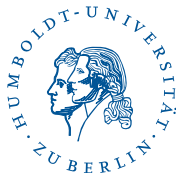


Semesterprojekt Verteiltes Genom Browsing

Projektstruktur, Werkzeuge und Use Cases



Institut für Informatik
Humboldt-Universität zu Berlin

27.10.2015

Inhaltsverzeichnis

- ▶ Mailinglists
- ▶ Git
- ▶ Teilprojekte
 - ▷ Integration
 - ▷ Middleware
 - ▷ GUI
- ▶ Projektleitung
- ▶ Use Cases

Mailinglists

- ▶ Drei Mailinglists:
 - ▷ `sp_vgb_all@lists.hu-berlin.de` [VGB-ALL]
 - ▷ `sp_vgb_integration@lists.hu-berlin.de` [VGB-IG]
 - ▷ `sp_vgb_middleware@lists.hu-berlin.de` [VGB-MW]
 - ▷ `sp_vgb_gui@lists.hu-berlin.de` [VGB-GUI]
- ▶ Externe können auch senden
- ▶ Nachrichten werden dann moderiert
- ▶ Mailtags verwenden!

Git

- ▶ Filehoster: GitHub
- ▶ Wird ebenfalls als Bugtracker und Milestonesverwalter verwendet
- ▶ Bis nächste Woche:
 - ▷ GitHub Account erstellen
 - ▷ Dem Tutorial Repository folgen
[github.com/hu-semesterprojekt-genombrowser/tutorial]
 - ▷ Repository clonen und eine Testdatei mit dem eigenen Namen pushen

Integration

Felix Scholze	f.scholze@gmx.de
Gabriel Krause	krausegv@informatik.hu-berlin.de
Kacper Siedlecki	kacperbajronsiedlecki@gmail.com
Marius Maass	le_schleicher@hotmail.de
Sonay Sengün	sonaysenguen@gmail.com

- ▶ Wöchentliches Treffen: Dienstags 11-13 Uhr, Bibiothek

Integration

Aufgaben:

- ▶ Herunterladen der Genomdaten und Metadaten von vier verschiedenen Quellen
- ▶ Parsing der Daten in einheitliches Format
- ▶ Erstellen einer Datenbank
- ▶ Überführen der Daten in die Datenbank

Middleware

Bastian Naber	b_naber@hotmail.de
Daniel Zyla	zyladani@informatik.hu-berlin.de
Martin Bauer	bauermax@informatik.hu-berlin.de
Martin Wackelbauer	paragumba@gmail.com
Nikita Rose	roseniki@cms.hu-berlin.de
Tobias Löffler	loeflert@informatik.hu-berlin.de

- ▶ Wöchentliches Treffen: Dienstags 11-13 Uhr, RUD 25 4.410

Middleware

Aufgaben:

- ▶ Schnittstelle zwischen UI und Daten
- ▶ Bereitstellung einer Datenstruktur, die eine effiziente Umsetzung der Suche ermöglicht
- ▶ Abarbeitung niedriger Detailstufen von UI-Anfragen aus der Datenstruktur im Hauptspeicher
- ▶ Umformung von UI-Anfragen in DB-Anfragen mit möglichst minimalem Suchaufwand
- ▶ „Zeit schinden“ durch geeignete Vorbereitung von Anfragen und Pufferung

Frontend

Aleksandar Atanasov	aleksandar_atanasov91@abv.bg
Erik Elisath	elisathe@informatik.hu-berlin.de
Jan Lelis	mail@janlelis.de
Johannes Rosswog	rosswogj@informatik.hu-berlin.de
Lucas Rebscher	lucas.rebscher@web.de
Robin Papke	robin_papke@web.de

- ▶ Wöchentliches Treffen: Dienstags 11-13 Uhr, RUD 25 3.321

Frontend

Aufgaben:

- ▶ Grafisches aufbereiten und darstellen der Daten
- ▶ Absprache mit Middleware

Projektleitung

- ▶ Malte Kruse (krusemal@informatik.hu-berlin.de)
- ▶ Ben Schumacher (ben.schumacher@informatik.hu-berlin.de)
- ▶ Treffen jeden Dienstag 8-9 Uhr (bzw. 11)

Ein Benutzer möchte sehen, welche Mutationen alle im Bereich von 144MB – 154MB auf dem Chromosom 7 auftreten können. Hierfür wählt er die 1000 Genomes Projekt-Datenbank aus. Nun sieht er das ganze Chromosom in einer eigenen „Lane“ im Vergleich zum Referenzgenom. Er sieht durch eine Markierung, dass im Bereich von 150MB-152MB gehäuft Mutationen auftreten können. Da er diesen Bereich nicht detailliert einsehen kann, da zu viele Basenpaare angezeigt werden, zoomt der Benutzer herein. Die Basenpaare, sowohl in der 1000 Genomes „Lane“ als auch vom Referenzgenom, sind immer besser zu erkennen, bis er genau einsehen kann, welche Mutationen auftreten können. Als er die maximale Zoomstufe erreicht, sind alle Basenpaare des Chromosoms 7 und des Referenzgenoms zu erkennen und er kann die auftretenden Mutationen betrachten.

Ein Benutzer möchte herausfinden, welche Mutationen bei einer bestimmten relativen Häufigkeit im Bereich eines Gens auftreten. Da er sich sehr für den Colorectalen Bereich interessiert, wählt er die TCGA aus. Da der Benutzer nicht den genauen Namen des Gens kennt und sich bei der Eingabe irrt, bekommt er kein Basenpaare angezeigt, sondern Vorschläge, welches Gen er gemeint haben könnte. Nun kann er sich auf Grund des Lesens erinnern und wählt einen der Vorschläge. Nun öffnet sich eine „Lane“ mit den Daten für den Bereich und der Benutzer sieht, wo die Mutation mit der bestimmten relativen Häufigkeit auftritt.

Ein Benutzer möchte sich darüber informieren, mit welcher Häufigkeit in einem bestimmten Bereich eines bestimmten Gens Mutationen auftreten können. Hierbei wählt er die HGMD mit einem Bereich von 140MB – 155MB, da der Bereich jedoch zu groß ist, wird ihm nur der Bereich von 140MB-150MB dargestellt.

Nun sieht er in dem Bereich, dass sehr wenig Mutationen auftreten. Ihn interessiert jedoch ebenfalls, ob bei Krebspatienten höhere Mutationsraten existieren. Dazu gibt er die gleiche Suche noch einmal ein, wählt jedoch zusätzlich TCGA für Lungenkrebs und TCGA für Colorectalkrebs.

Nun werden ihm die drei „Lanes“ und das Referenzgenom übereinander angezeigt. Nun kann er die Häufigkeit von Mutationen in den Bereichen vergleichen.

► Fragen?