CS Bioinformatics 2016

Dr. E.M. Bakker | LIACS

6-9-2016

Universiteit Leiden
The Netherlands

Discover the world at Leiden University

Programme

14.00 Welcome in Leiden
14.15 Overview CS Bioinformatics Program
14.45 Graduating in Leiden

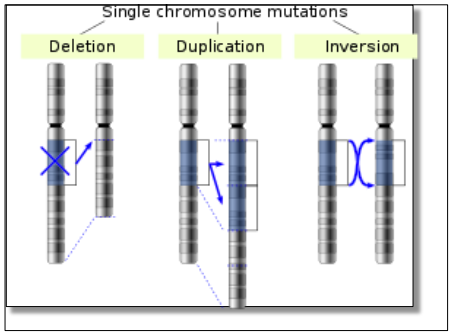
15.00 Break
15.15 Talks by (former) CS Bioinformatics and CS Students.
- *Reconstructing the subclonal evolution of tumors from targeted sequencing data.* Marleen Nieboer
- *Data analysis for the MinION nanopore sequencer.* Michael Liem
- *Gene Prediction Using Unsupervised Deep Networks.* Dimitris Sevastakis

16.15 Break

16.30 Required Knowledge and Skills
17.00 Closing

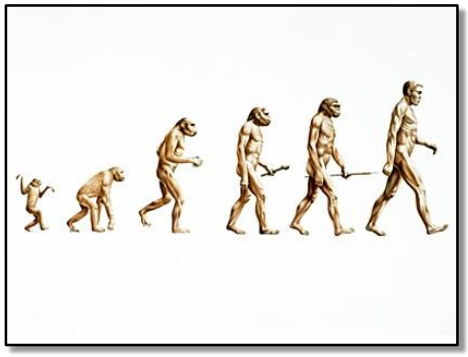
Discover the world at Leiden University

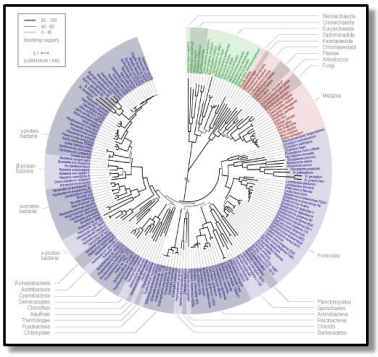
Understanding disease



3

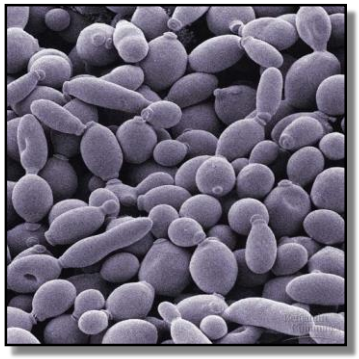
Understanding evolution



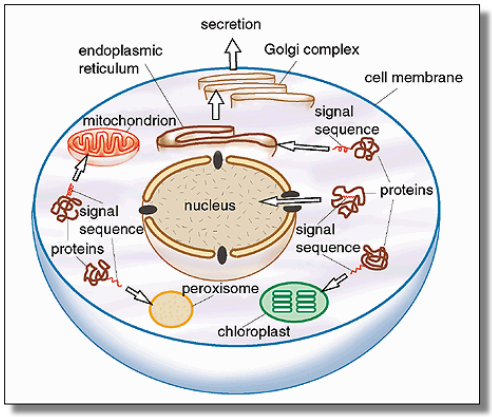


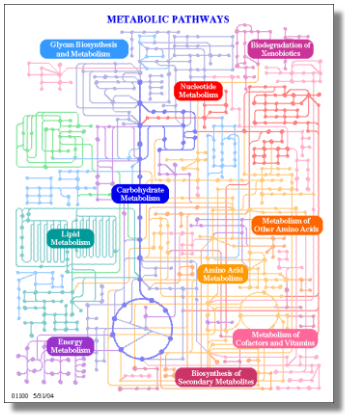
4

Understanding cell factories

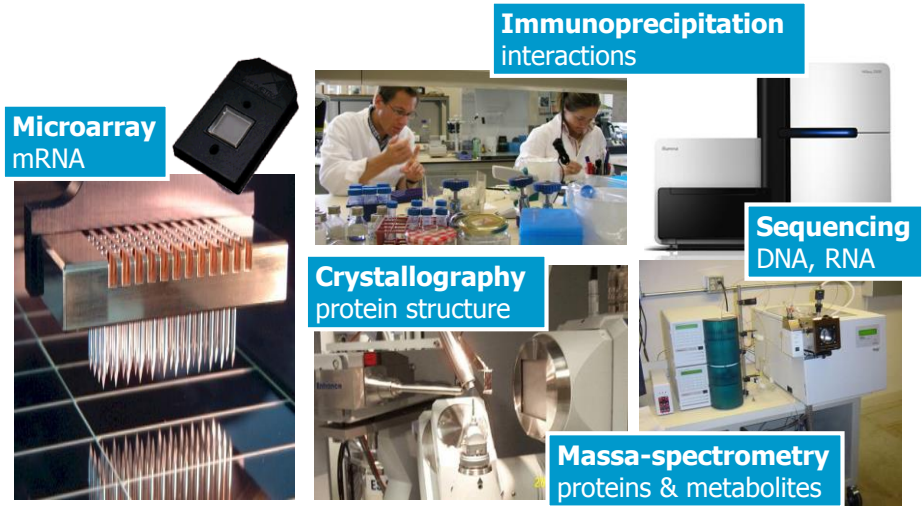


Understanding life of the cell



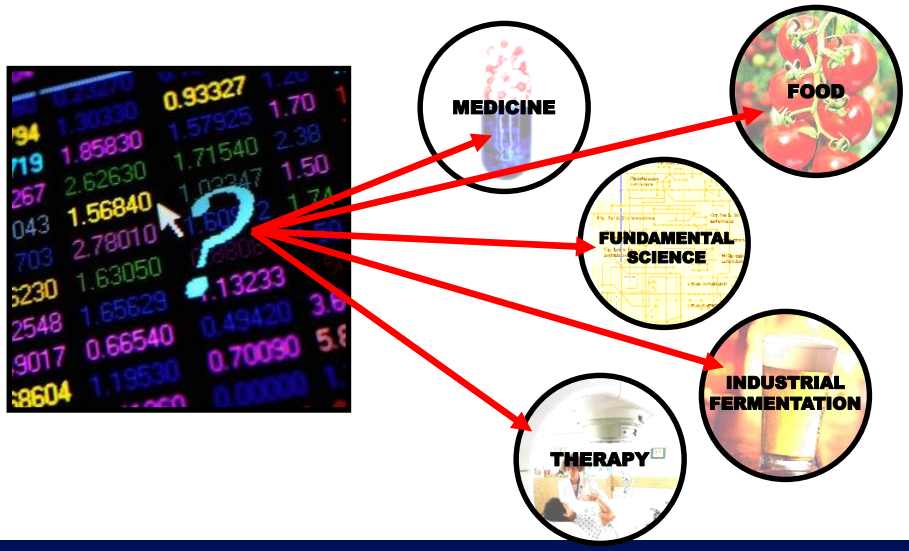


Measuring molecules in the cell

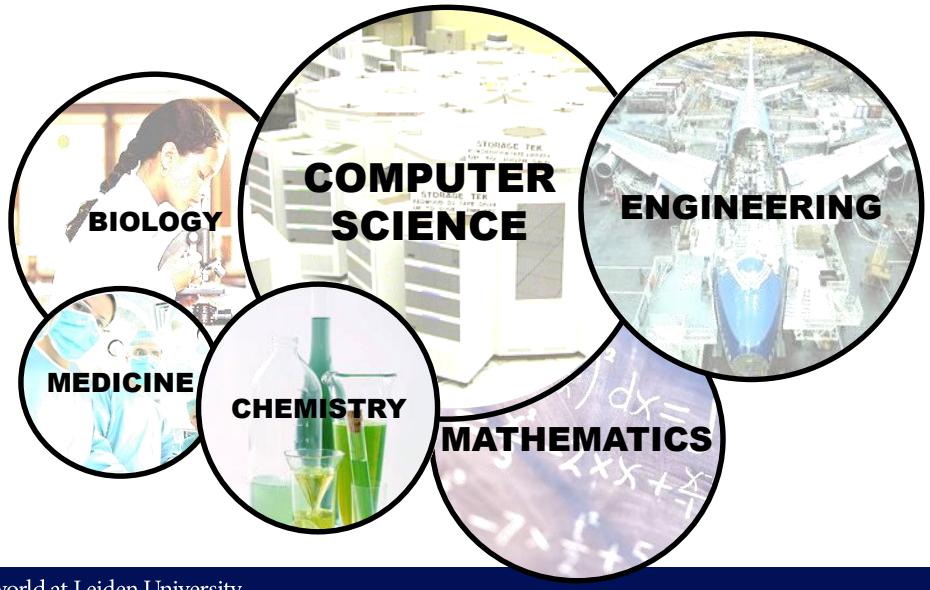


7

Bioinformatics: organize and interpret the data



A marriage of disciplines



Organization

- Bioinformatics is a track in MSc "Computer Science"

