

統合DB支援 高度DB利用者の養成

お茶の水女子大学・説明資料

瀬々 潤
2011/02/21

平成22年度・実施状況 (目的及び本年度成果)

- 目的
 - － 蓄積された生命科学データの活用
 - － Web上の生命科学データから解析や知識発見ができる人材の育成
- 本年度当初目標と成果
 - － 講義、演習の実施
 - 目標: 社会人2名以上を含む7名以上の修了者を輩出する
 - 成果: 社会人8名(4名部分受講), 学生8名受講
 - － 自習用教材の整備・改訂
 - 講義・演習で利用した資料の配布
 - 特に分野のイントロとなる所は, 独習用資料として整備
 - － 本事業参画期間のデータ解析・DB整備等業務経験
 - 統合DBセンターの文献情報抽出に従事: 2名
 - お茶大の教材公開用サーバ・実習用サーバの維持: 2名

期間全体を通しての成果

－ 講義、演習の実施

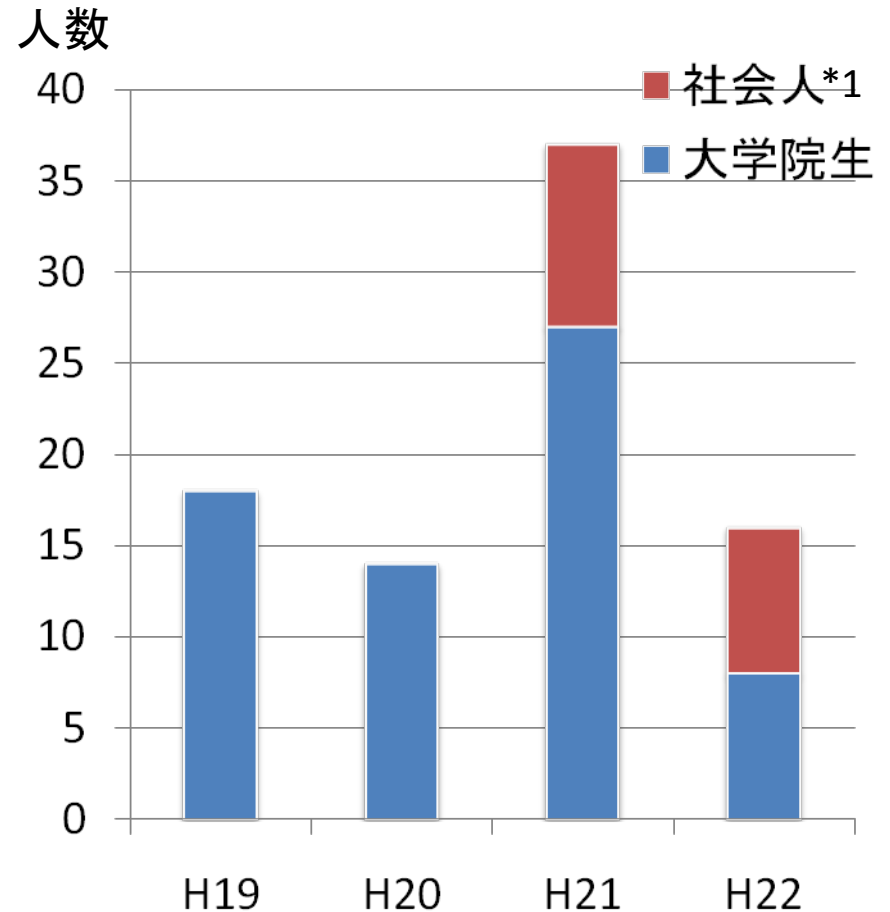
- 計93名受講
 - － 大学院生67名, 社会人18名
- 年度ごとの人数(右図)

－ 自習用教材の整備・改訂

- <http://togodb.sel.is.ocha.ac.jp/>
- 講義・演習で利用した資料の配布
 - － 特に分野のイントロとなる所は, 独習用資料として整備

－ 本事業参画期間のデータ解析・DB整備等業務経験

- 統合DBセンターの文献情報抽出に従事:2名
- 統合TV開発に従事: 2名
- お茶大の教材公開用サーバ・実習用サーバの維持:7名



*1: 部分受講生
(H21年5名, H22年4名)含む

平成22年度・実施状況

- 講義・演習の実施

- 基礎 (生物もしくは情報のいずれかの知識に不安のある人向け)

- 現代遺伝学基礎(7/26)
- ネットワーク基礎(7/28)
- データベース基礎(8/6)

- 専門

- データ収集の自動化と統計処理
 - 4/22-7/1 (毎週木曜日 16:40-18:10)
 - 8/9～12, 18
- データ処理と解析 (8/19,20,23-25)
- データマイニング技術 (毎週木曜日 10月～1月)

平成22年度・実施状況 (基礎講義・遺伝学)

● 現代遺伝学基礎(近藤るみ講師)

- 生物系以外が専門の人、あるいは、生命科学を離れていた人が生命科学関連のデータベースから知識発見するために必要最低知識の習得

分子遺伝学

遺伝情報とセントラルドグマ

分子進化

突然変異, 進化速度, アラインメントと系統樹

表現型からゲノムへ

標的遺伝子の探索, ゲノムアソシエーション



平成22年度・実施状況 (基礎講義・情報科学)

- 情報科学以外が専門の人、あるいは、情報分野から離れていた人が生命科学関連のデータベースを扱うために必要最低知識の習得
- ネットワーク基礎(小口正人教授)
 - コンピュータネットワークの歴史. ネットワークの分類. プロトコル. WWWとHTTP. ネームサーバ. IPアドレス. ルーティング.
 - ネットワーク設定実習.
- データベース基礎(渡辺知恵美講師)
 - データベースシステムの基本概念と構成. 関係モデルと関係代数. 基本的なSQL文法. 埋め込みSQLによるデータベースプログラミング.
 - MySQLサーバを用いた, SQL問い合わせ実習.



