

4. Übung Skriptsprachen in der Bioinformatik

Sommersemester 2014

Sascha Winter, Kai Dührkopp

Ausgabe: 14.08.2014

- Aufgabe 1.** Extrahieren Sie aus *pubchem.xml* alle Compounds mit Molekülformel, exakter Masse und Pubchem ID. Die Pubchem ID befindet sich im Tag "PC-CompoundType_id.cid". Masse und Molekülformel befinden sich in einem Tag InfoData_value_fval bzw. InfoData_value_sval. Ob es sich bei dem darin enthaltenen Wert um die Masse bzw. Molekülformel handelt, hängt davon ab ob im darüber liegenden PC-Urn_label der Wert "Mass" bzw. "Molecular Formula" steht. Schreiben Sie die drei Angaben für alle Moleküle in eine CSV Datei.
- Aufgabe 2.** Nutzen Sie den Pubchem REST Service um den inchikey und die MolecularFormula für die Compounds mit den IDs 2244,1000,1800, 30602 zu erhalten. Speichern sie die Ausgabe als CSV ab. <https://pubchem.ncbi.nlm.nih.gov/> Hinweis: Die Dokumentation von Pubchem REST enthält bereits Beispiele wie man diese Aufgabe löst.
- Aufgabe 3.** Nutzen Sie den ChemSpider SOAP Service um die Ids aller Compounds mit Masse 120.012 mit erlaubter Massenabweichung von 0.0025 zu erhalten. Speichern sie die IDs als CSV ab. <http://www.chemspider.com/> Hinweis: Nutzen Sie die Funktion SearchByMass.
- Aufgabe 4.** Nutzen Sie den CheBi SOAP Service und suchen sie nach Einträgen mit dem Namen "Flavonoid" die einen SearchScore von über 0.75 haben. Geben sie die Beschreibung dieser Ontologieeinträge zurück. <http://www.ebi.ac.uk/chebi/webServices.do> Hinweis: Mit **getLiteEntity** sucht man nach beliebigen Suchbegriffen. Die daraus erhaltenen IDs kann man in **getCompleteEntity** geben um andere Eigenschaften wie die Definition zu erhalten.