

情報検索演習

第1回:オリエンテーション

- オリエンテーション
 - 授業の目的・内容
 - 評価方法・指導方法
 - テキスト・参考書
 - USBメモリ
 - スケジュール(予定)
 - サーチャーの資格
 - 授業のWebサイトとE-mail
- 講義内容
 - 情報検索とは

2006年9月27日

後期 水曜4/5限

402教室

江草由佳

国立教育政策研究所

yuka@nier.go.jp

座席指定制:
座席表を確認してから座ること

授業の目的・内容

- 蓄積された大量情報の中から、必要な情報を迅速、的確に検索するための知識を習得する。
- また、各種情報メディアによる検索の特徴と検索システムの構造を理解する。
- 情報検索のプロセスや検索結果の評価方法についても学習する。
- オンディスク検索およびインターネットによる情報検索を実際に演習することにより、検索スキルを習得する。

データベース検索スキルを身につける

評価方法・指導方法

- 評価方法

- 50%:出席＋演習課題の提出結果
- 50%:最終試験の成績

- 指導方法

- 情報検索の理論についてテキストを使用して解説する。その後演習課題を行う。
- 実際にデータベース検索の演習を通じて、情報検索の基本を身につけられるようにする。

テキスト・参考書

- テキスト

- 原田智子編著「三訂情報検索演習」樹村房 2006年10月 ¥1,900?
- 今売っているものは古いため、まだ買わないこと!
 - 10月はじめに発売予定
- 講義や演習で使うので必ず毎回持参すること!

- 参考書

- 情報科学技術協会編「CD-ROM版情報検索の演習 新訂版」日外アソシエーツ 2003年 ¥2,200
- オンディスク検索で利用する。こちらで用意するので購入する必要はない

USBメモリ

- データを保存するメディア
- FDよりも大容量、安定している
- 金額:2,000円くらいから
- USBメモリを毎回授業に持参すること
 - USBメモリであればなんでもよい
 - 授業の資料、レポート作成のため
 - 402教室に保存したファイルは再起動すると消去される
 - 他の授業でも利用可能
 - さまざまなファイルの保存に利用できる
 - すでに持っているのなら、それでよい

スケジュール(予定)

1.	9/27	講義	□	<u>11/29</u>	演習 インターネット検索 (WWW)
□	<u>10/4</u>		7.	12/6	
2.	10/11		8.	12/13	
□	<u>10/18</u>		9.	12/20	
3.	10/25	演習 オンディスク検索 (CD)	□	<u>12/27</u>	テスト
4.	11/1		□	<u>1/3</u>	
5.	11/8		10.	1/10	
6.	11/15		11.	1/17?	
□	<u>11/22</u>		12.	1/?	

□ ← 休講 or 休み

来週(10/4)は休講です

サーチャー(情報専門家)の資格

- 資格

- ー 履歴書に書ける！（運転免許だけじゃさみしい）
- ー コンピュータ利用の基本能力があること、情報の専門化をアピールできる

2. 情報検索基礎能力試験

- 受けたい人は原田先生に連絡（今日なら私へ）
- 参考書：
情報検索の基礎知識 2006年発行 B5判 150ページ 定価：2,000円
購入希望者は原田先生へ（今日なら私へ）

3. 情報検索応用能力試験

- 1級/2級

授業のWebサイトとE-mail

- 授業のWebサイト
 - <http://momiji.mimoza.jp/lecture/2006/ir/>
 - 授業の資料がある
 - 授業の開始時にお気に入りに追加＋授業資料のダウンロードをすること
- 授業時間外の質問はE-mailで！ yuka@nier.go.jp
 - 使えない人は次回までに使えるようになっていること！
 - メールアドレス：学籍番号@stu.tsurumi-u.ac.jp
 - パスワードがわからない人は学生証を持って教務課へ
 - 鶴見大学Webメールシステム
 - <https://ems2.s-idc.net/eaajsp/login.jsp?email=1>

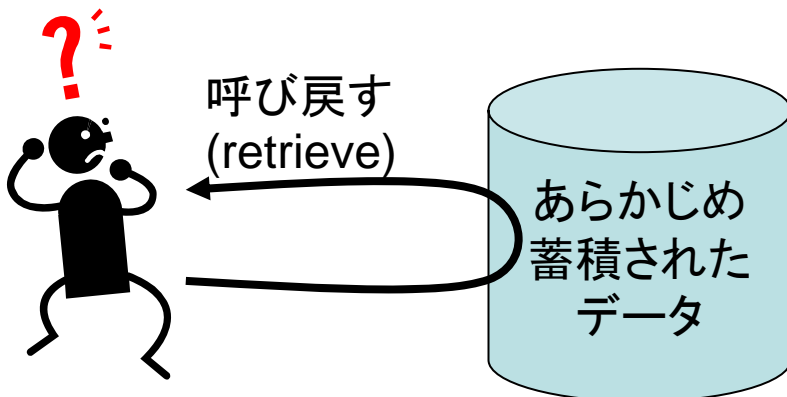
授業ホームページ利用方法 (次回以降の演習準備)

- 授業ホームページ閲覧＋お気に入り追加
 1. Internet Explore を起動
 - “スタート”→“全てのプログラム”→“Internet Explore”
 2. “アドレス” に以下を入力、“Enter” キー
<http://momiji.mimoza.jp/lecture/2006/ir/>
 3. “お気に入り”→“お気に入りに追加”→“OK”
- 授業資料取り寄せ(ダウンロード)方法
 1. リンクの部分の上のマウスポインタを持っていき、右ボタンをクリック
 2. “対象をファイルに保存”を選ぶ
 3. USBメモリの場所を選び、“保存”ボタンをクリックする
 4. (印刷したい人は)印刷
“ファイル”→“印刷”→“印刷対象”:「配布資料」→“OK”

