

情報検索サービス 第1回資料

オリエンテーションと講義

聖徳大学図書補講習
2008年8月29日、9月1日
江草由佳
国立教育政策研究所
yuka@nier.go.jp

第1回資料

1

テキスト・参考書

- テキスト
 - 情報科学技術協編「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
 - **演習で使うので必ず毎回持参すること!**
- 参考書
 - 原田智子編著「三訂情報検索演習」樹村房 2006年10月 ¥1,900

第1回資料

4

講義ホームページ利用方法 (次回以降の演習準備)

- 講義ホームページ閲覧＋お気に入り追加
 1. Internet Explorer を起動
 - “スタート”→“全てのプログラム”→“Internet Explorer”
 2. “アドレス” に以下を入力、“Enter”キー
<http://momiji.mimoza.jp/lecture/2008/ir-seitoku/>
 3. “お気に入り”→“お気に入り”に追加→“OK”講義資料取り寄せ(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メモリ等の場所を選び)“保存”ボタンをクリックする
 4. (印刷したい人は)印刷

第1回資料

7

今回のおしながき

- オリエンテーション
 - 講義の目的・内容
 - テキスト・参考書
 - スケジュール(予定)
 - 講義のWebサイトとE-mail
 - USBメモリ利用のすすめ
- 講義内容
 - 情報検索とは
 - 情報検索の理論
 - 論理演算子(AND、OR、NOT)
 - トランケーション(前方一致、後方一致、中間一致)
 - 情報検索の評価
 - データベースとは(起源、定義)
- 第1回演習課題:論理演算とトランケーション

第1回資料

2

スケジュール(予定)

注)進捗状況により多少前後します

- 9:00～10:30(90分)
 - オリエンテーション、情報検索とは
- 10:40～12:10(90分)
 - 新聞記事情報の検索
- 13:00～14:20(80分)
 - CD-ROM検索の準備と使い方
- 13:00～14:10(70分)
 - NDL-OPAC, Webcat, Cinii
- 14:30～15:30(60分)
 - 人物情報検索
- 14:20～15:30(70分)
 - 試験勉強
- 15:40～16:40(60分)
 - 雑誌情報検索
- 15:40～17:00(80分)
 - 試験
- 16:40～17:00(20分)
 - 総合演習、質疑応答

第1回資料

5

USBメモリ利用のすすめ

- USBメモリの利用をすすめます
 - USBメモリはあんなに安い
 - 講義の資料、課題の保存のため
 - 保存したファイルは再起動すると消去されるため
 - 他の講義でも利用可能
 - さまざまなファイルの保存に利用できる
- USBメモリとは
 - データを保存するメディア
 - FDより大容量、安定している
 - 金額1,000円前後～
- USBメモリの使い方
 - <http://momiji.mimoza.jp/lecture/2007/QA/#usb>
- USBメモリのはすし方
 - <http://momiji.mimoza.jp/lecture/2007/QA/#usb-exit>

第1回資料

8

講義・演習の目的・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。
- オンライン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

第1回資料

3

講義のWebサイトとE-mail

- 講義のWebサイト
 - <http://momiji.mimoza.jp/lecture/2008/ir-seitoku/>
 - 講義の資料がある
 - 講義の開始時にお気に入り追加＋講義資料のダウンロードをすること
- 講義時間外の質問はE-mailで！ yuka@nier.go.jp

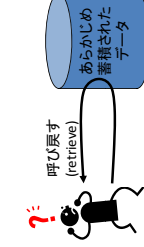
第1回資料

6

1.情報検索とは

情報検索

- **IR**: information (storage and) retrieval
 - 情報 (information) を呼び戻すこと (retrieval)
 - 元は information storage and retrieval 情報の蓄積と検索
- 1950年に**ムアーズ**(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



第1回資料

9

情報検索の理論(1) —p.19

- コンピュータ検索では論理演算の概念が基本
 1. データベース全体から合致するものを検索し
 2. 論理積、論理和、論理差の集合の概念をもちいて、広げたり、狭めたりして検索
- 論理演算
- トランケーション

p.19は
参考書のページ数を表す

第3回講義 10

論理積の例：教育 and 情報

- ヒット
 - 「教育」と「情報」の両方を含む
高等学校「情報」教員養成におけるメディア・リテラシー教育
- ヒット
 - 情報教育に関わる問題—高等学校の情報関係科目の履修の経験による検討
- 3. 高校生のコンピュータに対する意識調査
- 4. 高等学校教員免許取得の現状と課題—教科「情報」の免許状の取得
- ヒット
 - 高校普通教科「情報」とスキル教育
- 6. 高校生の情報活用に関する日中比較
- 7. 高校教育の多様化—高等学校現場からの報告
- 8. デジタルメディア利用教授不安の減少と高校生のコンピュータ不安



