書き

- お知らせ
- 第1回演習課題について:返却・コメント
- 講義
 - データベースの起源、発展、定義、分類、流通サイクル
 - 商用データベース
- 第1回レポート課題の出題
- CD-ROM検索
 1. 検索プログラムのインストール
 2. 検索の大まかな流れ
 3. 保存・印刷
- 第2回演習課題:CD-ROM検索・印刷の出題・提出₃

お知らせ

- 演習課題 → 提出していない人は提出すること
 - 第1回演習課題: 第1回目の講義に出題・提出
- パソコンについて
 - ファイルのコピー、ファイルの名前変更、ファイルの削除、ファイルの移動くらいはすらすらできるようになっておこう
 - 課題にはWord, Excel, PowerPointを使う予定
- 今日の配布資料:
 - 「CD-ROM版情報検索の演習 新訂版」
2006年 Tomoko Harada

第1回演習課題について

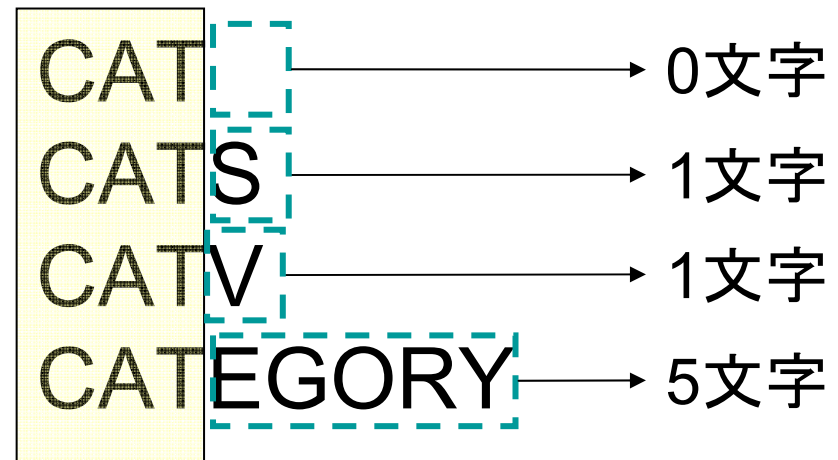
- よくできていました
- 評価: 全問正解: A, チョイミス: A⁻, 数問間違い: B, 大幅間違い: C
- 問題4: 多くの人が不正解

第1回演習課題：問題4の解説

- 問題4: CAT, CATS, CATV, CATEGORYのうち、CAT, CATS, CATVのみがヒットするような検索

- 基本的な考え方:

- トランケーションの
- 前方一致を使う

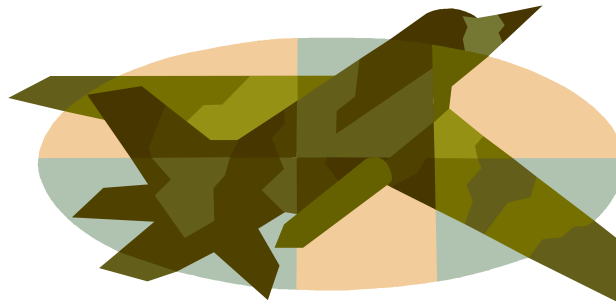


講義目次

- データベースの起源
- データベースの発展
- データベースの定義
- データベースの分類
- データベースの流通サイクル
- 商用データベース

データベースの起源 -27p.

- 1950年代
- 米国国防総省が戦力に関する**情報を保管、集中管理**するためコンピュータを使ったライブラリーを開発
- **データの基地**(data base)から由来



データベースの発展 --27p.

- リファレンス(文献)データベースの起源
 - 1950年代半ば頃:二次資料作成がコンピュータ化
 - 副産物としてデータが残る
 - 米国とソ連の宇宙開発競争 → NASA誕生
 - 後にDIALOG→1970年代はじめにオンライン情報検索サービスが開始
- 日本:
 - 1970年に日本経済新聞社による株価などの磁気テープ販売
 - オンラインデータベース:JOIS(現在JDreamII)登場
 - 1985年CD-ROMデータベース登場
 - 1990年代半ば以降:インターネット、Webの普及

データベースの定義(1)

- 著作権法二条十の三 –p.28
 - 論文、数値、図形その他の情報の集合物であって、それらの情報を電子計算機を用いて検索することができるように体系的に構成したもの
- 日本工業規格(JIS) –p.28
 - 適用業務分野で使用するデータの集まりであって、データの特性和それに対応する実態の間の関係とを記述した概念的な構造によって編成されたもの(X0017)
 - 特定の規則に従って電子的な形式で、一か所に蓄積されたデータの集合であって、コンピュータでアクセス可能なもの(X0807)