Universiteit Leiden

TU Delft

Cancer GENOMICS CENTRE The Netherlands Cancer Genome Network

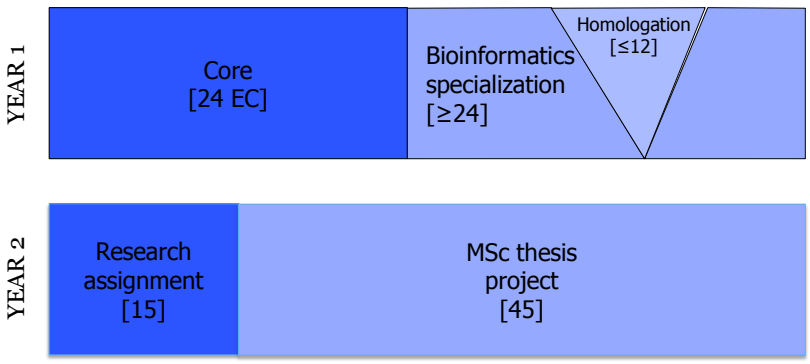
Kluyver CENTRE Kluyver Centre for Genomics of Industrial Fermentation

Educational programme

- General part
  - Core program
  - Homologation courses / free electives
- Specialization part
  - Bioinformatics
  - Courses from other programmes
- Research assignment
- MSc Thesis project

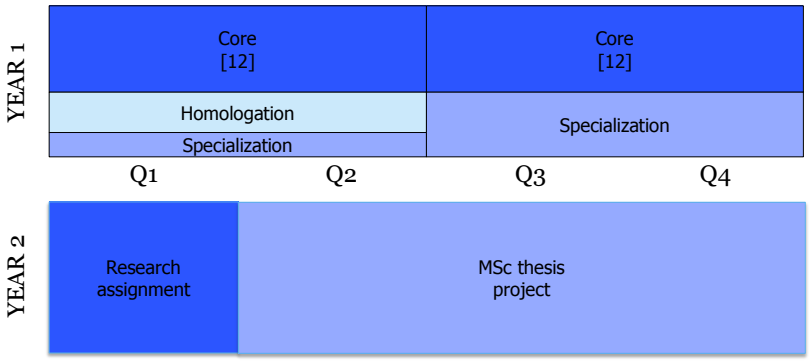
Educational programme

- Overview:



Educational programme

- Timing:



13

Educational programme

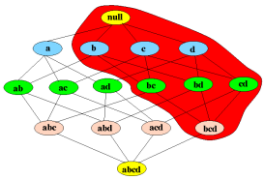
- General part
  - **Core program**
  - Homologation courses / free electives
- Specialization part
  - Bioinformatics
  - Courses from other programmes
- Research assignment
- Thesis project

[24 EC]

Pattern Recognition	[6]
Databases and Data mining	[6]
Computational Molecular Biology	[6]
Functional Genomics and Systems Biology	[6]

14

IN4170 Databases & data mining

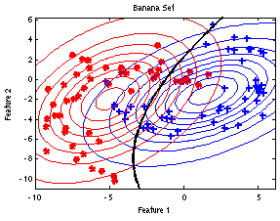


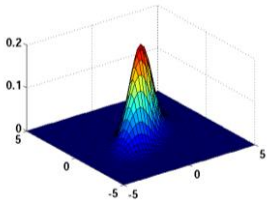
During the course we will discuss the following advanced database topics: the evolution of database technology, data preprocessing, data cleaning, data warehouses, OLAP, data cubes. Advanced data mining topics that will be handled during the course are: mining stream, time series and sequence data, graph mining, mining object, spatial, multimedia, text, and web data.

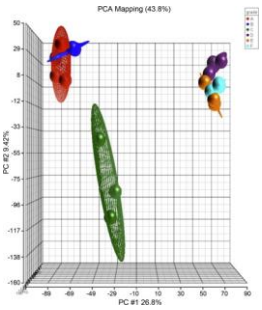
15

Discover the world at Leiden University

IN4085 Pattern Recognition





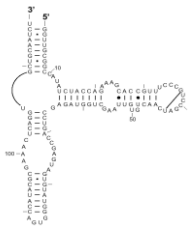


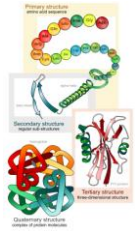
Recapitulation of multi-dimensional statistics, data visualisation, density esimation, cluster analysis. Representation of real world objects by features, prototypes and dissimilarities. Training pattern classifiers by examples. Feature extraction. Bayes' rule. Classification by statistical discriminants, neural networks, decision trees or support vector machines. Statistical learning theory. One-class classifiers. Combined appraoches. EM algorithm. Partially supervised learning. Evaluation procedures, cross validation Overtraining, regularisation.