1. 情報検索とは

- **情報検索**

- **IR**: **i**nformation (storage and) **r**etrieval
 - 情報(information) を呼び戻すこと(retrieval)
 - 元は**i**nformation storage and **r**etrieval 情報の蓄積と検索
- 1950年に**ムーアズ**(Calvin N. Mooers)が初めて定義
- 1960年代に広く使われるようになる
- search: データベース検索では、これも「検索」と訳す



retriever(レトリバー):
獲物をくわえて戻って
くるように訓練された猟犬

1.情報検索とは（種類）

(1)ディスク検索とWeb検索

(2)遡及検索とカレントアウェアネス検索

(3)文献検索とファクト検索

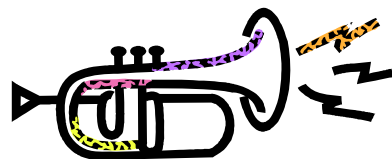
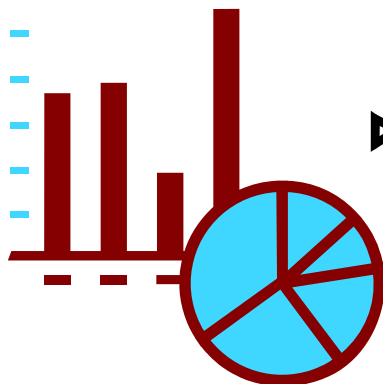
(4)自然語検索と統制語検索

(5)転置索引検索と順次検索

(6)概念検索と連想検索

(1) ディスク検索とWeb検索

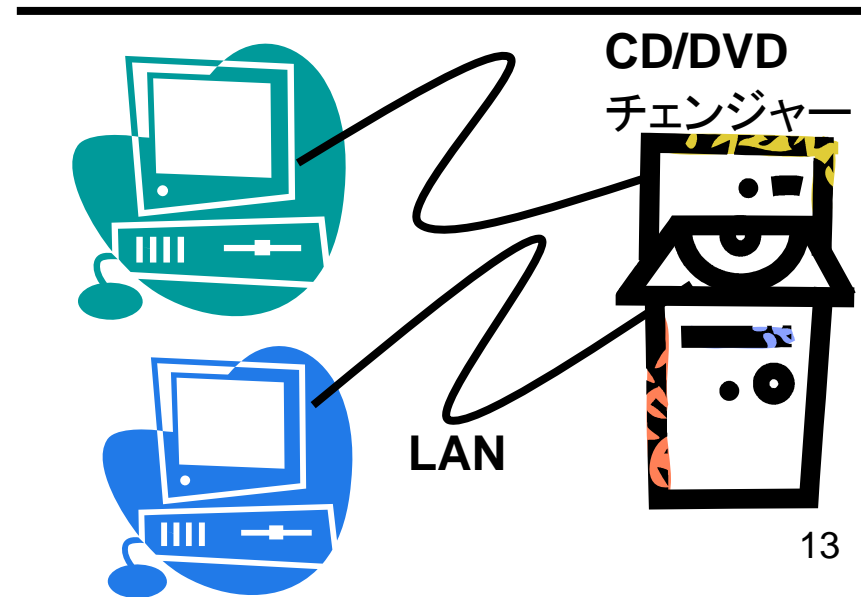
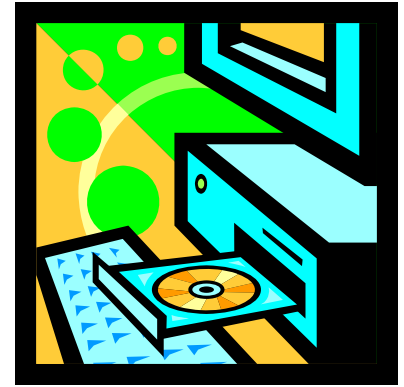
- ディスク検索とWeb検索
 - コンピュータを用いた検索
 - 長所 (印刷物を検索するのに比べて)
 - マルチメディア情報 (図表、写真、音声、動画)
 - 比較的簡単に検索可能
 - 短所 (印刷物を検索するのに比べて)
 - 個々の検索方法を覚える必要がある



