

**Laboratorio di Basi di Dati per Bioinformatica**  
**Esercitazione 11**

1. Eseguire sulla base di dati [genome](#) sul server dobserver almeno 3 interrogazioni che prevedano l'uso dei tipi `int_interval` e `stree` (con eventuali cast) e dei relativi operatori. Evitare di usare `select *` sulla base di dati (in particolare sulla tabella `sequence` dal momento che le sequenze sono di lunghezza non banale).
2. Analizzare attentamente i riferimenti ai siti indicati nei lucidi, con particolare attenzione ai dati in formato XML e ai relativi DTD: costruire una piccola base di dati con i riferimenti alla principali basi di dati di interesse bioinformatico, la presenza di dati XML, di DTD, gli ambiti scientifici e le dimensioni delle varie basi di dati, i possibili sistemi di interrogazione.
3. Scaricare dalla pagina del corso il file `NCI_CancerIndex_allphases_diseaseExcerpt.zip`. Fare riferimento alla pagina <https://wiki.nci.nih.gov/display/ICR/The+Cancer+Gene+Index+Gene-Disease+and+Gene-Compound+XML+Documents> per dettagli sul contenuto dei documenti XML relativi ai geni legati all'insorgenza di tumori. Analizzare il DTD fornito, provare a validare localmente (con `xmllint` – comando da terminale) il relativo documento XML. Effettuare almeno 3 query in Xpath sul documento XML in questione, per l'identificazione di sottoparti (piccole!) del documento considerato (usare PostgreSQL).