HMDB database

library(hmdbQuery)

lk1 = HmdbEntry(prefix = "https://www.hmdb.ca/metabolites/",

id = "HMDB0000001")

lk2 = HmdbEntry(prefix = "https://www.hmdb.ca/metabolites/",

name = "1-Methylhistidine")

data(hmdb1)

hmdb1

store(hmdb1)

store(hmdb1)[[28]]

biospecimens(lk1)

tissues(lk1)

st = store(lk1)

head(names(st))

length(names(st))

st$protein\_assoc["name",]

st$protein\_assoc["gene\_name",]

st$protein\_associations

unlist(s$pathways)

lk2 = HmdbEntry(prefix = "https://www.hmdb.ca/metabolites/",

id = "HMDB0304954")

st = store(lk2)

cat split.aa | awk -F'[<>]' '/accession/{print $3}'

cat split.aa | awk -F'[<>]' '/metabolite/accession/{print $3}'

cd /data/home/nis2034/Nish/hmdb/

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date/ {getline;print $3;f=0}' hmdb\_metabolites.xml >> accession.txt

wc -l accession.txt -> 56411 accession.txt

awk -F"[<>]" '/\<\/secondary\_accessions/ { getline;print $3}' hmdb\_metabolites.xml >> meta\_name.txt

wc -l meta\_name.txt. -> 56411 meta\_name.txt

awk -F"[<>]" '/direct\_parent/ {print $3}' hmdb\_metabolites.xml >> direct\_parent.txt

wc -l direct\_parent.txt -> 56328 direct\_parent.txt

awk -F"[<>]" '/kingdom/ {print $3}' hmdb\_metabolites.xml >> kingdom.txt

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date|\<\/secondary\_accessions/ {getline; printf "%-10s", $3OFS }' split.aa

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date|\<\/secondary\_accessions/ {getline; print $3;f=0}' split.aa

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date|\<\/secondary\_accessions/ {getline; printf ($3);}' split.aa

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date|\<\/secondary\_accessions/ {getline; print $3;}' split.aa

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date/ {getline;print $3;f=0}' split.aa

awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date/ {getline;print $3;f=0}' hmdb\_metabolites.xml >> accession.txt

accession=$(xmllint --xpath 'string(/metabolite /accession)' split.aa

printf '%s\n' "accession: $accession

acc = $(awk -F"[<>]" '/metabolite/ {f=1} f && /update\_date/ {getline;print $3;f=0}' split.aa)

echo "The acc is $acc"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*XMLStarlet\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

scp hmdb\_metabolites.xml nis2034@bastion:/data/home/nis2034/Nish/hmdb/

split -l 5000 hmdb\_metabolites.xml split.

mv split.\* \temp

xmlstarlet select -t -v '//File[0]/CustomerEmail' file.xml

xmlstarlet select -t -v '//File/@Name' file.xml

xmlstarlet select -t -v '//File/CustomerEmail' file.xml

xmlstarlet select -t -v '//root/ContactPerson/GivenName' file.xml

xmlstarlet select -t -v '//hmdb/metabolite/accession' split.aa

xmlstarlet sel -t -v "count(//GivenName)" file.xml

hmdb/metabolite/accession

tail -n +3 > file.stdout

awk '/metabolite/ && c==0 {c+=1; next} /metabolite/ && c>0 {c+=1; print " metabolite ="c; c=0; next} c>0 {c+=1}' split.aa

split -l 1131 split.aa sp\_test.

