BIOINFORMATIKA – CV. 3

SOAP rozhraní k NCBI a zadání první samostatné práce

NCBI (znáte z předchozího cvičení)

- http://www.ncbi.nlm.nih.gov/
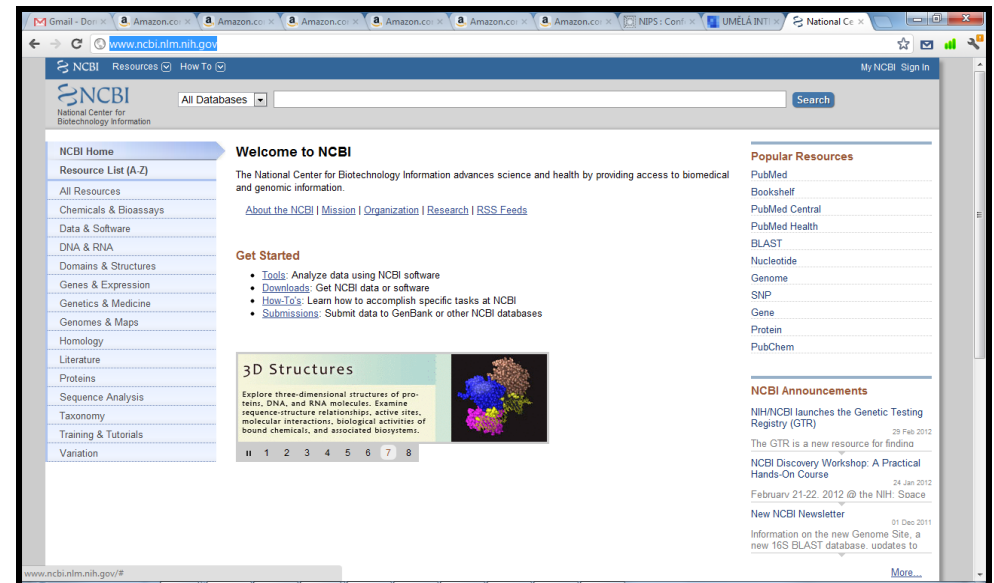
- Různé databáze:
  - Genome
  - Gene
  - PubMed
  - Protein
  - HomoloGene
  - …

- Většina z těchto databází umožňuje přístup pomocí SOAP

NCBI Gene

□ „Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.“

□ **Příklad:**

　□ http://www.ncbi.nlm.nih.gov/gene/5053

□ **Informace:** *Summary, Genomic context, Genomic regions, transcripts, and products, Bibliography, Phenotypes, Interactions, General gene info, General protein info, Reference sequences, Related sequences, Additional links*

SOAP = Simple Object Access Protocol

- **SOAP** je protokol pro výměnu zpráv založených na XML

- Formát SOAP tvoří základní vrstvu komunikace mezi webovými službami

```xml
<?xml version="1.0"?>
<soap:Envelope xmlns:soap="http://www.w3.org/2003/05/soap-envelope">
<soap:Header> </soap:Header>
        <soap:Body>
                <m:GetStockPrice xmlns:m="http://www.example.org/stock">
                        <m:StockName>IBM</m:StockName>
                </m:GetStockPrice>
        </soap:Body>
</soap:Envelope>
```

Přístup k NCBI pomocí SOAP

- Automaticky vygenerované Java třídy – Apache Axis2

  - Můžete si vygenerovat sami (pomocí nástrojů z Axis2) nebo stáhnout z

    ftp://ftp.ncbi.nlm.nih.gov/pub/eutils/soap/v2.0/java/axis2_1.5.2_jdk_6.0.12/)

  - Budete potřebovat knihovny (jar) z

    http://axis.apache.org/axis2/java/core/download.cgi - **Pozn.:** pokud budete používat stažené binaries z NCBI, pak použijte Apache Axis2 verze 1.5.2

Příklad: načtení informací z NCBI

```java
/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */
package bin.tutorial3;

import gov.nih.nlm.ncbi.www.soap.eutils.EUtilsServiceStub;

/**
 *
 * @author Ondra
 */
public class Example1 {

    public static void main(String[] args) throws Exception
    {
        // run_eGquery provides Entrez database counts for a single search
        try
        {
            EUtilsServiceStub service = new EUtilsServiceStub();
            // call NCBI eGQuery utility
            EUtilsServiceStub.EGqueryRequest req = new EUtilsServiceStub.EGqueryRequest();
            req.setTerm("PAH");
            EUtilsServiceStub.Result res = service.run_eGquery(req);
            // results output
            System.out.println("Search term: " + res.getTerm());
            System.out.println("Results: ");
            for (int i = 0; i < res.getEGQueryResult().getResultItem().length; i++)
            {
                System.out.println("  " + res.getEGQueryResult().getResultItem()[i].getDbName() +
                                    ": " + res.getEGQueryResult().getResultItem()[i].getCount());

                System.out.println("First search result in "+res.getEGQueryResult().getResultItem()[i].getDbName());
            }
        }
        catch (Exception e) { System.out.println(e.toString()); }
    }

}
```

Příklad 2: Načtení informací z NCBI Gene

```java
package bin.tutorial3;

import gov.nih.nlm.ncbi.www.soap.eutils.EFetchGeneServiceStub;
import gov.nih.nlm.ncbi.www.soap.eutils.EFetchGeneServiceStub.EntrezgeneSetSequenceE;
import gov.nih.nlm.ncbi.www.soap.eutils.EFetchGeneServiceStub.EntrezgeneSet_type0;
import gov.nih.nlm.ncbi.www.soap.eutils.EFetchGeneServiceStub.Entrezgene_type0;

/**
 *
 * @author Ondra
 */
public class Example2 {

    public static void main(String[] args) throws Exception
    {
        try
        {
            EFetchGeneServiceStub service = new EFetchGeneServiceStub();
            // call NCBI EFetch utility
            EFetchGeneServiceStub.EFetchRequest req = new EFetchGeneServiceStub.EFetchRequest();

            //ID of PAH
            req.setId("5053");
            EFetchGeneServiceStub.EFetchResult res = service.run_eFetch(req);

            // results output

            EntrezgeneSet_type0 egs = res.getEntrezgeneSet();

            EntrezgeneSetSequenceE[] seqs = egs.getEntrezgeneSetSequence();
            for (EntrezgeneSetSequenceE seq : seqs){
                Entrezgene_type0 gene = seq.getEntrezgene();
                //gene summary
                System.out.println("Gene summary: "+gene.getEntrezgene_summary());
            }
        }
        catch (Exception e) {
            e.printStackTrace();
        }
    }

}
```