情報量が増え続けるデータベースを使いこなすために：データベースからの知識発見

A course for biological knowledge discovery by handling genome-wide data

お茶の水女子大学　生命情報学教育研究センター
金子　聡子
お茶の水女子大学　大学院人間文化創成科学研究科
瀬々　潤

趣旨

実験技術の進歩によって、データ解析が研究の律速段階となりつつある。増え続ける多様かつ大規模なデータを効率よく生命現象の解明へとつなげるには、データ管理だけでなく、様々なデータの整理・統合を行い、知識発見へとつなげる「データベース (DB) 高度利用者」の役割が重要になってきている。しかし、DB高度利用者には、生命科学の知識のみならず、情報科学の技術など、多様なスキルが要求されるため、人材養成はあまり進んでいなかった。この人材の欠如に対し、統合データベースプロジェクトではDB高度利用者を養成するために必要な教育課程の策定と実施をお茶の水女子大学に委託し、人材養成を実施している。平成19年度よりDB高度利用者の知識を体系的に扱うカリキュラムを策定し、現在は、学外からの受講者も交え、講義・演習を実施している。

基礎教育

体系的な講義により情報科学と生命科学の知識を補完

a) ネットワーク基礎

コンピュータネットワークおよびインターネットの歴史と仕組み／階層プロトコルとTCP/IP／ネームサーバ／TELNETとFTP／ルーティングなど

b) データベース基礎

生物学におけるデータベースの歴史／リレーショナルデータベースについて／MySQLによる演習など

c) 現代遺伝学基礎

細胞やゲノムの構成 ／ 転写調節／翻訳／タンパクの立体構造／分子生物学実験の原理／遺伝学の歴史／進化学の概論など

専門教育

実践的な演習により生命情報のデータ利用や解析ツール開発を行う

d) ネットワークを通じたライフサイエンスデータの利用

ライフサイエンス統合データベースセンターが提供するWebAPIサービス(TogoWS)の利用とデータ処理で頻出する作業の効率化と自動化を学ぶ

UNIXの基礎／ゲノムビューアの使い方／相同性検索(Web, ローカル, EMBOSSの利用)／WebAPIの利用(Google Chart API, TogoWS(SOAP, REST)) ／相同性検索の自動化 ／統計処理の基礎など

e) データ解析の基礎

ライフサイエンスデータを扱うためのプログラム手法と基礎的なデータの統計処理を学ぶ

プログラミングの流れ／変数と型／制御文／メソッド／配列／文字列／ArrayList／JFreeChartを利用したグラフ描画／相同性検索など

f) データマイニング技術

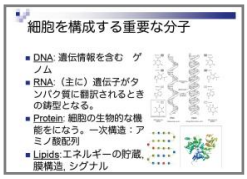
高度なデータ処理技法を学ぶ事で、新しいデータベースの創発を目指す

決定木／Naive Bayes／AdaBoost ／k-Nearest Neighbor／Support Vector Machine ／k-means／クラスタリング／ Association Ruleなど

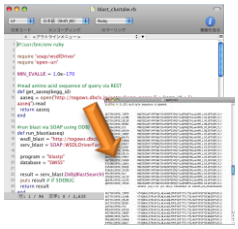
講義資料(例)

ネットワーク基礎

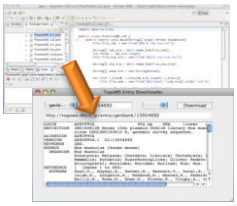
データベース基礎



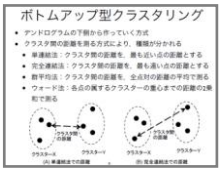
現代遺伝学基礎



ネットワークを通じたライフサイエンスデータの利用

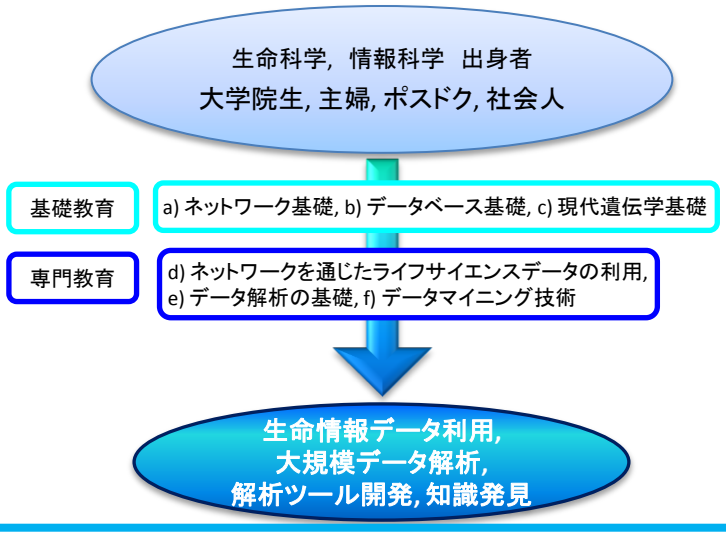


データ解析の基礎



データマイニング技術

概念図



実績、感想、進路

-実績-（延べ人数)

平成19年度	平成20年度	平成21年度	平成22年度
	基礎　14名	基礎　42名	基礎　32名
専門　18名	専門　31名	専門　65名	専門　32名

平成21年度　18名、平成22年度　14名の社会人(国立機関研究員、主婦、大学教員、民間企業社員)の方々が受講しています。

-受講者の感想-（平成22年度実施分)

・短い時間で役に立つ基礎知識を実習という形でやって頂き、社会人としては大変ありがたかったです。
・ついていくのが大変でしたが、TAの補助もあり丁寧に教えて頂いたので、何とか仕事に使うことができそうです。
・この分野(プログラミング)を一から自分で勉強するのはとてもハードルが高いので本当に助かりました。
・資料が充実していたので何とか着いていくことができました。

-受講者の進路など-

卒業生は、製薬会社やシンクタンクのSE、化粧品メーカーなどに就職しています。在学生は、統合データベースセンター内にてリサーチアシスタント（文献解析2名，統合TV開発2名）、本人材育成で公開している教材開発，サーバ管理に活躍しています。

講義風景

平成22年度　専門
(ネットワークを通じたライフサイエンスデータの利用, データ解析の基礎)

自習教材

本活動で作成した教材は、**http://togodb.sel.is.ocha.ac.jp/**にて閲覧可能です。

Licensed under a Creative Commons表示2.1日本 license
(c)2010 金子聡子　瀬々潤（お茶の水女子大学）