head -1131 split.aa

# awk -F'[<|>]' ' # Set the field separator to < or >

/tns/ { # If line contain word "tsn"

t=$3 # Assign the value of third field to variable **n**

}

/user/ { # If line contain word "user"

u=$3 # Assign the value of third field to variable **u**

}

/password/{ # If line contain word "password"

printf \ # I use printf to print the line

"sqlplus %s/%s@%s\n",\ # Format the print output

t,$3,u # Print variable **t**, value of **$3** and variable **u**

}

' file # Input file

# awk -F'[<|>]' '/accession/ { t=$3 }/name/{ u=$3 } /kingdom/{ printf "hmdb %s/%s@%s\n",t,$3,u }' split.aa

awk -F'[<|>]' '/update\_date/ {getline; t=$3 ; printf "%s\n", t}' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/class/{c=$3}/sub\_class/{l=$3}/direct\_parent/{d=$3;printf "%s\t%s\t%s\t%s\t%s\t%s\t%s\n",a,n,k,s,c,l,d}' split.aa

awk -F'[<|>]' '/<\/direct\_parent/{getline;k=$3; printf "%s\n",k}' split.aa

awk -F'[<|>]' '/sub\_class/{ k=$3; printf "%s\n",k}' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/class/{c=$3}/sub\_class/{l=$3}/direct\_parent/{d=$3;printf "%s\t%s\t%s\t%s\t%s\t%s\t%s\n",a,n,k,s,c,l,d}' split.aa

Accession Name direct\_parent Kingdom Super\_class class

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,d,k,s,l,w }' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,d,k,s,l,w ; a="";n="";d="";k="";s="";l="";w="" }' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,d,k,s,l,w ; a="";n="";d="";k="";s="";l="";w="" }'

wc -l meta\_anno\_hmdb.txt

awk -F'[<|>]' '/ContactPerson/{f=1}/GivenName/{a=$3}/FamilyName/{b=$3} f && /GivenName/{printf "%s\t%s \n",a,b ;f=0;a="";b=""}' file.xml

<?xml version="1.0"?>

<root>

<ContactPerson>

<GivenName>John</GivenName>

<FamilyName>Wayne</FamilyName>

<PhoneFix>78787878</PhoneFix>

<EmailList>

<Email>john.wayne@gmail.com</Email>

</EmailList>

</ContactPerson>

<ContactPerson>

<GivenName>James</GivenName>

<PhoneFix>90909090</PhoneFix>

<EmailList>

<Email>james.stewart@gmail.com</Email>

</EmailList>

</ContactPerson>

<ContactPerson>

<GivenName>Jafer</GivenName>

<FamilyName>Khan</FamilyName>

<PhoneFixList>

<PhoneFix>90909090</PhoneFix>

</PhoneFixList>

<EmailList>

<Email>james.stewart@gmail.com</Email>

</EmailList>

</ContactPerson>

</root>

John

James Wayne

Jafer Wayne

awk -F'[<|>]' '/GivenName/{a=$3}/FamilyName/{b=$3}/PhoneFix/{c=$3}/Email/ {d=$3; printf "%s\t%s\t%s\t%s \n",a,b,c,d }' file.xml

awk -F'[<|>]' '/<\/PhoneFixList/{getline;c=$3;printf "%s\n",c}' file.xml

awk -F'[<|>]' '/<\/GivenName/{getline;c=$3;printf "%s\n",c}' file.xml

awk -F'[<|>]' '/Email/{d=$3; printf "%s\n",d }' file.xml

awk -F'[<|>]' '/GivenName/{d=$3; printf "%s\n",d }' file.xml

awk -F'[<|>]' '/GivenName/{a=$3}/FamilyName/{b=$3}/PhoneFix/{c=$3}/Email/ {d=$3; printf "%s\t%s\t%s\t%s \n",a,b,c,d }' file.xml

awk -F'[<|>]' '/GivenName/{a=$3}/FamilyName/{b=$3;printf "%s\t%s \n",a,b }' file.xml

awk -F'[<|>]' '/GivenName/{a=$3}/FamilyName/{b=$3}/PhoneFix/{c=$3;printf "%s\t%s\t%s \n",a,b,c; a="";b="";c="" }' file.xml

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,d,k,s,l,w ; a="";n="";d="";k="";s="";l="";w="" }' hmdb\_metabolites.xml >> meta\_anno\_hmdb.txt