第3回講義 13

情報検索の理論(4) —p.20 (1)論理演算子 -p.20

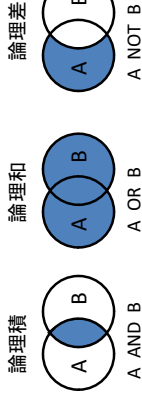
- 論理差 (NOT) -p.20
 - Aという検索語を持つ情報の集合からBという検索語をもつ集合部分を除いた部分を検索すること
 - 複数の言語で書かれていることはすくないので、ロシア語文献を抜くといった検索に有効
 - 例：「トラック事故以外の自動車事故」
 - 検索式：”自動車事故 NOTトラック事故”



第3回講義 15

情報検索の理論(1) —p.19 (1)論理演算子 -p.19

- 論理演算
 - 論理積 (AND)、論理和 (OR)、論理差 (NOT)
- 論理演算子
 - 説明ではAND, OR, NOTを使うが、演算子の書き方はシステムによって様々である。



第1回講義 1-9図 p.19 11

情報検索の理論(3) —p.19 (1)論理演算子 -p.19

- 論理和 (OR) -p.20
 - Aという検索語を持つ情報の集合とBという検索語のいずれか一方の検索語をもつ集合部分と、両方をもつ集合部分全てを検索すること
 - 同義語などの検索に有用
 - 例：「高校におけるコンピュータ教育」
 - 検索式：”高校 OR 高等学校”



p.19は
参考書のページ数を表す

14

論理差の例：自動車 not トラック

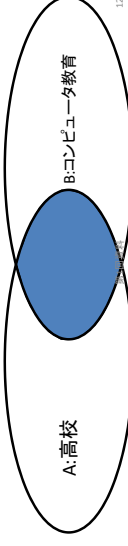
- ヒット
 - 「自動車」を含みかつ「トラック」を含まない
- ヒット
 - アジアの自動車産業
- 3. 新しい自動車の構造と運転取扱法
- ヒット
 - 自動3輪トラックの構造と運転
- 5. アメリカの自動車会社ビッグ3の復活
- 6. はたらく自動車：トラック・工車用車両
- 6. 大型トラック・トレーラの安全対策の研究



第1回講義 17

情報検索の理論(2) —p.19 (1)論理演算子 -p.19

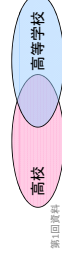
- 論理積 (AND) -p.20
 - Aという検索語を持つ情報の集合とBという検索語をもつ集合の両方を含む部分を検索すること
 - 情報を絞り込んでいくときに有用
 - 例：「高校におけるコンピュータ教育」
 - 検索式：”高校 AND コンピュータ教育”



12

論理和の例：高校 or 高等学校

- ヒット
 - 「高校」か「高等学校」のどちらかを含む
高等学校「情報」教員養成におけるメディア・リテラシー教育
- ヒット
 - 情報教育に関わる問題—高等学校の情報関係科目の履修の経験による検討
- ヒット
 - 高校生のコンピュータに対する意識調査
- ヒット
 - 高等学校教員免許取得の現状と課題—教科「情報」の免許状の取得
- ヒット
 - 高校普通教科「情報」とスキル教育
- ヒット
 - 高校生の情報活用に関する日中比較
- ヒット
 - 高校教育の多様化—高等学校現場からの報告
- ヒット
 - デジタルメディア利用教授不安の減少と高校生のコンピュータ不安



第3回講義 15

情報検索の理論(8) —p.20 (3)トランケーション -p.21

- トランケーション
 - 検索語を入力する場合に、語の一部が任意であるように指定して検索すること
 - 指定方法としては、大きく分けて2種類ある
 - 任意の部分を指定する方法 (マスク文字を使用)
 - 決まっている位置を指定する方法

図書館?

ここはなんでもいい(任意)と指定する方法
任意の部分の特殊文字と指定する方法
この(前方)は「図書館」と決まっていますよと任意の部分を特殊文字と指定する方法
のことをマスク文字という

18

マスク文字

- マスク文字(ワイルドカード)
 - 任意文字とする部分に使用する入力文字
 - マスク(mask): 覆い隠すから由来
- マスク文字はシステムによって異なる

以下は説明で使用

 - ? : 0文字以上、何文字でもよい
 - # : 0文字または一文字
 - ! : ちょうど一文字



第10回資料

19

情報検索の理論(10) — p.20 (3)トランケーション — p.21

- 2) 後方一致検索 — p.22
 - 検索語の **後方** を一致させる検索
 - 検索語の **始まり** を任意文字に指定する方検索
- 例)
 - **情報** / **情報**
 - **情報**、**安全情報**、**特許情報**

p.? は
参考書のページ数
を表す

第10回資料

22

情報検索結果の評価(1) — p.23

- 検索結果の評価
 - 検索終了後、求める情報が適切に検索できているか、検索漏れはやノイズがないかどうかをチェックする
- 検索漏れ
 - 本来必要な情報でデータベースに存在するにもかかわらず検索されなかった情報
- ノイズ
 - そのテーマに不要な情報が入り込んで検索された情報

