

TÖL303G - Gagnasafnsfræði

Vikublað 11

1 Heimadæmi

Skilafrestur er til mánudagsins 21. nóvember kl. 23:59. Gradiance skilafrestur er til 23:59, þriðjudaginn 22. nóvember

Gradiance æfingar

Klárið dæmi merkt HW 13 á Gradiance. Athugið að Namespaces, xmlns er í kafla 2.6 og DTD er í kafla 3 undir XML kaflanum (11).

Dæmi

1. Dæmi 2.1 bls. 484 í bók. Athugið að skoða Fig. 5 bls. 481.
2. Finnið 'PMID' og 'ArticleTitle' fyrir allar greinar í gagnagrunninum medsamp2014.xml sem hafa fleiri en einn höfund og orðið 'gene' kemur fyrir í annað hvort ArticleTitle eða Abstract. Lýsing á grunninum er gefin í skjalinu.nlmmedlinecitationset_130101.dtd. Hér er dæmi um hvernig má vinna með XML í Python með ElementTree pakkanum. Sjá frekari leiðbeiningar á <https://docs.python.org/2/library/xml.etree.elementtree.html>

```
import xml.etree.ElementTree as ET
tree = ET.parse('medsamp2014.xml')
root = tree.getroot()

# print all children
for x in root:
    print x.tag, x.text

# search for tag by name
y = root[0] # get first child
z = y.find('./PMID')
print z.tag, z.text

for y in root[:100]:
    for z in y.findall('./KeywordList/'):
        print y.find('PMID').text, z.text
```