16

Discover the world at Leiden University

IN4173 Comp. Molecular Biology

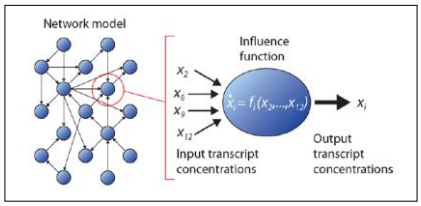


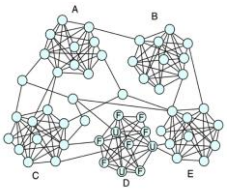


- *Algorithms: Pair wise alignment; Phylogeny; Physical Mapping; Gene Finding; Genome Rearrangements; DNA Chips and Gene Networks; Sequence Alignment Heuristics; Multiple Sequence Alignment; Bioinformatics Tools; Linkage Analyses.*
- *Protein structure prediction: homology modeling; fold recognition; knowledge-based potentials; ab initio structure prediction.*
- *RNA structure prediction: energy minimization; folding simulations; comparative analysis; non-canonical base pairs; 3D-modeling; RNomics; search for RNA and RNA motifs in genomic sequences.*

17

IN4176 Functional genomics & Systems biology





The goals and methodology of systems biology will be discussed first, followed by a brief overview of the most important types of biological measurements used. Networks (the main data type used) and their properties (small-worldness, motifs) will then be discussed, as well as a number of network models often used: linear, Boolean and Bayesian. The last half of the course will show how various -omics levels can be described using these models. Finally, a few examples of fully integrated models will be given.

18

Educational programme

- General part
  - Core program
  - **Homologation courses / free electives**
- Specialization part
  - Bioinformatics
  - Courses from other programmes
- Research assignment
- Thesis project

[at most 12 EC]

Introduction to life science, statistics or computer science

19

Educational programme

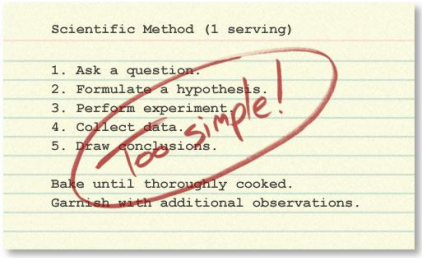
- General part
  - Core program
  - Homologation courses / free electives
- Specialization part
  - **Bioinformatics**
  - Courses from other programmes
  - Research assignment
- Thesis project

[at least 24 EC]

Image Analysis in Microscopy	[6]
Multimedia Information Retrieval	[6]
Mathematical Biology, Virtual cell	[6]
Mathematical Biology, Metabolic network	[6]
Bio-modeling and Petri-Nets	[6]
Optimization (Swarm-based Computation with Applications in Bioinformatics	[6]
Advanced Bioinformatics	[4]
Advanced Digital Image Processing	[6]

20

WM0332IN Methodology of Science & Engineering



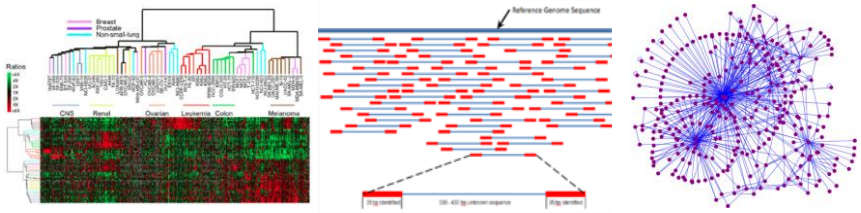
This course introduces the student to:

1. *the aims and character of science and engineering as human practices;*
2. *the distinction between facts and values or norms, and the role they play in science and engineering;*
3. *ways of arguing in support of factual and in support of normative claims;*
4. *aspects of empirical research: induction, deduction, measurement, evaluation;*
5. *the character and status of theories and models in science and engineering;*
6. *similarities and differences between the natural sciences and the human or social sciences.*

21

Discover the world at Leiden University

IN4329 Adv. bioinformatics



After successfully completing this course, the student is able to:

- *understand several high-throughput experiments, such as microarrays, and next generation sequencing, and discuss the benefits and limitations of these methods*
- *comprehend the statistical and computer science issues in analyzing high-throughput data*
- *discuss the basic systems biology approach, and the role of*
- *high-throughput measurements, gene selection and classification therein*
- *read and comprehend a current paper on systems biology*

22

Discover the world at Leiden University

ET4283 Adv. digital image processing



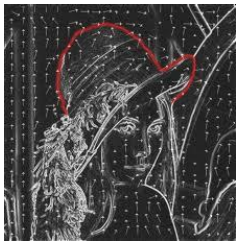
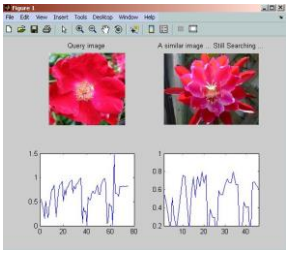


