



In-Memory Technology for Life Sciences Organization and Seminar Details

Cindy Fähnrich,
Dr. Matthieu-P. Schapranow,
Dr. Mariana Neves, and
Dr. Matthias Uflacker

April 8, 2014

- Seminar Organization
- Seminar Topics
- Introduction Analyze Genomes Project

Seminar Organization Setup

- Supervisors:
 - Cindy Fähnrich,
 - Dr. Matthieu-P. Schapranow,
 - Dr. Mariana Neves,
 - Dr. Matthias Uflacker

- Location: HPI Campus II, Room D.E-9/10 (former SNB)
- When: Tuesdays and Wednesdays 9:15-10:45 a.m. (s.t.)
- Periods: 4 SWS
- Credits: 6 graded ECTS
- Enrollment: until April 17, 2014
- Website:
<https://epic.hpi.uni-potsdam.de/Home/InMemoryForLifeSciencesSoSe2014>

Seminar Organization

What you can expect from us

- Broaden your horizon in the fields of
 - in-memory technology,
 - life sciences, and
 - your project's research topic
- Get in touch and work with real-world data
- Work together with experts from industry / research
- Work with latest beta hard-/software resources, e.g. in the Future-SOC laboratory at HPI
- Get experienced in collaborative project work
- Improve your presentation skills (English)
- Enhance your skills in scientific working and writing



<http://i.kinja-img.com/gawker-media/image/upload/s--cREIB5AZ--/1865smw5hbtt6jpg.jpg>

Seminar Organization

What we expect from you

- Commitment on your selected research topic
- Perform autonomously research to acquire required knowledge about your selected research topic
- Work together in interdisciplinary teams
- Participate in seminar meetings
- Systematic use of software design and engineering methods
- Contribute with your expertise also to your colleagues / other teams
- Update supervisors regularly on your progress / issues
- Handle sensitive data, e.g. from partners, confidentially



<http://i.kinja-img.com/gawker-media/image/upload/s--cREIB5AZ--/1865smw5hbbt6jpg.jpg>

Seminar Organization

Grading

- The grading of the seminar works as follows (Leistungserfassungsprozess):
 - 40%: Seminar results, e.g. prototype, and research article
 - 40%: Methodic research approach and individual commitment
 - 20%: Intermediate and final presentations
- All individual parts have to be passed to pass the complete seminar



http://www.hpi.uni-potsdam.de/fileadmin/hpi/presse/Fotos/campus_und_gebaeude/20111017_HPI_Hoersaal.jpg

Seminar Topics

Overview

- Clinical Trials Recruitment Process
- Search and Information Extraction in Unstructured Medical Documents
- Identification of Mutations in Man and Mouse
- Analysis of Medical Side Effects
- Real-time Analysis of Sensor Data for High-risk Patients
- Evaluating Influence Factors for Identifying Genetic Variants
- Annotating Genome Data with Context-specific Information

Seminar Topics

A: Clinical Trials Recruitment Process

Issue:

Identification of suitable therapies for chronic diseases is a time-consuming and complex task.

Idea:

- Work in interdisciplinary teams with our partner Cytolon AG in Berlin
- Support interactive identification of suitable clinical trials for chronic disease patients and vice versa
- Provide tools for automatic assessment of structured and unstructured patient data from clinical systems
- Create together with experts from industry a working research prototype on in-memory technology (mobile app or web app)
- You will gain access to real-world patient data that you are working with, e.g. design of database schemas and data import/export
- Identify relevant criteria and define similarity metrics for patient data together with medical experts
- Apply clustering to patient data based on your similarity metric



<http://pubs.acs.org/subscribe/archive/mdd/v04/i01/figures/clinical.gif>

Seminar Topics

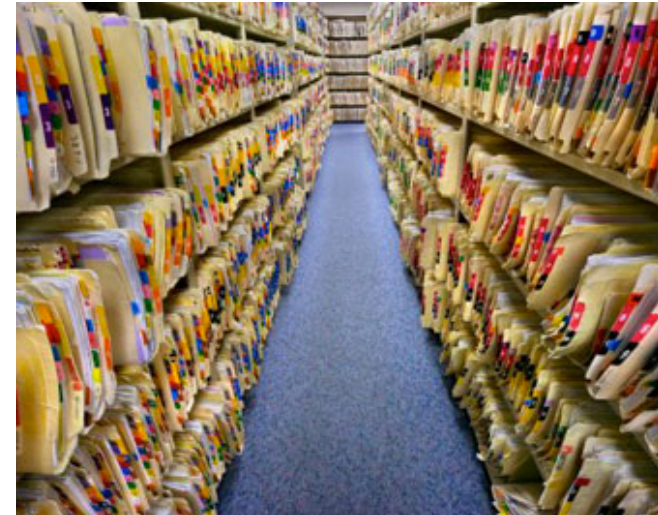
B: Search and Information Extraction in Unstructured Medical Documents

Issue:

Many medical documents are stored as unstructured texts. Thus, identification and assessment of relevant content becomes a time-consuming task binding medical experts.