データベースの定義(2) -28p. --日本のデータベースの特徴--

- データベースとは“コンピュータを用いて検索できる”ことが重要である。情報が電子メディアに蓄積され、コンピュータ、携帯情報端末(PDA)、地上波テレビ端末などを使用して検索できる状態になっている。
- データや情報がコンピュータ処理できるように体系的に整理され、統合化・構造化されて蓄積・保存されており、必要な情報だけを部分的に取り出せる。
- 蓄積情報の検索や更新が容易に行えるよう、効率化を図ったものである

一方、ヨーロッパにおけるデータベースの定義では、コンピュータを使用するかしないか、電子的であるかどうかについては特に限定していない

データベースの分類 (1)

p.30 2-1図 参照

- p.30 2-1図 データベースの分類

1)データの形態別分類-p.30

- データの形態に注目して分類
- 2-1図:データの形態別分類を参照

2)データの分野別分類 -p.30

- データベースを扱っている主題分野によって分類
- 一般、自然科学・技術、社会・人文科学、ビジネス、その他

データベースの分類 (2)

p.30 2-1図 参照

3)提供形態別分類 –30p.

– オンライン

- 遠隔地にあるWWW上のデータベースを手元のコンピュータからネットワークを介して利用するもの

– オフライン(スタンドアロン)

- 手元のコンピュータのみで利用できるもの。CD-ROMやDVDなどで提供されるパッケージ型のデータベースなど。

データベースの分類 (3)

p.30 2-1図 参照

4)用途別分類 —p.30

- データベースの用途に応じて分類
- 商用データベース
 - 誰でも利用できるが、課金される
- インハウス・データベース
 - 機関内で構築され、利用に制限があるもの
 - 社内、業界、学術にさらに分類される
- パーソナルデータベース
 - 個人が構築し、個人が利用するもの
- オープン
 - インターネット上に提供され、一部を除いて無料で広く一般に開放されている

データベースの分類(4)

図にはない

5)更新頻度別分類—31p.

- データベースの更新頻度で分類
- イミディエート
 - 時々刻々と変化更新されるもの
 - 株価情報や気象情報などを扱ったデータベース
- ヒストリカル
 - 一か月、3ヶ月などある程度まとまって更新されるもの
 - 書誌データベースなど

データベースの分類(5)

図にはない

6)利用料金別分類 -31p.

- 利用料金の有無により分類
- 有料
 - 商用のデータベース、契約・ID必要
- 無料
 - 商用のデータベースであっても、図書館が一括して契約し、利用者自身には料金がかからないものはこちらにあてはまる
 - インハウスデータベースもこれに該当する場合がある
 - 機関内であれば無料だが、他者が利用する場合は課金されるなど
 - 無料のデータベース

データベースの分類(6)