Image restoration (inverse filtering, Wiener filtering, geometric transformation), advanced morphological image processing and extension to grey-scale images, data-driven image segmentation (boundary detection, region-based segmentation, watersheds), model-based image segmentation (Hough transform, template matching, deformable templates, active contours), representation and description of image objects, image features (structure tensor, local shape), camera calibration (intrinsic and extrinsic parameters, projection matrix), stereopsis (correspondence, epipolar geometry, essential and fundamental matrix), motion estimation (optical flow, feature-based techniques)

23

Discover the world at Leiden University

IN4174 Multimedia inf. retrieval



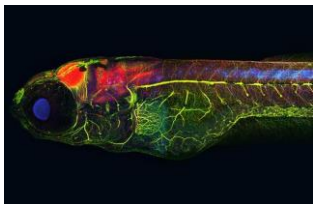
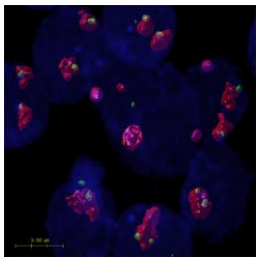
Google

Extending beyond the borders of culture, art, and science, the search for digital information is one of the major challenges of our time. Digital libraries, bio-computing & medical science, the Internet, streaming video, databases, cultural heritage collections and peer-2-peer networks have created a worldwide need for new paradigms and techniques on how to browse, search, and summarize multimedia collections. This course focuses on the area of searching and retrieving multimedia information from digital databases and collections. Examples of multimedia would be X-Ray and MRI scans, general photos, and video.

24

Discover the world at Leiden University

IN4395 Image analysis in microscopy

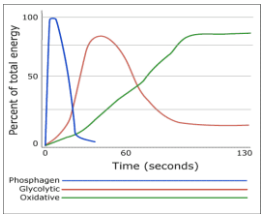




In a series of lectures all important aspects of imaging along the line of the characteristic sequence of image analysis are dealt with. Concepts of image processing will be introduced and it will be discussed how set of image features is compiled in measurements. Subjects will use the 2D imaging as a means of explaining the principles and the switch to multi-dimensional imaging to illustrate the implications of imaging in research and connect to current topics in bio-medical research. Presenting results through visualization and modeling is an ingredient found in applications that are discussed.

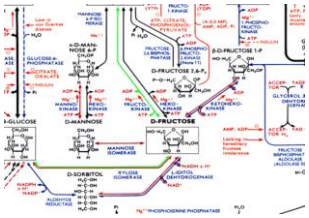
25

Discover the world at Leiden University

IN4322 Mathematical biology – Metabolic network analysis



2014-2015

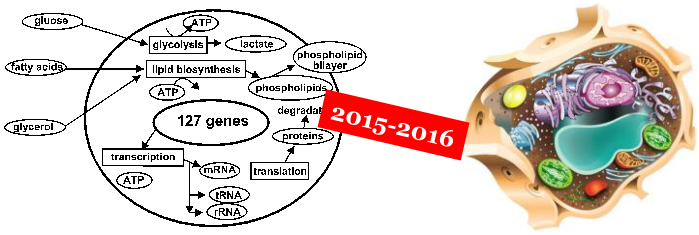


The course discusses the mathematical modelling of large biochemical networks, metabolic networks in particular, and the subsequent contrained-based analysis of their dynamic properties. We introduce the fundamental concepts of the stoichiometric matrix and flux vector and show what information can already deduced from the first, e.g. concerning possible steady state flux vectors for the system: extreme currents, extreme pathways, elementary modes and the relationships among them. Several algorithms will be explained for computing them together with software packages that implement these (e.g. FluxAnalyzer). The concepts are applied to the problem of optimal metabolite production for a model organism. This is of importance in the production of e.g. pharmaceuticals in plant cell cultures or bacteria. If time permits, parametric sensitivity is discussed.

26

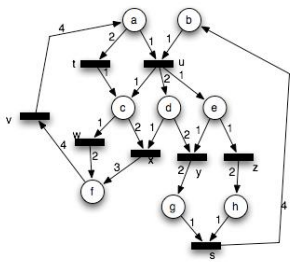
Discover the world at Leiden University

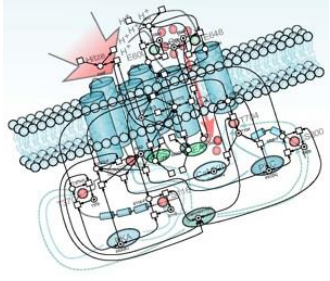
IN4177 Mathematical biology – Virtual cell



In this course it is shown how methods from statistics and analysis can be used to design and analyze biological models. For highly complex systems, such as those encountered in the living cell, it is often not possible to give an adequate system description of the dynamics of the different sub-processes. However, different mathematical techniques exist that can give insight in the dynamics of those systems using concrete data sets. These mathematical techniques are also used in the Virtual Cell, a computational environment designed for the construction of cell biological models and simulations.