(1) ディスク検索とWeb検索

1) ディスク検索

- CD-ROM、DVD-ROM記録メディアに収録したデータベースを検索する方法



(1) ディスク検索とWeb検索

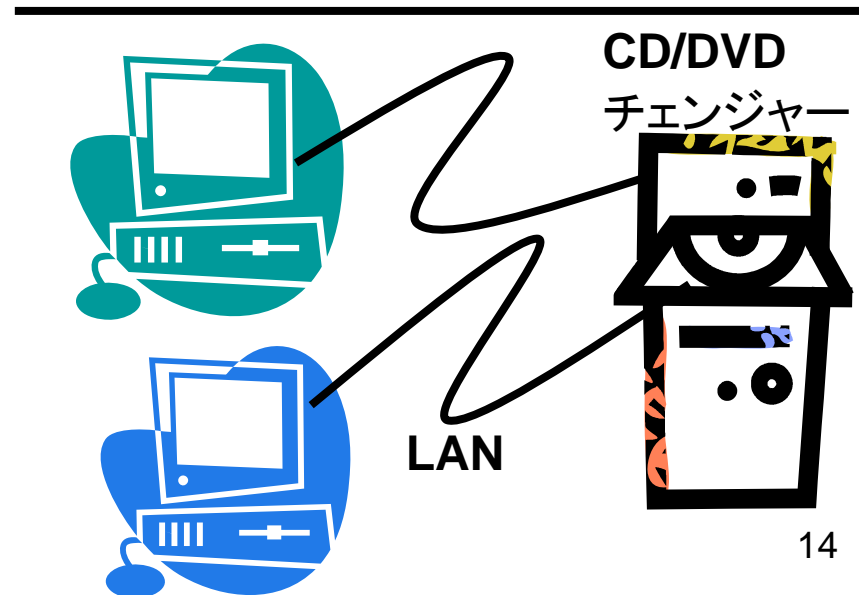
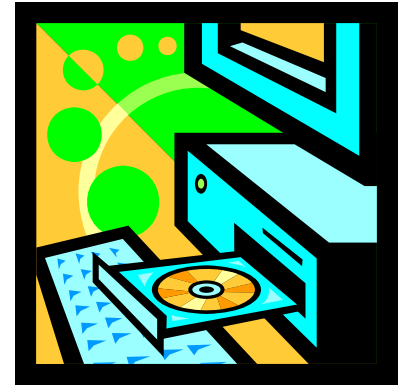
1) ディスク検索

ー 長所

- 外部通信回線の影響を受けない
- 購入料金が固定で予算化しやすい

ー 短所

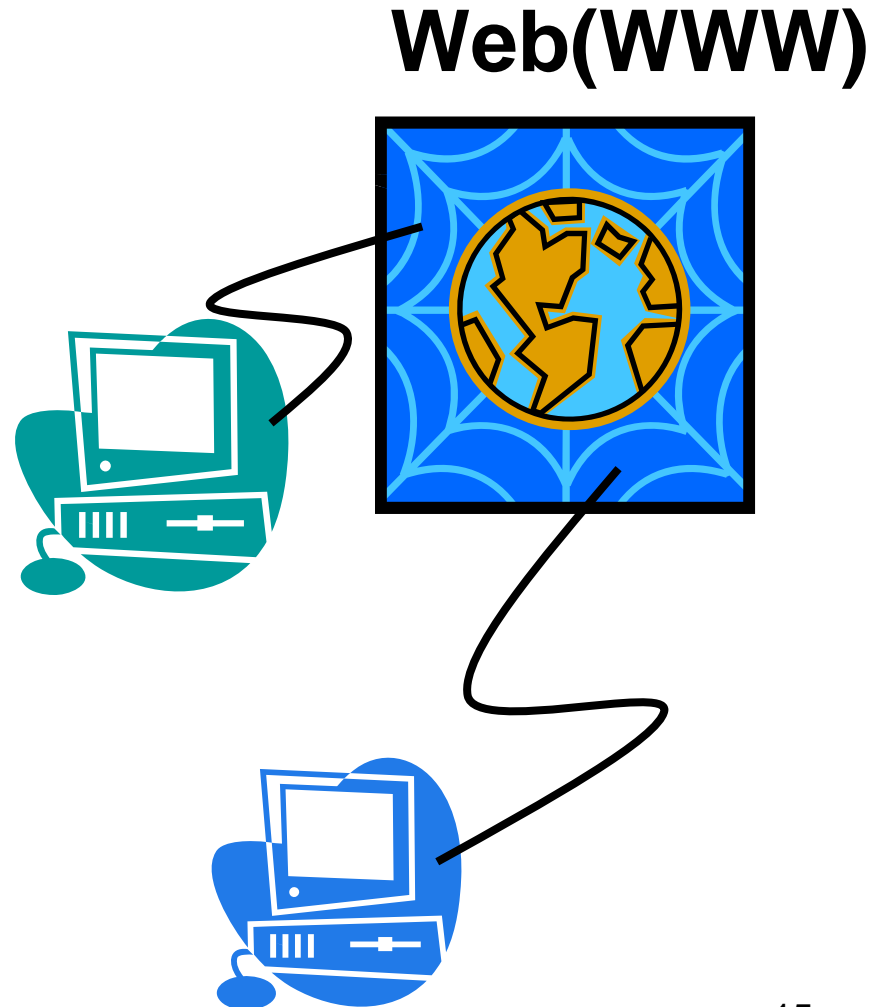
- ディスクの切り替え操作の必要
- 記憶容量の限界 (CD-ROMだと700MB程度、DVD-ROMだと約5GB)