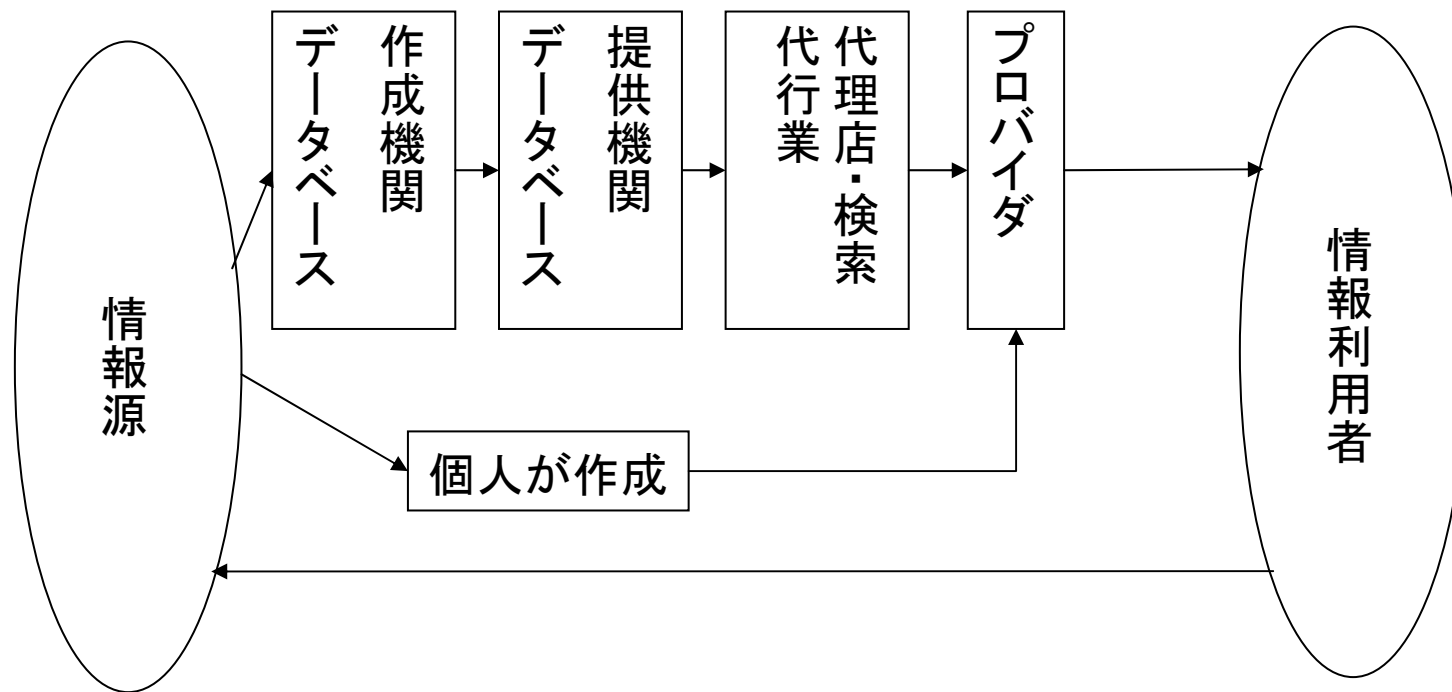
まとめ

p.30 2-1図 参照

- データベースの分類
 - データの形態別分類
 - データの分野別分類
 - データの提供形態別分類
 - 用途別分類
 - 更新頻度別分類
 - 利用料金別分類

データベースの流通サイクル(1)

--p.32 2-2図--



データベースの流通サイクル(2)

P.32 2-2図

- データベース作成機関 —32p.
 - データベース作成者(機関)のこと
 - 文献データベースの場合は情報源収集から検索語を付与し蓄積するまでを行う

データベースの流通サイクル(3)

P.32 2-2図

- データベース提供機関 –32p.
 - データベース作成機関から委託されたデータを、自社のコンピュータによってサービスを提供する機関
 - データベース作成機関と提供機関を兼ねることもある
 - 日本ではこのタイプが多い
 - 例：科学技術振興機構（JST）、日本経済新聞社、帝国データバンク

データベースの流通サイクル(4)

- プロバイダー—32p.
 - インターネット接続業者：インターネット・プロバイダ（単にプロバイダとも呼ぶ）
 - Webでの検索では、プロバイダと契約してインターネット接続できるようにする

データベースの流通サイクル(5)

P.32 2-2図

- 代理店 (agent) –32p.
 - 有料の商用データベースを利用する際に契約
 - データベース作成機関の著作権事務の代行
 - データベース提供機関の営業事務の代理を行う
 - 具体的には
 - 情報検索の講習会
 - 新規データベースの案内
 - 検索マニュアルの日本語化
 - 例: 紀伊国屋書店、丸善、ジー・サーチ、化学情報協会

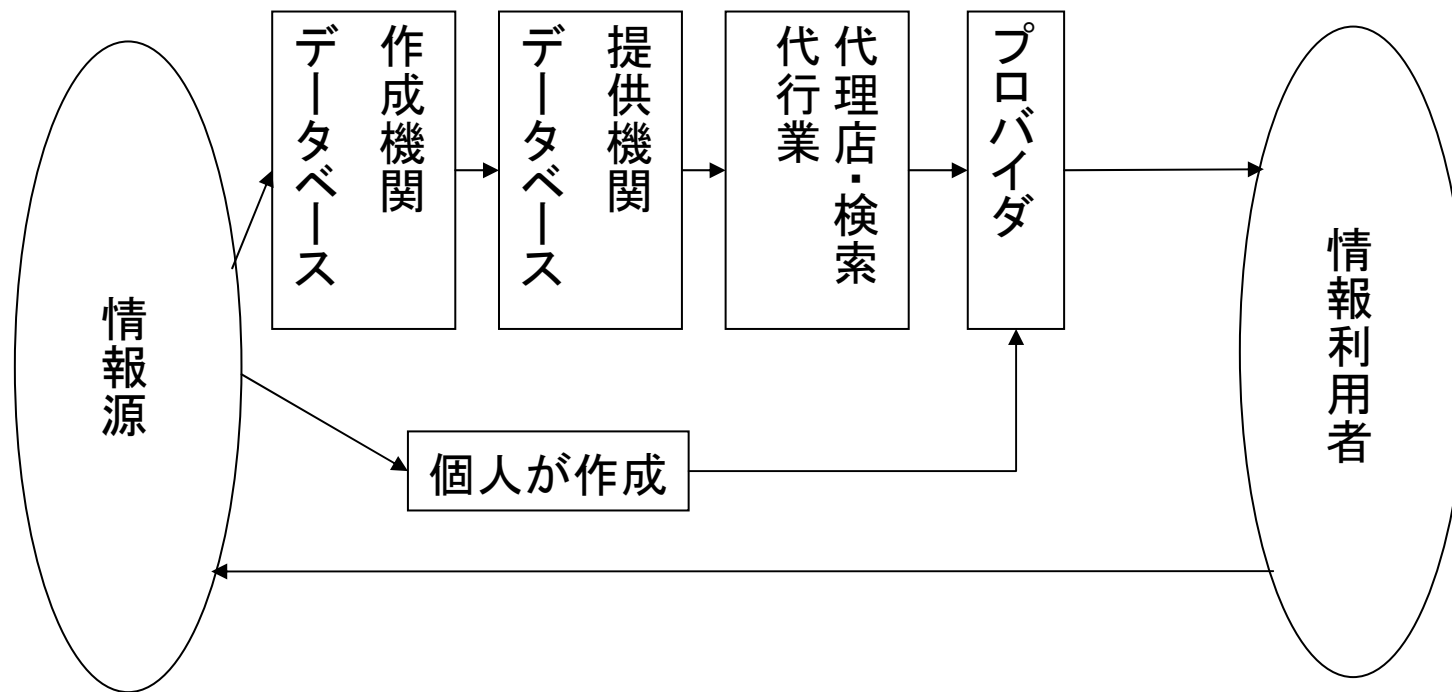
データベースの流通サイクル(6)