IN4396 Biomodeling & Petri nets

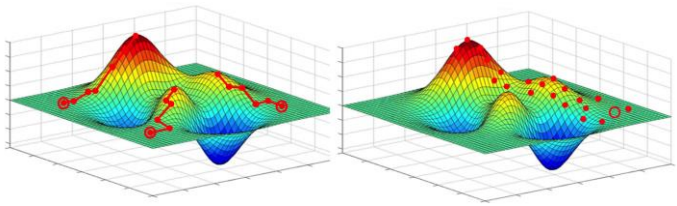




This course aims at making students aware of formal, mathematically precise, approaches to the faithful modeling of biological processes. We will investigate the possibilities provided by the framework of Petri nets and computer tools for the analysis and representation of developmental processes. We will use biological case studies to clarify and practice construction of Petri nets for biology.

IN4178 Optimization

(Swarm-based Computation with Applications in Bioinformatics)



- Swarm intelligence is a modern artificial intelligence discipline that is concerned with the design of multi-agent systems with applications. Swarm intelligence is embedded in the biological study of self-organized behaviors in social animals, e.g., the collective behaviour of social insects such as ants and bees, as well as flocks of birds and schools of fish. Instead of a sophisticated controller that governs the global behaviour of the system, the swarm intelligence principle is based on many unsophisticated entities that cooperate in order to exhibit a desired behaviour. For example, without any master blueprint bees are able to build complex hive in cooperation.

29

Educational programme

- General part
  - Core program
  - Homologation courses / free electives
- Specialization part
  - Bioinformatics
  - **Courses from other programmes**
  - Research assignment
- Thesis project

From other tracks or MSc programmes in:
- Electrical engineering
- Mathematics
- Computer Science
- Life Science & Technology
- LIACS

30

Other Specializations (Leiden)

- Bioinformatics
- Advanced Data Science
- Algorithms and Software Technology
- Computer Systems, Imagery and Media

Other Specializations (Leiden)

Selected Courses

- Evolutionary Algorithms
- Complex Networks
- Multicriteria Optimization and Decision Analysis
- Social Network Analysis for Computer Scientist
- Neural Networks
- Bayesian Networks
- Advances in Data Mining
- Quantum Computing

Other specializations (Delft)

- Life Science & Technology
- Computer engineering
- Electrical Engineering

- Applied mathematics
  - WI4201 Scientific computing
  - WI4207 Continuous optimization
  - WI4219 Discrete optimization
  - WI4201COSSE Scientific computing (COSSE)
  - WI4017 Parallel computing
  - IN4049TU Introduction to high performance computing
  - …

Other specializations (Delft)

- CS – Information architecture
  - IN4324 Web & semantic web engineering
  - IN4331 Web data management
  - …

- CS – Software technology
  - IN4150 Distributed algorithms
  - IN4301 Advanced algorithms
  - IN4026 Parallel algorithms and parallel computers
  - IN4389 Complex networks from nature to man
  - …

Other specializations (Delft)

- CS - Media & knowledge engineering
  - IN4086 Data visualization
  - IN4144 Data science
  - IN4320 Machine learning
  - AP3231 Medical imaging
  - …

Educational programme

- General part
  - Core program
  - Homologation courses / free electives
- Specialization part
  - Bioinformatics
  - Courses from other programmes
- **Research assignment**
- Thesis project

[15 EC]

Research assignment
at company/university internship

Educational programme

- General part
  - Core program
  - Homologation courses / free electives
- Specialization part
  - Bioinformatics
  - Courses from other programmes
  - Research assignment
- **Thesis project**

[45 EC]

In collaboration with biological partners:
- medical domain
- biotechnology domain
- food domain

37

Graduation at Leiden University

Erwin M. Bakker

Universiteit Leiden
The Netherlands

Second Year

- Research Project (15 EC)
  - At Leiden University
  - At TU Delft
  - At research institutes, etc.
- Master Thesis (45 EC)
  - idem
- Graduation
  - At Leiden University

Requirements

- Topic related to Bioinformatics
- (Challenging) Scientific Research
- Research plan

Important:

Always at least one supervisor from LIACS

- responsible for approval of research plan and final grading.

Deliverables

Research Project

- Research Report
- Software and documentation (if any)

Master Thesis

- Research Report
- Software and documentation (if any)
- Final presentation (+ defense)

Assessment and Grading Master Thesis

- At least two assessors
- Assessment Form

Supervisors at LIACS

Research Groups

- Algorithms and Software Technology (AST)
- Computer Systems, Imagerus & Media (CSI)

The Foundations of Software Technology

prof.dr. Farhad Arbab, prof.dr. Joost N. Kok

The Algorithms and Software Technology (AST) program performs fundamental research in the areas of algorithm design and analysis with an emphasis on algorithms and architectures for mining large data volumes as well as on natural computing.

