

Examensarbete

Harmonisera distribuerad biologisk information "big data" för bättre genomisk analys

Introduktion

Teknikutvecklingen inom genomisk forskning går med rasande fart och man har i princip nått den magiska gränsen '1000 dollars-genomet', dvs. möjligheten att bestämma hela arvsmassan hos en person för ca 1000 dollar. Denna teknikrevolution, kallad "den nya generationens sekvenseringsteknik" (NGS), kommer att innebära ett paradigmskifte för hur biomedicinsk forskning och analys kommer att utföras inom akademi, industri och hälso- och sjukvård. Genom detta skifte har flaskhalsen flyttats från generering av data till analys och tolkning. Qlucore utvecklar en tilläggsmodul till programmet Qlucore Omics Explorer som skall kunna hjälpa användare att analysera och tolka NGS data.

Inom den biologiska forskningen tas det fram och har tagits fram en stor mängd med databaser som innehåller information som är viktig för olika grupper av forskare. Nedan kallar vi det publik data. Exempel är information om mutationer, vad som är sjukdomsgenererande osv

Projekt

För att erbjuda användare av Qlucore Omics Explorer ett bättre stöd i analysfasen skall projektet **undersöka, utvärdera och ta fram en prototyp för ett nytt system för hantering av biologisk publik data "big data", i en molnmiljö, med data från många olika oberoende publika källor.**

Projektet spänner över områden såsom "big data", "genomics", "health care" och förstås molnbaserad teknologi.

Målet med systemet är att åstadkomma ett stabilt gränssnitt mellan Qlucore Omics Explorer och den publika data som ändras och som per definition är utanför Qlucores kontroll.

Projektet är tänkt för två examensarbetare och består troligen av följande moment:

- Design av en modell för att hämta data från multipla heterogena publika "big data" databaser med olika struktur, uppdateringsstakt, rättigheter och gränssnitt.

- Skapa och definiera en molnbaserad middleware-databas.
- Specifikation och definition av ett objektorienterat gränssnitt (API) mellan middleware-databasen och Qlucore Omics Explorer som skall kunna hantera framtida nya dataset utan förändring.
- Test och utvärdering av att modellen klarar av aktuella uppdateringsintervall och datastorlekar

Från ett tekniskt perspektiv bedömer vi att arbetet minst inkluderar följande:

- Genomföra en förstudie på befintliga publika databaser (innehållande annoteringsinformation i form av kända mutationer, kända sjukdomar och annat) med fokus på bland annat datalagringsformat, uppdateringsintervall, säkerhet mm.
- Undersöka olika typer av lösningar för databaser i molnet. Vad finns det för initiativ inom bioinformatikvärlden redan? Finns det lösningar som går att använda eller anpassa?
- Specifikation, definition och utveckling av den modell och det ramverk som skall gälla för att hämta uppdatera "big data" informationen
- Kravställning och design av middleware molnlösningen.
- Verifiera valda lösningar i en prototyp.
- Ta fram en specifikation på APIet och verifiera lösningen i en prototyp.

Förkunskaper

Gärna erfarenhet från databaser, molnteknologi och C++

Övrigt

Ersättning utgår.

En stor del av arbetet utförs lämpligast i Qlucore's lokaler på Ideon.