(1) ディスク検索とWeb検索

2) Web検索

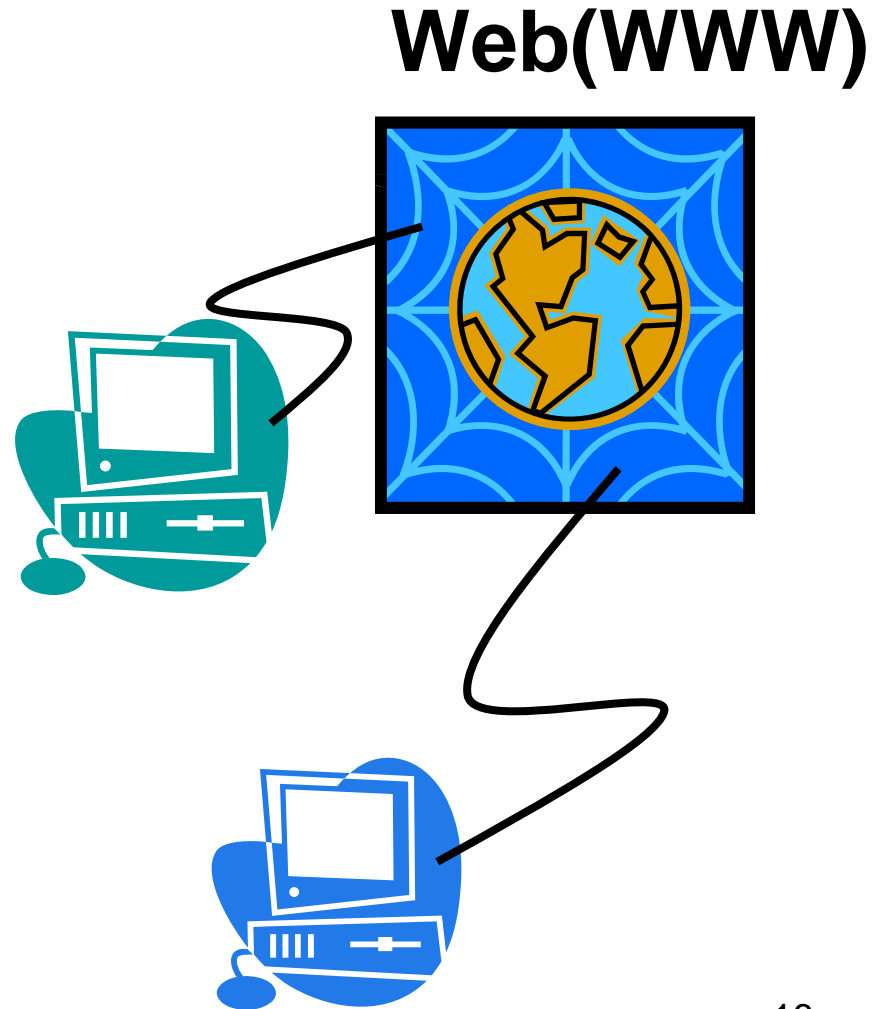
- Web上に公開されているデジタルコンテンツを検索する方法
- インターネット検索、ネット検索



(1) ディスク検索とWeb検索

2) Web検索

- 多くの無料検索Webサイトが存在
- 有料のものは契約、ID・パスワード取得が必要
- ページ間のリンク機能を活用可能



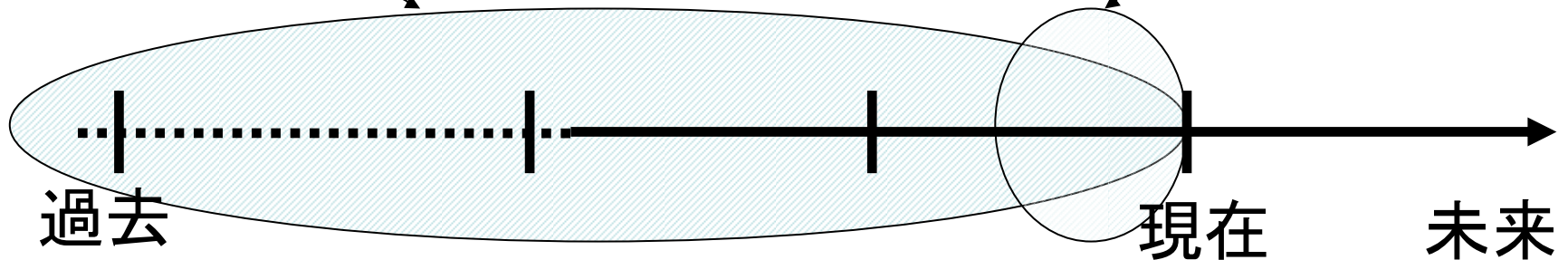
(2) 遡及検索と カレントアウェアネス検索

1) 遡及検索

- 過去に遡(さかのぼ)って検索すること

2) カレントアウェアネス検索

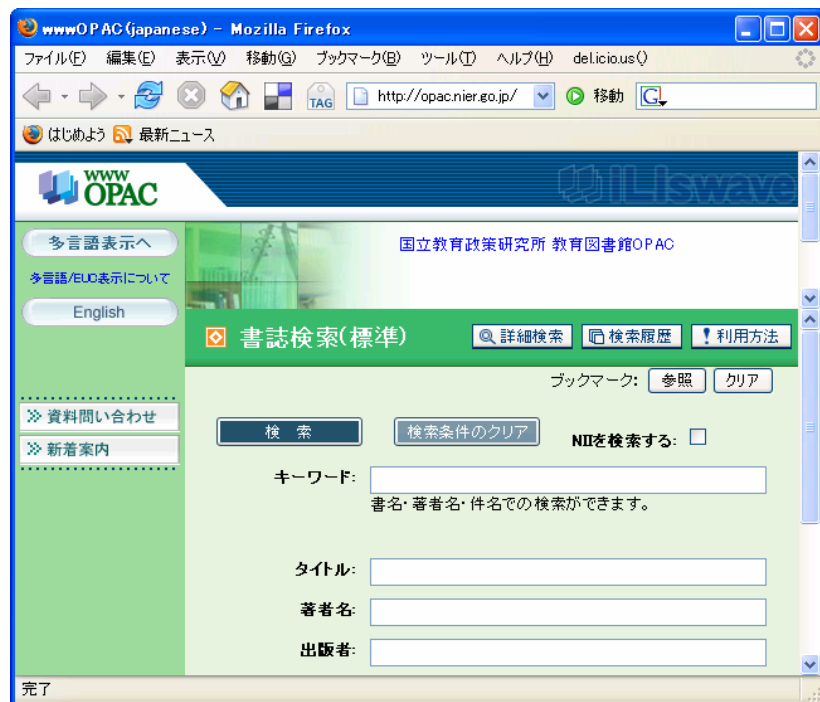
- 現状における最新情報を検索すること
- SDIサービス:定期的に最新情報を入手