Furthermore, focus is on the development of formalisms, methods, techniques and tools to design, analyze, and construct software systems and components.

Applications in medicine, bio- and chemoinformatics, engineering, and physics.

High Performance Computing prof. dr. Harry A.G. Wijshoff

Research focus:

- Large-scale applications, grid computing, problem solving environments
- Parallel and distributed computing, optimizing compiler technology, embedded software development
- Large-scale database systems, data compilation, data integration, and data mining

Imagery & Media

Imaging and BioInformatics Dr.ir. Fons J. Verbeek The research focus of this group is on bio-imaging and integration of the analysis of images and image information with other bio-molecular information resources; in addition we study new ways of interacting with these data.

Media Research Dr. Michael S. Lew

The Media Research (MR) Group is concerned with the scientific investigation of novel directions and paradigms in the interaction and understanding of diverse media, such as images, video, and audio.

Date of Graduation

- The formal date of graduation is the date on which you have met all the requirements of the MSc in Computer Science program. This date is recorded in your transcripts and also on your diploma.

- This date coincides with the date on which you have passed the last required component of your approved program of study.

Graduation Procedure

Students, who are close to graduation, must apply for the Master's Graduation 5 weeks before the graduation date.

If you are planning to graduate in August or September, you should apply for your Masters Graduation before **1 July.**

Important:
Application for your Masters Graduation is done by sending an e-mail to E.M. Bakker (Programme Director).

Programme

14.00 Welcome in Leiden
14.15 Overview CS Bioinformatics Program
14.45 Graduating in Leiden

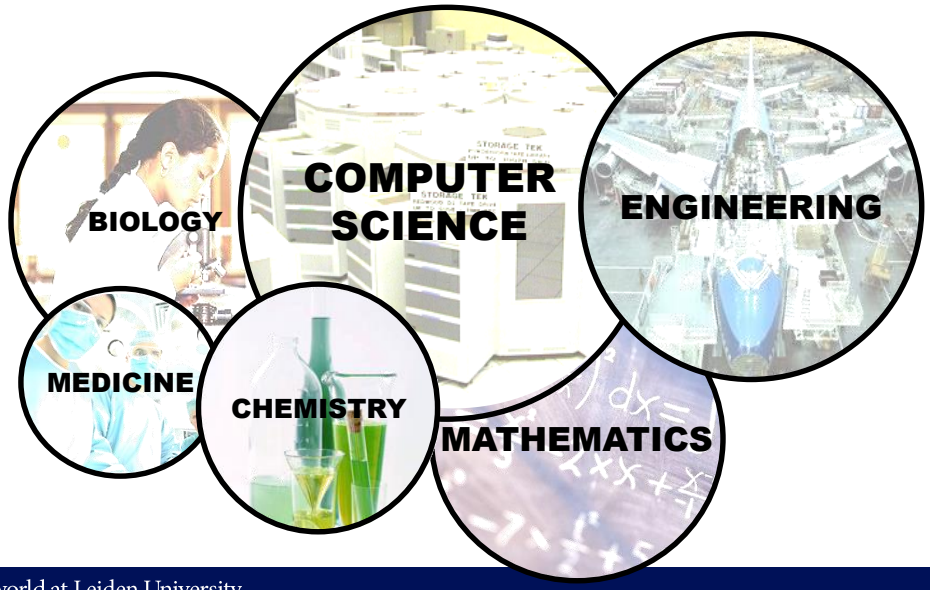
15.00 Break
15.15 Talks by (former) CS Bioinformatics and CS Students.
- *Reconstructing the subclonal evolution of tumors from targeted sequencing data.* Marleen Nieboer
- *Data analysis for the MinION nanopore sequencer.* Michael Liem
- *Gene Prediction Using Unsupervised Deep Networks.* Dimitris Sevastakis

16.15 Break

16.30 Required Knowledge and Skills
17.00 Closing

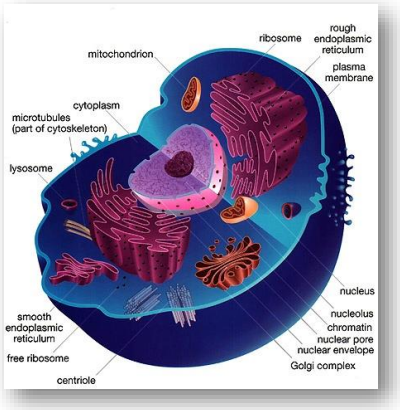
50

A marriage of disciplines



Life science

- Universal features of life and evolution; cell biology, elementary biochemistry and biosynthesis; protein shape, structure and function; RNA; DNA and chromatin structure; DNA replication; transcription and translation; control of gene expression;membrane structure