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/smpdb\_id/{p=$3}/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,p,d,k,s,l,w ; a="";n="";p="";d="";k="";s="";l="";w="" }' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/smpdb\_id/{p=$3}/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,p,d,k,s,l,w }' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/<pathways>/{p=$3; printf "%s\t%s\t%s\n",a,n,p; a="";n="";p=""}' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<smpdb\_id>/{p=$3; printf "%s\t%s\n",a,p}' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<smpdb\_id>/{p=$3} /<pathway>/{getline;q=$3; printf "%s\t%s\t%s\n",a,p,q}' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<pathway>/{getline;q=$3}/<smpdb\_id>/{p=$3; printf "%s\t%s\t%s\n",a,p,q; }' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}

/direct\_parent/{d=$3}/<\/direct\_parent/{getline;k=$3}/super\_class/{s=$3}/<\/super\_class/{getline; l=$3}/sub\_class/{w=$3}/<pathway>/{getline;q=$3}/<smpdb\_id>/{p=$3; printf "%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,d,k,s,l,w,q,p }' hmdb\_metabolites.xml >> meta\_anno\_hmdb.txt

awk -F'[<|>]' '/<pathway>/{p=$3; printf "%s\n",p}' split.aa

Another better logic to print- NOt implremented

}/sub\_class/{w=$3} if any of a,d,l or k is not null { printf "%s\t%s\t%s\t%s\t%s\t%s\t%s \n",a,n,d,k,s,l,w ; a="";n="";d="";k="";s="";l="";w="" }' hmdb\_metabolites.xml >> meta\_anno\_hmdb.txt

echo -e "FNAME\tLNAME" > file2.txt && cat file.txt >> file2.txt

Accession Name direct\_parent Kingdom Super\_class class sub\_class

echo -e " HMDB\_id\tname\tdirect\_parent\tkingdom\tsuper\_class\tclass\tsub\_class\tpathway\_name\tSMP" > meta\_anno\_hmdb\_h.txt && cat meta\_anno\_hmdb.txt >> meta\_anno\_hmdb\_h.txt

scp nis2034@bastion:/data/home/nis2034/Nish/hmdb/meta\_anno\_hmdb\_h.txt /Users/nis2034/Desktop/nis2034/Desktop/Nish/RProjects/autonomics/nis2034/beef\ metabolomics/

grep -rnw '/temp' -e 'HMDB0000538'

HMDB\_id/name/synonym

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3}/<synonyms>/ {f=1} f&&/<synonym>/{s=$3; printf "%s\t%s\t%s \n",a,n,s }' split.aa

awk -F'[<|>]' '/update\_date/{getline; a=$3}/<\/secondary\_accessions/ {getline;n=$3} /<\/description>/ {f=1} f&&/<synonym>/{s=$3; printf "%s\t%s\t%s \n",a,n,s }' hmdb\_metabolites.xml >> meta\_name\_hmdb.txt

echo -e " HMDB\_id\tname\tsynonym " > meta\_name\_hmdb\_h.txt && cat meta\_name\_hmdb.txt >> meta\_name\_hmdb\_h.txt

scp nis2034@bastion:/data/home/nis2034/Nish/hmdb/meta\_name\_hmdb\_h.txt /Users/nis2034/Desktop/nis2034/Desktop/Nish/RProjects/autonomics/nis2034/beef\ metabolomics/

xmlstarlet select -t -v '//hmdb/metabolite/accession' split.aa

xmlstarlet sel -t -m "//hmdb/metabolite/accession" -t -m "/hmdb/metabolite/name" -t -m "//hmdb/metabolite/synonyms/synonym" -n split.aa

xmlstarlet sel -T -t -m "//hmdb/metabolite/accession" -v "concat(current(),',',//hmdb/metabolite/name)" -n split.aa

xmlstarlet sel -T -t -m "//hmdb/metabolite/accession" -v "concat(current(),',',//hmdb/metabolite/name,',',//hmdb/metabolite/synonyms/synonym)" -n split.aa