p.? は
参考書のページ数
を表す

第10回資料

23

一致指定文字*

- 一致指定文字*
 - 一致する部分がかかを示す特殊文字
 - 一致指定文字はシステムによって異なる

以下は説明で使用

 - / : ここから始まる、もしくは、ここで終わる

*マスク文字と異なり、この「一致指定文字」は
江草が説明のために作った造語です。

第10回資料

20

情報検索の理論(11) — p.20 (3)トランケーション — p.21

- 3) 中間任意検索 — p.23
 - 検索語の **途中** を任意文字に指定する検索
- 例)
 - **情報**? システム ⇒ 情報システム、情報管理システム、情報検索システム
 - **ログ!** ⇒ ログイン、ログオン
 - 同義語が同時に検索できる
 - **WOMIN** ⇒ WOMAN, WOMEN
 - 単数形、複数形が同時に検索できる
 - **GR!Y** ⇒ GREY, GRAY
 - 英米綴りの違いを同時に検索できる

p.? は
参考書のページ数
を表す

第10回資料

23

情報検索結果の評価(2) — p.23

- 再現率 (recall ratio)
 - データベースに合致する適合情報のうち、どれだけ適合情報が検索されたかという割合を表す
 - 検索漏れの程度を表す指標となる
 - データベース全体の適合情報を実際に調べるのは不可能なため、普通、正確にはわからない
- 精度 (precision ratio) (適合率 (relevance ratio))
 - 実際に得られた検索結果の情報全体のうち、どれだけ適合情報が検索されたかという割合を表す。
 - ノイズの程度を表す指標となる
 - 検索結果から容易に計算可能
- 両方とも高いのが理想だが、どちらかを上げるとどちらかが下がるという関係にある

第10回資料

23

情報検索の理論(9) — p.20 (3)トランケーション — p.21

- 1) 前方一致検索 — p.22
 - 検索語の **前方** が一致する検索
 - 検索語の **末尾** を任意文字に指定する検索
 - 大抵どのシステムにもある
- 例)
 - **情報**? / **情報**
 - **情報**、**情報検索**、**情報検索システム**
 - **CAT#** ⇒ CAT, CATS, CATV
 - 猫を検索したいときに、CATVまで検索されてしまう
 - **DIS!** ⇒ DISK, DISC
 - 英米綴りの違いを同時に検索できる

p.? は
参考書のページ数
を表す

第10回資料

21

情報検索の理論(12) — p.20 (3)トランケーション — p.21

- 4) 中間一致検索 — p.23
 - 検索語の **中間** が一致する検索
 - 検索語の **両端** を任意文字に指定する検索
 - インターネットの検索エンジンでは中間一致していることが多い
 - 一般に3文字以下の略字ではノイズを招くので、トランケーションを使わず、完全一致させたほうがよい
- 例)
 - **?情報**? ⇒ 情報、交通情報、情報システム、交通情報システム

p.? は
参考書のページ数
を表す

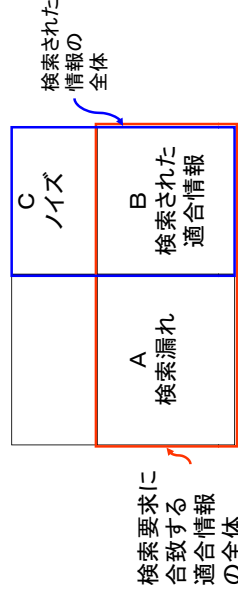
第10回資料

24

情報検索結果の評価(3) — p.23

再現率 $R = B \div (A + B) \times 100\%$
 精度 $P = B \div (C + B) \times 100\%$

図1-10 — p.33



第10回資料

27

情報検索結果の評価(4) -p.23

- 例) 10件の文献を出力、10件全部が適合文献、データベース全体には100件の適合文書が存在
 - 精度: $10/10 \times 100=100\%$
 - 再現率: $10/100 \times 100=10\%$
 - 残りの90%が検索漏れ
- 例) 50件の文献を出力、30件が適合文献、データベース全体には70件適合文書が存在
 - 精度: $30 \div 50 \times 100 = 60\%$
 - ノイズは40%
 - 再現率: $30 \div 70 \times 100 = 43\%$
 - 検索漏れは57%

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

データベースの起源 -27p.

- 1950年代
- 米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



第1回演習課題: 論理演算とトランゼーション

- 第1回演習課題を行い、提出しなさい
 - 第1回演習課題は講義Webサイトにもあります。
 - 学籍番号、名前、提出日を忘れなさいこと
- この課題のねらい:
 - 論理演算やトランゼーション機能を使った検索式を組み立てることができるようになる。
 - 論理演算子の書き方やマスク文字が、システムによって異なることを体験する。

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つて、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である
- 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つて、データの特性と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つて、コンピュータでアクセス可能なもの(X0807)