

情報検索演習  
第2回資料  
講義の続き、  
論理演算の演習

鶴見大学

2010年9月29日

江草由佳

国立教育政策研究所

yuka@nier.go.jp

# 本日のお品書き

- 課題の提出
- 情報検索の評価
- データベースとは(起源、定義)
- 第1回レポート課題出題
- 論理演算の演習
  - 第2回演習課題: 論理演算(CinNii)

# 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

# 情報検索結果の評価(2) -p.23

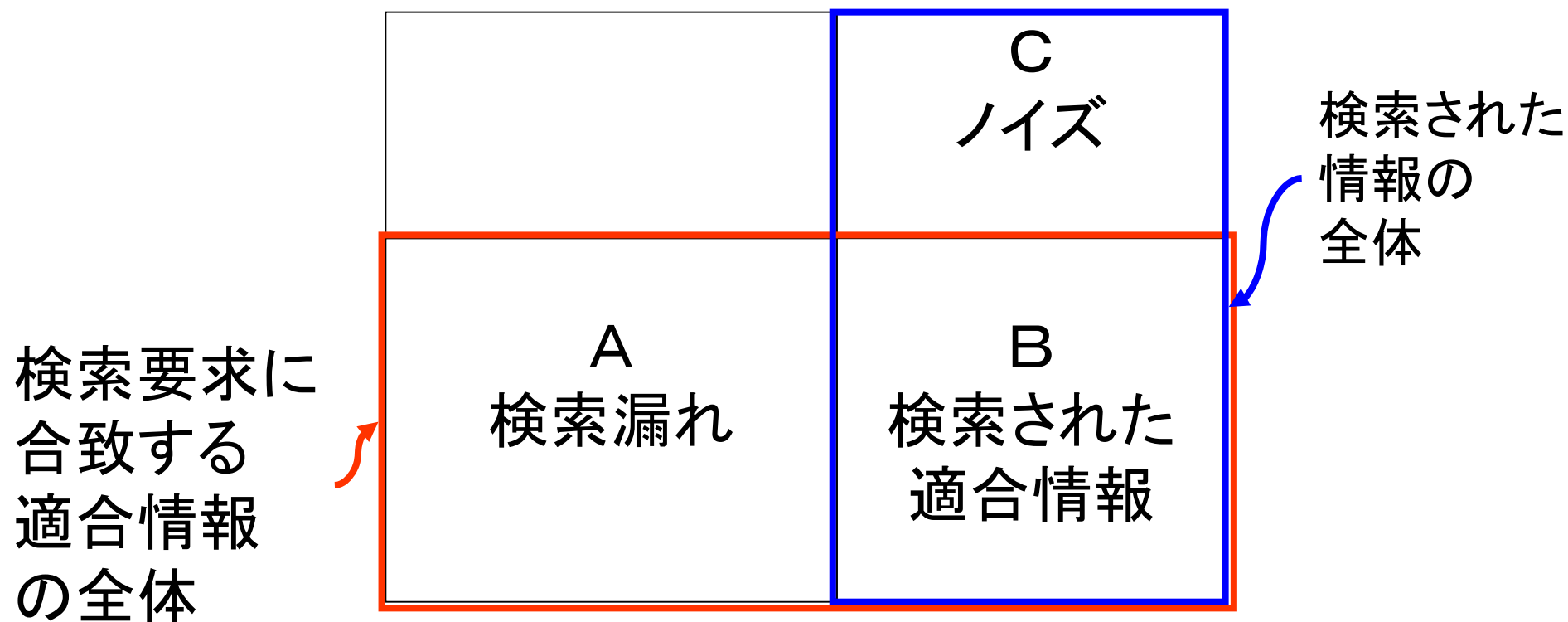
- 再現率 (recall ratio)
  - データベースに合致する適合情報のうち、どれだけ適合情報が検索されたかという割合を表す
  - 検索漏れの程度を表す指標となる
  - データベース全体の適合情報を実際に調べるのは不可能なため、普通、正確にはわからない
- 精度 (precision ratio) (適合率 (relevance ratio))
  - 実際に得られた検索結果の情報全体のうち、どれだけ適合情報が検索されたかという割合を表す。
  - ノイズの程度を表す指標となる
  - 検索結果から容易に計算可能
- 両方とも高いのが理想だが、どちらかを上げるとどちらかが下がると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

