

教員名	安達 淳 教授	研究場所	情報研	研究分野	データ工学・情報検索
-----	---------	------	-----	------	------------

対象とする研究領域は、インターネットの上で動作する大規模な分散情報システムです。なかでも、データベースや情報検索など**デジタルコンテンツ**を中心に研究してきました。現在の関心のある研究関連のキーワードを図に示します。本務である国立情報学研究所（NII）では「電子図書館システム/Digital Library」の研究開発や情報検索システムなどコンテンツの活用を目指したシステムに関する研究を軸に仕事をしてきました。

現在、東大の博士課程2人、修士課程3人の院生の他、ポスドクとともに、研究所内の同僚の先生方の協力もおおいで、以下のような研究を行っています。研究所には情報システム関係の大きなプロジェクトが複数走っており、院生も適宜参加しています。なお、ソフトウェアのみならず、ハードウェア、ネットワーク、通信システムなど広い分野に関心を持って研究に取り組もうという姿勢を期待します。

**データベースのモデルや情報空間の構築** 情報の持つ意味を的確に表現するために、どのような構造でデータベースを作ればよいかの議論です。関係データベースのような単純な構造ではなく、**Web 文書**などを念頭においた新しいアプローチがいろいろ提案されています。デジタルドキュメントの大規模データベースに関しメタデータを中軸にした柔軟な情報空間の実現を検討しています。

**フルテキストデータベース** 大量の文書をデータベースとして蓄積する場合には、効率的な検索処理、文書の構造の表現、検索の利用者インターフェースなど多方面の課題があります。**Web 文書**や**XML 文書**など現実的なデータを対象にして、効果的な処理手法を模索しています。また、電子ジャーナルや文書画像を**ORC**処理して作る大量のテキストデータベースも課題の一つです。

**情報発見・情報検索** ネットワーク上に存在する膨大な情報から効率よく情報を引き出すための新しい検索方法に関する研究を行っています。テキスト情報を対象とした研究が中心です。**新しい検索手法**の提案、多言語情報検索、記憶システムや索引システムの設計の手法、性能解析などが課題になります。このような研究では、**文書クラスタリング**、**文書分類**、**重要語や固有名抽出**や**トピック抽出**、情報の**リンケージ判定**や**同定**など、いわゆる**テキストマイニング**という分野の基礎的研究も必須です。