52

Computer science

- Programming (high-level language, object oriented)
  - data types, operators, expressions, type conversions, declarations, assignments; conditional and repetition statements
  - class specification and implementation; inheritance, type casts, redefinition, dynamic binding
  - use of libraries (APIs); implementing, testing, debugging
- Elementary algorithms and data structures
  - searching and sorting, recursion, backtracking; iterators; time and space complexity
  - arrays, linked lists, stacks, queues, ordered structures, binary (search) trees, priority queues, hashtables and graphs

Mathematics

- Calculus:
  - limits, differentiation, integration, differential equations, complex numbers, series and sequences, multivariate functions
- Linear algebra:
  - vector and matrix algebra, solving linear equation systems, linear projections, subspaces, least squares methods, determinants, eigenvectors and eigenvalues, diagonalising
- Probability and statistics:
  - combinatorics, conditional probability, Bayes' theorem, probability density functions and cumulative distribution functions; CLT; expected value, variance, moments; multivariate distributions, covariance, correlation; estimators, confidence intervals; hypothesis testing

Tutor Molecular Genetics
(aka Introduction to Life Science) (10 EC)

Format:

Self-study, problem and discussion sessions starting at the beginning of November 2016.

Materials:

B. Alberts et al., 'Molecular Biology of the Cell', 6th edition
Publcation Date: November 18, 2014
ISBN-13: 978-0815344322
ISBN-10: 0815344325

Hereby a provisional list of the materials to be (self-)studied.
Ch 1. all
Ch 2. all, except pp 75-78 and 118-119
Ch 3. all
Ch 4. all
Ch 5. pp 263-304
Ch 6. pp 331-400
Ch 7. pp 411-454
Ch 10.pp 617-636
Please note, some extra material will be added later.

It is expected that you can and will study these materials
mainly by yourself. You can start as soon as you have the book.

Examination: Final Exam

Tutor Computer Science (10 EC)

Description: The tutor Computer Science will consist of a tailor made program that depends on your previous studies. As a general rule, it may consist one or more of the following three parts:

Programming in C++

- Assignments and exams will be similar to the (Dutch LIACS) course 'Programmeermethoden'
- Book used for this part is: W. Savitch, Absolute C++, fourth edition, Addison-Wesley, 2009.
- Slides and code, see: ftp://ftp.awl.com/cseng/authors/savitch/cpp4e

Algorithms and Data Structures

- Book used for this part is: T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to Algorithms, Third Edition, MIT Press / McGraw-Hill Book Company, 2009. (Click the link for videos, slides and further information.)

Databases

- This course will be given during the second semester in English during the first 2 hours on Mondays.
- Book: Ragu Ramakrishnan, Johannes Gehrke: Database Management Systems, McGraw Hill, third edition, paperback version; ISBN: 0071151109; 936 blz.;
- See also the website: https://studiegids.leidenuniv.nl/courses/show/29451/databases

Format:

Self-study, discussion and problem-sessions. Please note: Available lectures can be followed only if the schedules and course-language allow this.

Examination: Assignments and Final Exam (for some of the parts)

Tutor Statistics (10 EC)

Description: The tutor Statistics will consist of a tailor made program that depends on your previous studies.

Book:

[1] J.A. Rice, Mathematical Statistics and Data Analysis 3rd Edition (Paperback)

Further Materials:

- C.M. Grinstead, J.L. Snell, Introduction to Probability, 2nd Edition, American Mathematical Society, 2006.
- Slides used during the lectures.

Format:

Self-study, discussion and problem-sessions. Please note: Available lectures can be followed only if the schedules and course-language allow this.

Examination: Final Exam

Tutor Statistics (10 EC)

Subjects:

- Introduction to Discrete Math and Calculus
- Discrete Probability Distributions
- Continuous Probability Densities
- Combinatorics
- Conditional Probability
- Distributions and densities
- Expected Value and Variance
- Sums of Random Variables
- Law of Large Numbers
- Central Limit Theorem
- Hypothesis Testing
- Selected Topics from: Survey Sampling, Parameters Estimation, Hypotheses Testing, Summarizing Data, Comparing Two Samples

Discover the world at Leiden University

Registration at TU Delft

This is to inform you about the registration procedure
that you have to follow in order to obtain the TUDelft
blackboard account:

1) Register through 'studielink' as a (bijvak-)student at TUDelft
2) Obtain a proof of payment for your study at the
student-administration at Leiden University
3) Send or deliver the proof of payments to Prof. dr. ir M.J.T. Reinders at TUDelft
4) Prof. dr. ir M.J.T. Reinders will inform the student administration at TUDelft.
5) After this you will obtain a TUDelft blackboard account that you need
to activate at the TUDelft student administration

Note: this of course assumes that you completed your registration at
Leiden University first.

If you have any questions or encounter any problems please contact me.

Discover the world at Leiden University

Questions?

erwin@liacs.nl

Universiteit
Leiden
The Netherlands

Discover the world at Leiden University