(3) 文献検索とファクト検索

- 文献検索

- 文献情報(図書、雑誌、論文、新聞記事)の検索
- 図書館などでよく扱う検索
- 書誌データや本文内容に関する文字情報の検索



(3) 文献検索とファクト検索

- ファクト検索 (事実型検索)
 - 数値情報: 株価、為替レート、統計データ
 - 画像情報: 写真、設計図、地図
 - 動画情報: テレビ、ニュース
 - (フルテキスト: 論文本文、新聞本文)



Yahoo!ファイナンス - 8268.t - Microsoft Internet Explorer

アドレス http://quote.yahoo.co.jp/q?s=8268.t&d=t

(株)西友 (東証1部:8268) - 関連情報: ニュース・企業情報・株主優待・掲示板・レポート

取引値 15:00 194	前日比 -4 (-2.02%)	前日終値 198	出来高 3,683,000	時価総額 172,560百万円
始値 198	高値 198	安値 194	売気配 ---	買気配 ---
発行済株式数 889,486,797株	純資産倍率 (連)-7.06倍	1株株主資本 (連)-27.48円	1株利益 (連)-25.43円	株価収益率 (連)-7.63倍
1株配当 ---	1株配当 ---	1株株主資本 (連)-27.48円	純資産倍率 (連)-7.06倍	株価収益率 (連)-7.63倍
株主資本比率 (連)16.8%	株主資本利益率 (連)-32.16%	総資産利益率 (連)-2.74%	調整1株益 ---	決算年月 2005年12月
単元株数 1,000株	単元株数 1,000株	単元株数 1,000株	単元株数 1,000株	単元株数 1,000株

8268.t 2006/9/23 (C) Yahoo!

チャート(小): [1日](#) | [1年](#) | [なし](#)

チャート(大): [1日](#) | [3か月](#) | [6か月](#) | [1年](#) | [2年](#) | [5年](#) | [10年](#)

気配

(4) 自然語検索と統制語検索

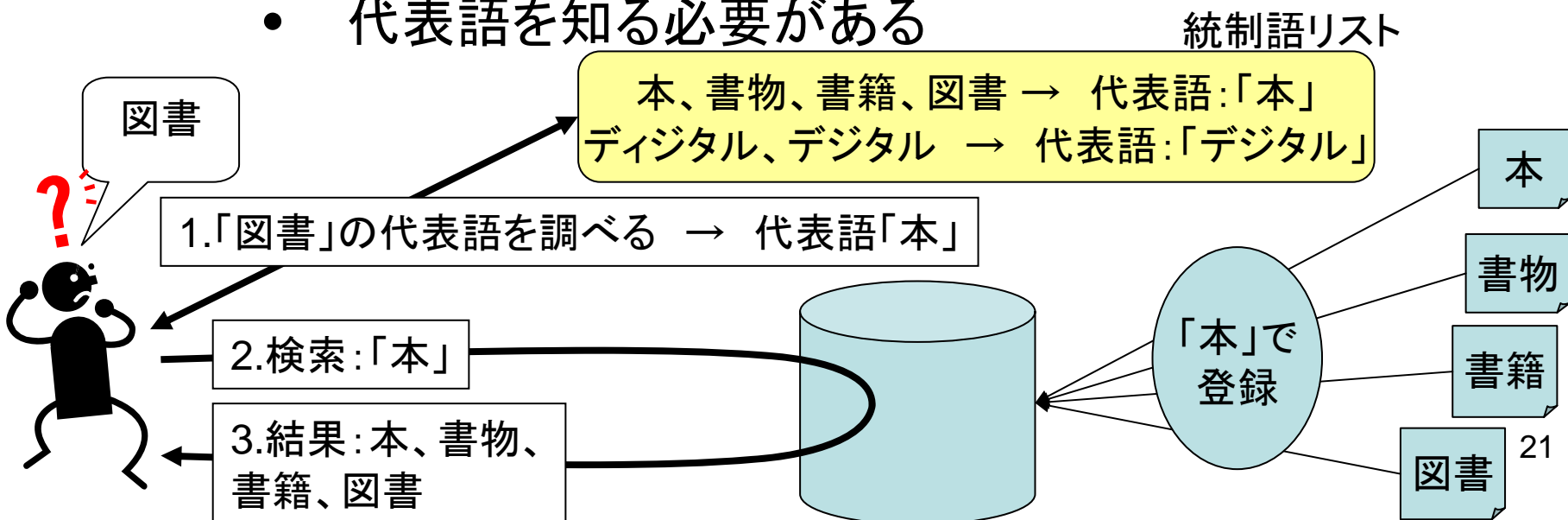
- 自然語検索

- 日常的につかっている言葉をそのまま検索語として使った検索
- 検索漏れを防ぐには以下に注意
 - 同義語：本、書物、書籍、図書、ブック
 - 表記のゆれ：デジタルとディジタル
 - 略語と完全表記：SBEと牛海綿状脳症
 - 日本語と英語：図書館とライブラリー
 - 表現のちがひ：日本とわが国