平成22年度・実施状況 (専門講義 1/2)

- 講義・演習内容

- データ収集の自動化と統計処理

- ゲノムビューアの利用とデータの利用

- ゲノム情報がWeb上に大量に存在し、DBが構築されていることのイントロダクションとして、NCBI Genome Browser, EnSEMBLの利用

- BLAST及びBLATを用いたアラインメントの実行と結果の見方

- プログラムを通じたデータの利用

- 各自のコンピュータへ、BLASTのセットアップから実行まで

- プログラムを通じたネットワーク上のデータ利用

- WebAPIの利用

- 統合データベースの提供するTogoWebServiceの利用

- 既存ソフトウェアとの組み合わせによるデータの利用

- 基礎統計の導入. 統計処理ソフトRによる実行.

- 進化距離の計算、分子時計のグラフ作成

平成22年度・実施状況 (専門講義 2/2)

– データ処理と解析

- Java言語の習得
- 逆相補鎖を求めるプログラムの作成
- 塩基配列の頻度計算と可視化
- 相同性領域の発見と可視化
- 進化系統樹の計算と可視化



平成22年度・実施状況 (専門講義 2/3)

– データマイニング

- クラス分類手法の導入
 - 決定木、ナイーブベイズ、ブースティング、SVM(サポートベクターマシン)
- クラスタリング
 - K-meansクラスタリング、SOM(自己組織化マップ)、階層型クラスタリング
- 相関ルール
 - Apriori
- 統計処理ソフトRを用いた各アルゴリズムの実習
 - 医療診断例としてコンタクトレンズ処方データ
 - 画像診断の基礎問題として数字画像認識

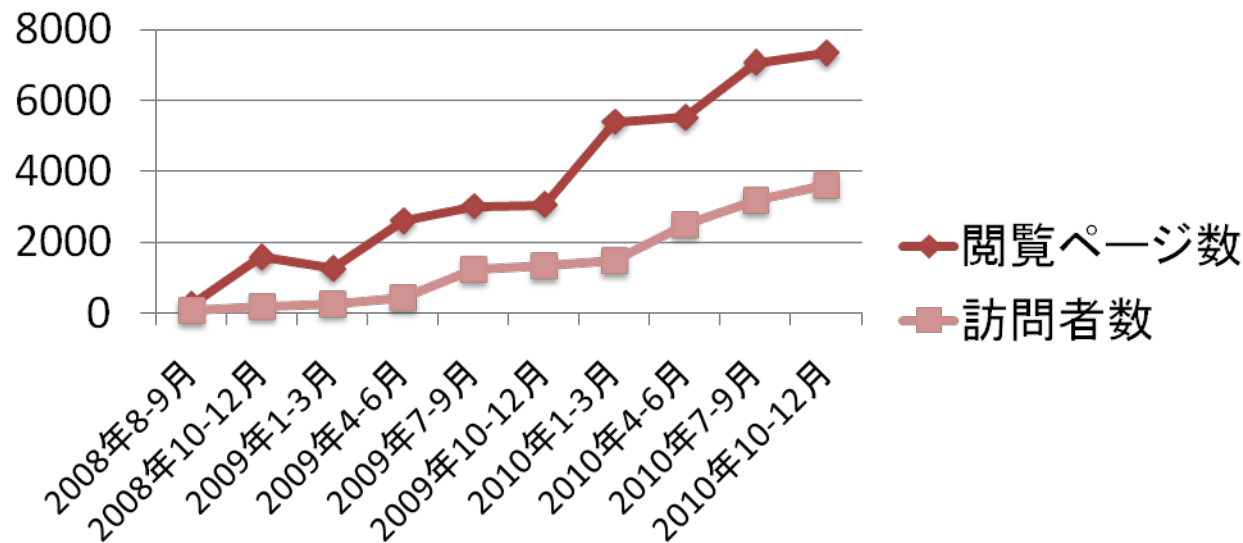
授業資料の配付

- 授業資料及び資料に自習用のテキストを加えたものをWebから公開しています。

- <http://togodb.sel.is.ocha.ac.jp/>

– 閲覧人数及びページ数

- まだ飽和しておらず、今後も需要の高まりが予想される



今後の維持態勢，問題点

- 講義・演習の維持
 - － 全学の大学院生が取れる正規授業となっており，継続予定。
 - － お茶大・医科歯科・北里大・学習院大の4大学で作る「学際生命科学東京コンソーシアム」中で単位互換ができる科目に含まれているため，お茶大のみならず，他大学からの受講も可能な態勢を整えている。
- 教材の公開に関して
 - － スライド等は大学で保持せず，slideshare社開発のサイトslideshareに登録することで，大学とは独立して公開継続可能
- 育成した学生及び社会人の活躍について
 - － 学生の多くは，一般企業へ就職し統合DB事業に関わりが薄い。統合DB内でインターンシップの様な形態を作り，継続的に人を育成し，バイオインフォマティクス分野で従事してもらうシステムも必要だろう。
 - － 企業や研究所から受講した社会人は，バイオインフォマティクス関連，あるいは，活用する部署に勤めており，知識を得る場として活用して頂けた。
 - － 子育て等で企業や大学から離れていた方が，企業や大学へと戻ってくるための機会として受講した方も数名おり，中には能力の高い方もいた。統合DBプロジェクトだけでなく，日本経済の為にも，このような受講生を活用する態勢を整える必要がある。