P.32 2-2図

- 検索代行業者(information broker) –33p.
 - 利用者に代わってデータベースを検索し、結果を利用者に提供する機関
 - 形態はさまざま
 - データベース販売代理店、公共図書館のレファレンスサービスなど
 - 企業や情報センターなどではサーチャー(検索技術者)が検索を行い、検索結果の評価、分析、報告書作成などを代行する

データベースの流通サイクル(7) まとめ

P.32 2-2図



商用データベース –33p.

- まずはじめに、Webで無料で検索することが多い
- 無料で利用可能なWeb資源には量と質が問題になることもある
- 商用データベースを利用する必要もある
 - p.34 2-1表:おもな国内外のサービス提供機関とサービス内容
 - 契約・利用方法や最新情報を入手できる
 - p.35 2-2表:おもな商用データベース
 - 司書が知っておいたほうがよい商用データベース

講義のまとめ

- データベースの起源
 - 米軍のコンピュータシステム
- データベースの発展
 - 二次資料作成のコンピュータ化の副産物、NASA
- データベースの定義
 - 著作権法、JIS、コンピュータを使う
- データベースの分類 p.30 2-1図 参照
 - 形態別、提供形態別、分野別、用途別、更新頻度別、利用料金別
- データベースの流通サイクル P.32 2-2図
 - データベース作成機関、データベース提供機関、プロバイダ、代理店、検索代行業
- 商用データベース

第1回レポート課題

- 今日までの講義をA4用紙1ページ分にまとめなさい
 - ヒント:スライドの「今日のまとめ」、目次、章タイトル
- ✕切:次回の演習の最初に集めます
- 以下をレポートの冒頭に必ずつけること
 - タイトル:第1回レポート課題:情報検索演習の講義の要約
 - 講義名:情報検索演習
 - 出題日:2007-10-10(第3回)
 - 学籍番号
 - 氏名
 - 提出日

CD-ROM版情報検索の演習

1. 検索プログラムのインストール: 資料p.1~
 - 次回以降も毎回する必要がある
 - ↑再起動するとアンインストールされてしまうため
2. 検索の大まかな流れ: 資料p.8
3. 印刷(保存): 資料p.4
 1. 検索画面の保存方法
→ <http://momiji.mimoza.jp/lecture/2007/QA/#screen-shot>
4. 検索の基本
 1. 論理検索(検索フィールド内/間)
 2. トランケーション
 3. 範囲検索

第2回演習課題： CD-ROM検索・印刷

1. 検索プログラムのインストール
2. データベースを検索
 1. データベースを選択：**雑誌記事情報**
 2. 検索実行(なんでもよい) **ただし2件以上ヒットするもの**
例) 論題中の単語:「印象」、出版年月日:「2002*」
 3. **2件目**の検索結果の表示
 4. Wordに保存(ファイル名:**kadai02.doc**)
 - ヘッダに「**第2回演習課題: CD-ROM検索・印刷**」「**情報検索演習**」「**学籍番号**」「**氏名**」「**提出日**」をつけること
 - 必ず**1ページ**にまとめること!
 - **検索画面、一覧表示、詳細表示**を含めること(1ページに収まるように適度に省略すること)
5. 印刷

本日のまとめ

- 講義
 - データベースの起源、発展、定義、分類、流通サイクル
 - 商用データベース
- 第1回レポート課題出題、来週提出すること
- CD-ROM検索
 1. 検索プログラムのインストール
 2. 検索の大まかな流れ
 3. 保存・印刷
 4. 検索の基本
- 第2回演習課題：CD-ROM検索・印刷の出題・提出