(4) 自然語検索と統制語検索

● 統制語検索

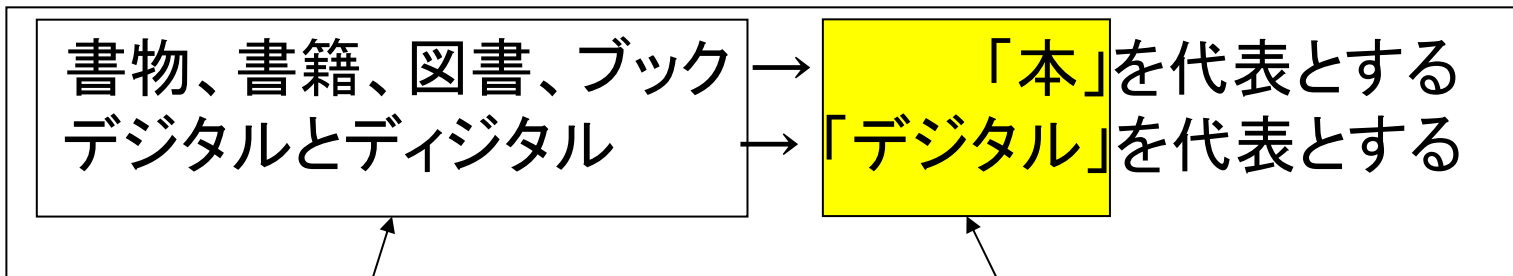
- ー 類義語、表記のゆれなど自然言語での語に関するわずらわしさを解消するため考案された
- ー 統制語リスト(統制語彙)を使った検索
 - 代表語を使って検索すれば、検索漏れがなくなる
 - 代表語を知る必要がある



(4) 自然語検索と統制語検索

- 統計語リスト(統計語彙)
 - 件名表目標(subject headings)
 - 代表例:「基本件名標目表」(BSH: Basic Subject Headings)
 - 図書館の目録作成時の件名付与に使用される
 - シソーラス(thesaurus)
 - 自然語検索時にうまく検索できないときに活用できる
 - OR検索をつかうための同義語探しなど

統計語リスト(統計語彙)



