

SELEZIONE PUBBLICA PER TITOLI E COLLOQUIO PER IL CONFERIMENTO DI N. 1 ASSEGNO DELLA DURATA DI 12 MESI PER LO SVOLGIMENTO DI ATTIVITA' DI RICERCA AI SENSI DELL'ART. 22 DELLA LEGGE 30.12.2010, N. 240 PRESSO IL DIPARTIMENTO DI INGEGNERIA GESTIONALE, DELL'INFORMAZIONE E DELLA PRODUZIONE DELL'UNIVERSITA' DEGLI STUDI DI BERGAMO (SETTORE CONCURSALE 09/G1 – AUTOMATICA) SSD ING-INF/04 – AUTOMATICA (FONDO: DIGIPFUTURMAN)

bandito con Decreto del Rettore Rep. n. 722/2017 del 11.12.2017, pubblicato all'Albo di Ateneo il 11.12.2017

PROGETTO DI RICERCA

Il processo di analisi del genoma è composto di diverse fasi. Alcune richiedono la preparazione, manipolazione ed analisi dei campioni fisici ed hanno durate non comprimibili. Altre riguardano solo il trattamento dell'informazione con numerose diverse attività (accessi a DB esterni, accessi memorie di massa locali, accessi al cloud, calcolo, etc...) su file "pesanti" (il file annotato completo, alla fine del processo, occupa all'incirca 20 GB). Inoltre, nel processo di diagnostica genetica è necessario accedere a numerosi DB che contengono le informazioni sulle patologie associate alle mutazioni genetiche. Le informazioni recuperate dalle query a questi DB vengono poi associate alle mutazioni del genoma in esame, generando il cosiddetto "File annotato". Infine, una volta che è stata effettuata la diagnosi, il caso clinico viene archiviato associando la sintomatologia alla base del quesito diagnostico alla sola mutazione per cui è stata effettuata la diagnosi (o la ricerca per rispondere al quesito diagnostico). Può accadere che successivamente arrivino da altri centri medici richieste di ricerca nella casistica clinica su sintomatologie simili a casi già visti ma con quesiti o ipotesi diagnostiche differenti. Attualmente, ricercare in modo automatico e veloce tra i casi clinici già archiviati sulla base di ipotesi diagnostiche differenti è un'operazione complessa e non standardizzata.

Gli obiettivi del lavoro di ricerca sono:

- a) Effettuare un'analisi funzionale delle diverse attività del processo con l'obiettivo di individuare quelle che sono maggiormente time-consuming e chiarirne i motivi tecnici.
- b) Proporre una nuova architettura di sistema informativo "locale" (per il solo servizio di Genetica Medica) evidenziandone chiaramente i benefici dal punto di vista delle prestazioni e le modalità di interazione con il sistema ospedaliero, con particolare attenzione a garantire la sicurezza e l'integrità del dato.
- c) rendere completamente automatica la generazione del file annotato, per esempio attraverso lo sviluppo di uno script Ruby on Rails/Python che prova ad automatizzare il processo di query e di download dei risultati dal sito web del DB, sia attraverso la modifica del programma Bash esistente che interpreta i risultati ritornati dall'interrogazione e genera il file annotato completato.
- d) la progettazione di un nuovo sistema di archiviazione del dato genetico che consenta di effettuare query per creare nuove associazioni tra i dati contenuti nei file annotati e la sintomatologia. Il candidato dovrà prima definire la tecnologia database più adatta a gestire il dato genetico e, successivamente, svilupparne l'implementazione. Esempio di tecnologie RDBMS: Oracle DB, MySQL, IBM DB2, PostgreSQL. Esempio di tecnologie non relazionali: Cassandra, HBase.