# 情報検索結果の評価(3) -p.23

図1-10 -p.33

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

精度  $P = B \div (C + B) \times 100\%$

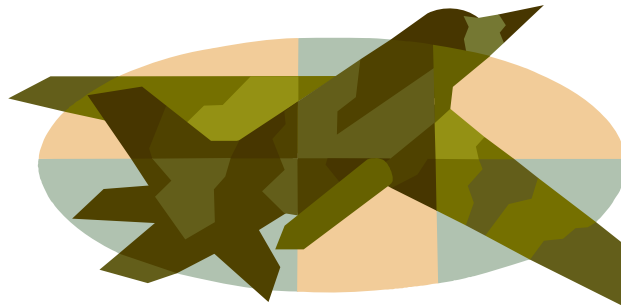


# 情報検索結果の評価(4) -p.23

- 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

# 第1回レポート課題:第1,2章のまとめ

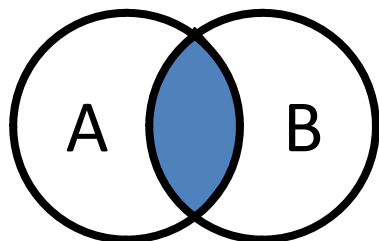
- 教科書の第1章、第2章(p.1～p.36)の内容をWordで2ページピッタリにまとめなさい。
  - 10/20の授業開始時に提出しなさい。
  - 講義名、第？回レポート課題、学籍番号、名前、提出日を最初に記述すること
  - 2分割印刷で1枚に印刷して提出しなさい

# 第1回レポート課題： この課題のねらい

- 情報検索の理論について理解する
  - 情報検索の定義(何の訳語？誰が言ったの？何年ごろ？など)
  - 情報検索の種類
  - データベースの定義・・・どの法律？どんなもの？
  - データベースの分類
  - データベースのファイル構造
    - どんなファイルがあった？(大きくわけて2種類)
  - 情報検索の理論
    - 論理演算(論理積、論理和、論理差、**図も描けるように**)
    - トランケーション(前方一致、後方一致、中間一致、中間任意)
  - 検索結果の評価(再現率、精度)

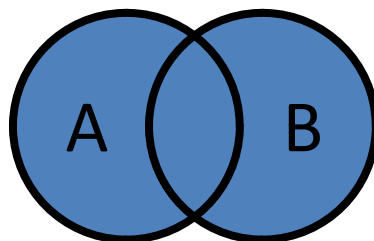
- 論理演算
  - 論理積 (AND)、論理和 (OR)、論理差 (NOT)
- 論理演算子
  - 説明ではAND, OR, NOTを使うが、演算子の書き方はシステムによって様々である。

論理積



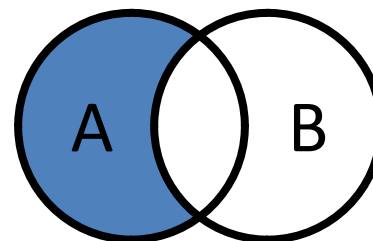
A AND B

論理和



A OR B

論理差



A NOT B

# 論理演算の演習

- CiNii(サイニイ)
  - <http://ci.nii.ac.jp/>
  - 日本で出版された多くの学術論文が検索できる

- CiNiiを使って論理演算を使った検索をしてみましょう
  - 第2回演習課題：論理演算(CiNii)

# 本日のまとめ

- 課題の提出
- 情報検索の評価
- データベースとは(起源、定義)
- 第1回レポート課題出題
- 論理演算の演習
  - 第2回演習課題: 論理演算(CinNii)