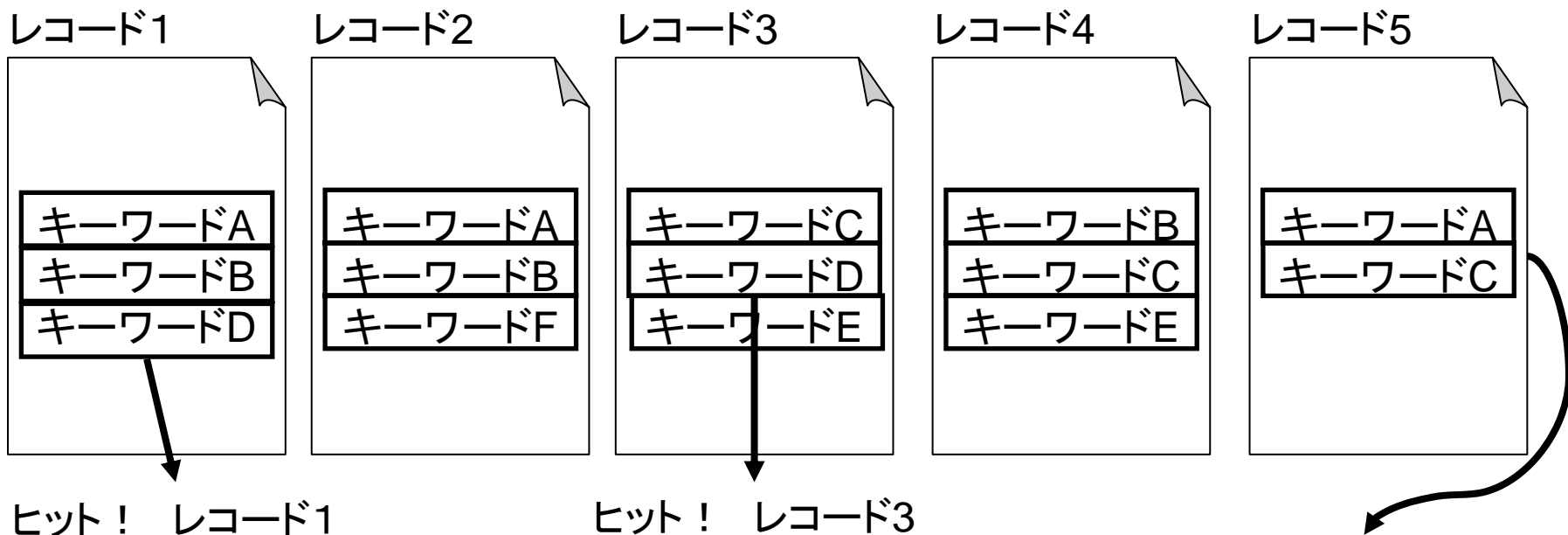
非ディスクリプタ

ディスクリプタ: 代表語のこと

(5) 転置索引検索と順次検索

● 順次検索

- レコードを最初から最後まで順番に検索
- レコードが多くなると時間がかかってしまう



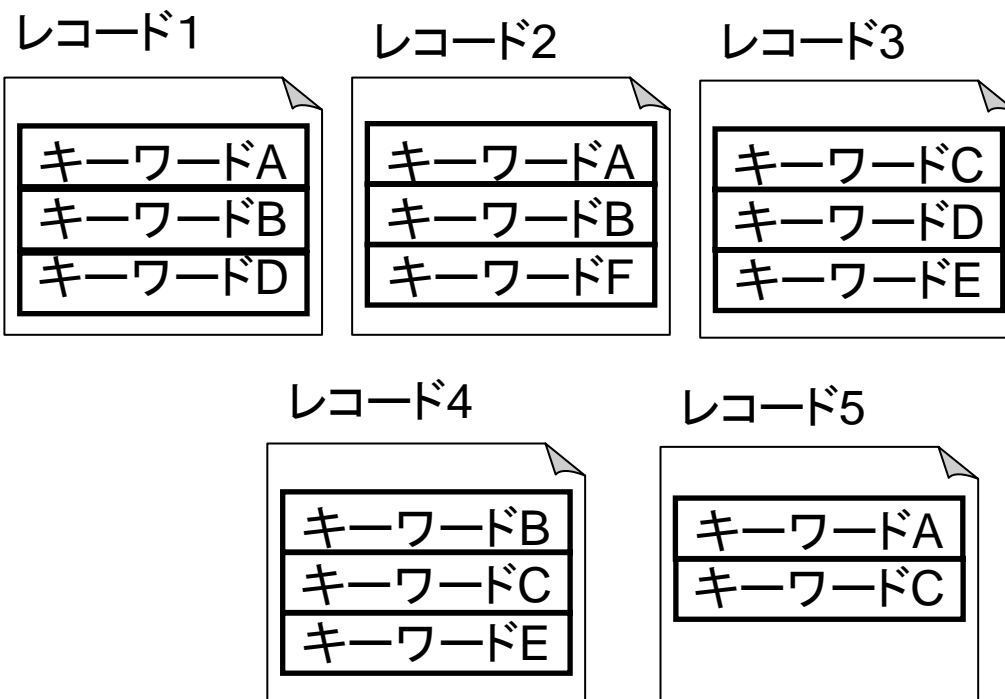
例: 「キーワードD」を検索する

ヒットした件数は2件、
ヒットしたレコードは
レコード1とレコード3

(5) 転置索引検索と順次検索

- 転置索引検索

- インバーテッド・ファイル(転置ファイル、倒置ファイル)を使用した検索
- 順次索引検索の時間の遅さを解消するために考案された



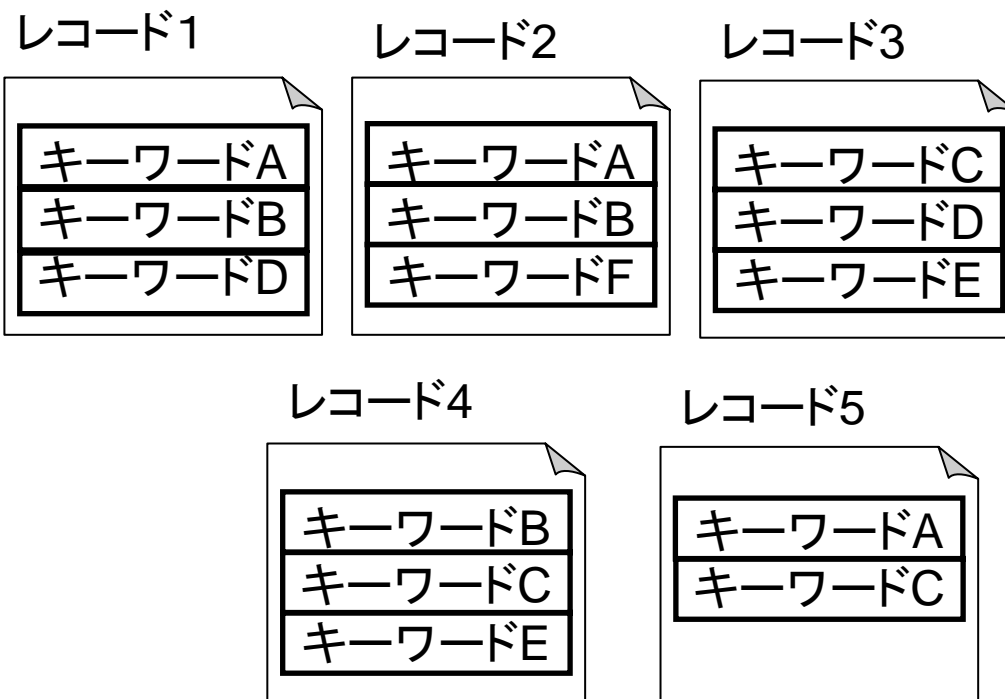
インバーテッドファイル

キーワード	レコード	件数
A	1 2 5	3件
B	1 2 4	3件
C	3 4 5	3件
D	1 3	2件
E	3 4	2件
F	2	1件

(5) 転置索引検索と順次検索

- 転置索引検索

- 例: 「キーワードD」を検索する
- 欠点: あらかじめインバーテッドファイルを作らなければならない、ファイルの容量増



インバーテッドファイル

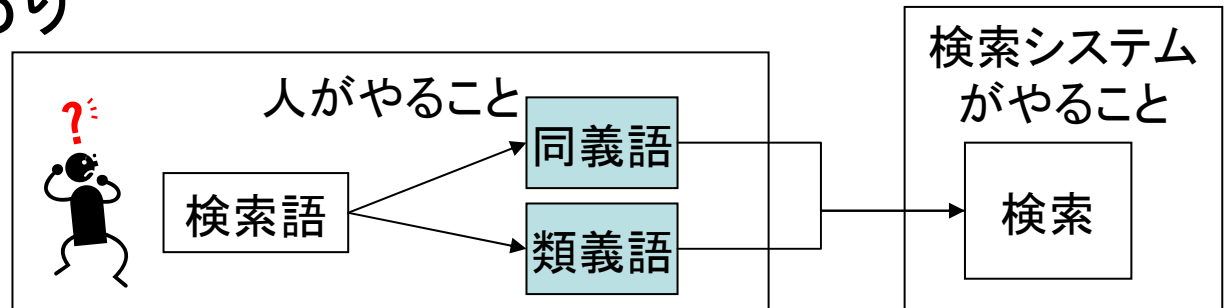
キーワード	レコード	件数
A	1 2 5	3件
B	1 2 4	3件
C	3 4 5	3件
D	1 3	2件
E	3 4	2件
F	2	1件

