

第3・4回
(2009/4/24, 5/1)

『分類の一般概念と分類理論』

黒橋禎夫教授（情報学研究科）

【【第3回 講義】 2009/4/24 吉田南1号館311 出席者82名

配布資料：パワーポイントスライド（A3二つ折り1部）。演習ページにも掲載。

● 導入

- 「分類は知のはじまり」。無意識の分類から体系化された分類へ。

● 分類の演習

- 挙げられた単語を分類してみる。（解答の提示）大項目から小項目へのトップダウン、類似性から帰納的に行うボトムアップ、いずれの場合も階層構造を持つ。

● 分類の諸問題

- 分類は視点・観点によって異なる

- 「家族的類似性」（ウィトゲンシュタイン）：事物は特性のどの部分を共有をするかで異なった分類が可能で、各々が網の目のようにつながっている。

- 言葉、文化との関係：言葉と概念との関係は必ずしも一意的ではない。むしろ言語化により文節化されるとも言える。またその文化の中での概念の位置づけなどが分類に影響を与える。

- 「オーバーゼーニング」：目的に応じて旧来の分類を縦断するように新たに分類しなおす。例）「スキー」をテーマにスキー用品だけでなく、ツアー予約、道路地図などの全くジャンルの異なる（と思われる）商品を集めて陳列する。

● 動植物の分類

- 人為分類（アリストテレス）→類型分類（アンダソン）→系統分類（ラマルク）

● 図書の分類

- 歴史：個人的・小規模の図書館から蔵書数が多く公共性の高い図書館に変化していく中で体系的かつ普遍性の高い分類が求められるようになった。

- 様々な分類：十進分類法（デューイ、国際、日本）、コロソ分類法（主題＋ファセット）

● ことばの分類—シソーラス

- 一般用語については上位下位よりも同義語関係について用いることが多い。

- 専門用語ではその分野での体系を明らかにする＝ある程度分野的な成熟が前提

- 自然言語処理でのシソーラスの利用

- 用例ベース翻訳への応用

● 分類の数学的理論

- 人為分類から類型分類へ（より多くの形質の共有を調べる）。

- 数量分類学。例）クラスタ分析。特徴ベクトルによる個体表現→各個体間の類似度を算出→クラスタ分析を行うことで樹形図に。

● 情報検索

- （個体の代わりに）テキストを特徴ベクトル表現、類似度計算。

- インターネット（コンピュータのネットワーク）とWWW（テキストのネットワーク）の違い。

- 検索エンジンの仕組み—クローラーが収集したテキストを類似度計算（人間の分類との違い：文字コード、辞書式順序、転置インデックス、Term Frequency、Inverse Document Frequency）。

- 検索の精度を上げるその他の工夫：PageRankによる重要度判別、アンカーテキストをインデクス対象に含める等。

【第4回 演習】 2009/5/1 学術情報メディアセンター203、204

配布資料：演習問題（WebCT、黒橋先生のウェブサイト掲載のものと同内容）、

WebCT操作マニュアル。問題Bの図書リストはGoogle Book Searchへのリンクで提示。

出席者：89名

時間配分：開室～13:10 WebCT接続指導

13:10～13:40 問題、提出方法等解説

13:40～14:25 演習

14:25～14:30 提出方法、アンケート等説明、解散