

## Prova pratica 22/03/2019

Vanno consegnati sia il codice software che l'output generati durante la prova pratica. Il codice, in particolare, deve essere riproducibile e fornire lo stesso identico output.

- Accesso al computer con utente "user" e password "lab".
- Usare il browser web per le pagine di riferimento.
- Il computer dispone di un server Sun Grid Engine (SGE) locale in cui è disponibile una coda denominata "local" per sottomettere job di calcolo.
- Il computer dispone di un database MongoDB accessibile tramite console (comando "mongo").

### 1. Accesso programmatico a servizi web con API REST

UniProtKB fornisce accesso programmatico ai propri dati tramite una API REST (<https://www.ebi.ac.uk/prot eins/api/doc/>). Utilizzando uno script implementato p.es. in Python o Perl, fare una ricerca per identificare le proteine "Cytochrome C" in UniProtKB che abbiano associato almeno un dominio Pfam e una struttura PDB. Per ogni elemento restituito va stampato l'identificativo UniProtKB e la lista di domini Pfam e strutture PDB associati.

*Alto Tardito*

*user*

*passwd*  
*lab*

## Prova pratica 22/03/2019

Vanno consegnati sia il codice software che l'output generati durante la prova pratica. Il codice, in particolare, deve essere riproducibile e fornire lo stesso identico output.

- Accesso al computer con utente "user" e password "lab".
- Usare il browser web per le pagine di riferimento.
- Il computer dispone di un server Sun Grid Engine (SGE) locale in cui è disponibile una coda denominata "local" per sottomettere job di calcolo.
- Il computer dispone di un database MongoDB accessibile tramite console (comando "mongo").

### 2. Calcolo distribuito con Sun Grid Engine

Scaricare le prime 25 strutture della proteina "Cytochrome C" dal sito del PDB. Se necessario estrarre la catena corrispondente e generare un nuovo file PDB per ogni record scaricato.

Eseguire un allineamento strutturale tutti contro tutti con il programma TM-align tramite uno script SGE utilizzando la coda "local". Generare un file di output contenente per ogni riga l'identificativo della coppia dei file PDB confrontati, il TM-score e il valore di RMSD riportati da TM-align. Le colonne vanno separate dal carattere "tab".

*Antonio Turchio*

*prof*

*RMSD*

*prof*

## Prova pratica 22/03/2019

Vanno consegnati sia il codice software che l'output generati durante la prova pratica. Il codice, in particolare, deve essere riproducibile e fornire lo stesso identico output.

- Accesso al computer con utente "user" e password "lab".
- Usare il browser web per le pagine di riferimento.
- Il computer dispone di un server Sun Grid Engine (SGE) locale in cui è disponibile una coda denominata "local" per sottomettere job di calcolo.
- Il computer dispone di un database MongoDB accessibile tramite console (comando "mongo").

### 3. Banche dati MongoDB

Effettuare una ricerca in UniProtKB delle proteine "alpha synuclein" di "Homo sapiens". Dalla pagina della ricerca cliccare su "Columns" e selezionare: "Entry name", "Protein names", "Gene names" e "Length". Scaricare il risultato in formato "tab-separated". Convertire il file scaricato in formato JSON ed inserirlo nella banca dati MongoDB su localhost (porta 27017, nome DB "prova", collection "uniprot"). Generare uno script p.es. JavaScript che stampi in standard output l'intero documento della proteina più lunga.

Roberto Bortolotto  
prof

Roberto Bortolotto  
prof