(6)概念検索と連想検索

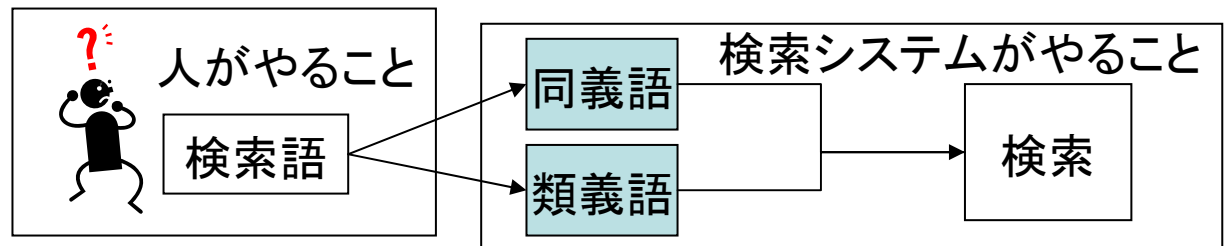
1)概念検索

- 単純な文字列一致ではなく概念で検索
- システム内に統制語彙を組み入れた検索
- 同義語や類義語検索に関してはある程度期待できる成果あり

通常検索
(文字列一致)



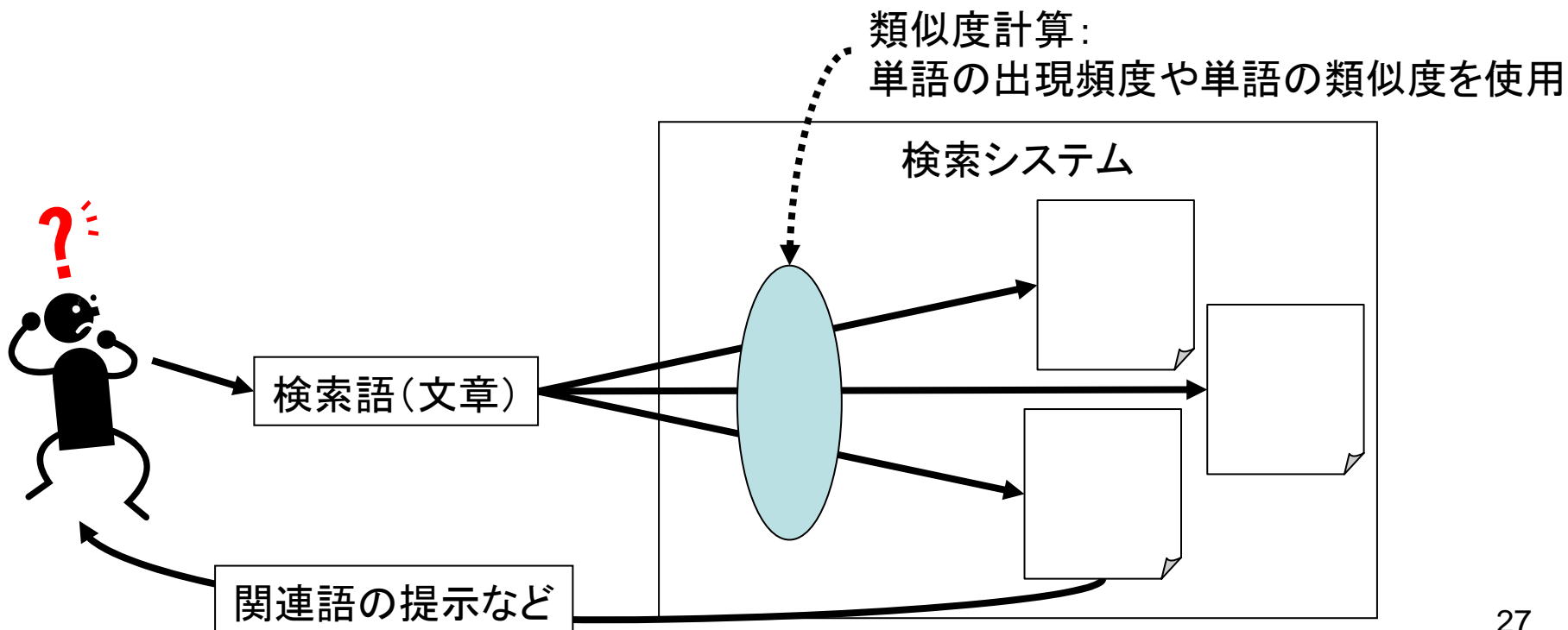
概念検索



(6)概念検索と連想検索

2)連想検索 GETA

– 文章・文書間の類似性を使った検索



今日のまとめ

- 講義内容
 1. 情報検索とは
 - (1) ディスク検索とWeb検索
 - (2) 遡及検索とカレントアウェアネス検索
 - (3) 文献検索とファクト検索
 - (4) 自然語検索と統制語検索
 - (5) 転置索引検索と順次検索
 - (6) 概念検索と連想検索
- 連絡事項
 - － 来週は休講
 - － USBメモリを持参すること
 - － 教科書はまだ買わないこと次回の授業時に連絡します

第一回 課題提出方法

