

情報知識ネットワーク特論

(Information Knowledge Network)

種別・単位：講義・2単位（週1講時）

開講期：二学期

担当者：有村博紀（コンピュータサイエンス専攻・知識ソフトウェア科学講座・内線7678）
喜田拓也（コンピュータサイエンス専攻・知識ソフトウェア科学講座・内線7679）

主題と目標

高度なネットワークと大規模記憶装置の急速な発達によって、現在、膨大な量のデータがネットワーク上で利用されている。本講義では、ネットワークを介して情報や知識を伝達・共有するための技術について講義する。具体的には、ウェブデータやテキストデータなどの大規模非定型データ処理のための基礎技術として、情報検索やデータマイニング、ウェブ情報システムに焦点をあてる。特にアルゴリズムとプロトコルの視点から、パターン照合技術・パターン発見技術・HTMLやXML等の半構造データ処理・情報セキュリティ技術について学ぶ。

授業計画（項目、授業実施回数、内容）

項目	回	内容説明
情報検索とパターン照合	6	索引構造を用いた検索と文字列照合による検索の違い。Namazuなどの全文テキスト検索システムの索引構造として用いられている転置ファイル。ウェブ検索システムGoogleの仕組みと理論的土台。KMP法やBoyer-Moore法とその変種、ビットパラレル手法などの各種パターン照合アルゴリズム。正規表現パターン照合アルゴリズム。近似文字列照合。圧縮テキストに対するパターン照合。
データマイニングとパターン発見	6	データマイニングとは何か。データマイニングの歴史。データマイニングと機械学習・統計との関係。データマイニングの応用。テキストマイニングやゲノム解析のための系列マイニング。決定木構築や素朴ベイズ法などの機械学習アルゴリズム。
ウェブとセキュリティ	3	ウェブ情報システム。XMLとXML情報検索。情報セキュリティと暗号。トランザクション管理。アクセス管理。

評価・教材・受講条件等

《評価》 「情報検索とパターン照合」「データマイニングとパターン発見」「ウェブとセキュリティ」のテーマごとにレポートを課し、学習の達成度を評価する。成績はレポートの評価で判定し、試験は行わない。

《教材》 教科書は特に指定しない。各テーマの最初の授業で参考書を指定する。教材として、適宜、基礎的な技術と最近の研究動向に関する資料を配布する。

《受講条件等》 計算機プログラミングの経験を仮定する。アルゴリズムとデータ構造の基礎および、データベースを学部において履修していることが望ましい。