Idea:

- Applying text extraction methods of in-memory technology
- Enable users to conduct an interactive Google-like search to obtain relevant results for their work/research
- Identify and extract specific entities from medical documents, e.g. drugs, diseases, and medications.
- Identify and extract relationships between these entities, e.g. side effects for drugs, attributes related to a certain disease
- Develop a ranking metric to list documents according to their particular relevance for a search in a web application



<http://static.ddmcdn.com/gif/hippa-4.jpg>

Seminar Topics

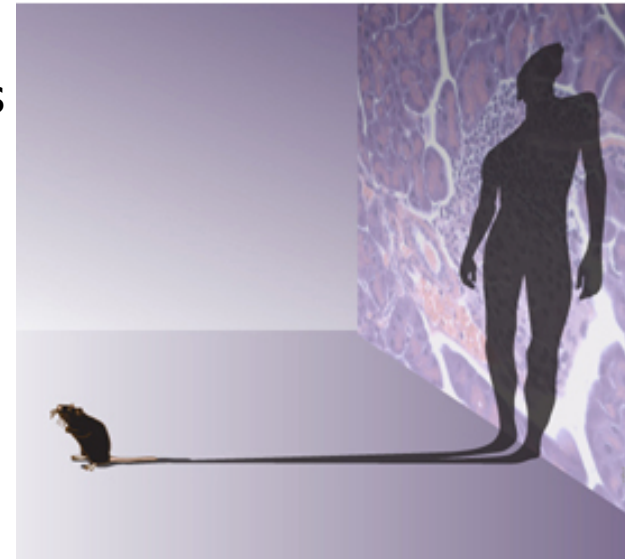
C: Identification of Mutations in Man and Mouse

Issue:

To obtain knowledge about genes and their functions in human beings, a variety of research is conducted in Xenograft models. However, the results cannot be applied 1:1 to human beings.

Idea:

- Enable researchers to transparently access knowledge from other experiments, e.g. Xenograft models, when investigating genetic variants
- Combine heterogeneous data sources about known variants for different species, e.g. Mouse Genome Database (MGD), to enable a holistic view on genetic variants
- Identify relevant data sources and integrate them into an in-memory database to enable interactive analysis
- Design and implement associations between the individual data sources in a web application



http://hms.harvard.edu/sites/default/files/assets/Shay_mouse-and-man_360.jpg

Seminar Topics

D: Real-time Analysis of Sensor Data for High-risk Patients

Issue:

Intensive-care patients require 24/7 monitoring, e.g. on ICUs. However, it is known that recovery is accelerated when people are hosted in their used environments, e.g. at home. At the same time, an emerging market of wearable sensors exists.

Idea:

- Evaluate wearable sensors, e.g. accordingly to their technology, functionality, programmability
- Use wearable sensor technology to monitor health constitution of real-world persons and obtain wave forms
- Design and implement an in-memory database prototype to regularly acquire data from selected wearable sensor devices (web / mobile app)
- Analyze wave form patterns to identify emergency events in advance, e.g. with machine learning or pattern recognition



<http://www.technologyreview.com/sites/default/files/images/ces.monitorsx519.jpg>

Seminar Topics

E: Analysis of Medical Side Effects

Issue:

The interaction of different medical substances at the same time may result in reduced activity of active ingredients or side effects in patients. Today, national databases exist that document patient cases and observed side effects.

Idea:

- Enable researchers and physicians to instantly analyze and explore existing side effect reports to eliminate known interactions
- Integrate existing databases into our in-memory database, e.g. pharmaceutical product details, U.S. FDA, or German Bfarm
- Enable real-time analysis / graphical exploration of these data
- Automatically combine and update newly released data
- Design and implement a web application



<http://www.cancer-clinical-trials.com/p/clinical-trials-in-cartoons.html>

Seminar Topics

F: Identification of Genetic Variants with Population-specific Data

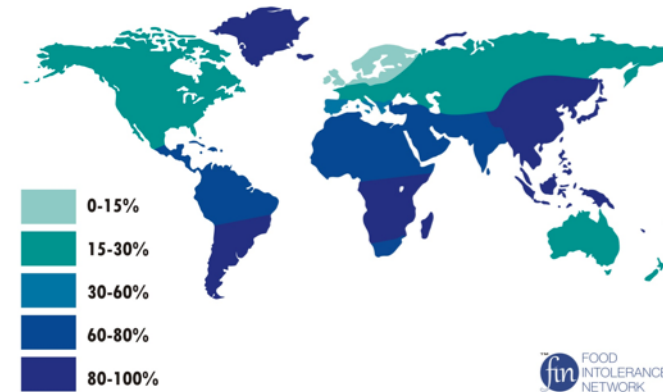
Issue:

To identify genetic variations, a DNA sample is analyzed and compared to a reference during Variant Calling. However, this rather simple method leads to inaccurate results because the data is error-prone and there are other factors that should be considered in calculations.

Idea:

- Identify and integrate data sources that contain information about genetic variants
- Extend the statistical model of variant calling to incorporate additional data sources, e.g. population-specific differences
- Evaluate the influence of different aspects on the quality of variant calling results

Worldwide prevalence of lactose intolerance in recent populations (schematic)



http://upload.wikimedia.org/wikipedia/commons/2/27/Worldwide_prevalence_of_lactose_intolerance_in_recent_populations.jpg

Seminar Organization

Enrollment for Seminar Topics

- How to apply for a topic?
 - Mail **prioritized** list of your 3 favorite topics to Cindy Fähnrich
 - 1 (very high adherence): ...
 - 2 (high adherence): ...
 - 3 (medium adherence): ...
 - Deadline: Tuesday April 15, 10:45 a.m.

- Assignment of seminar topics by Wednesday April 16



http://faithlifewomen.com/files/2012/06/wishlist_edit-630x320.jpg