

# Entrez Utilities (E-Utills)

- ✓ NCBI — это организация, у которой есть много баз данных;
- ✓ Entrez — это система 40 баз данных и поиска по ним,
- ✓ E-Utilities (Entrez Utilities) — это набор программ для работы с Entrez.

## Доступ к E-Utilities:

1. <https://eutils.ncbi.nlm.nih.gov/entrez/eutils/>
2. [Entrez Direct](#): UNIX command line;
3. [eutils](#) в Python или [Модуль Entrez в Biopython](#);
4. [reutils](#) в R (CRAN).

# Инструменты E-Utils

- ✓ ESearch — поиск по запросу;
- ✓ ESummary — подробные сведения по UID;
- ✓ EFetch — скачивание по UID;
- ✓ ELink — ссылки из одной базы в другую;
- ✓ EPost — загрузка данных на сервер NCBI для последующего использования;
- ✓ ...

ESummary есть для всех баз, EFetch — не для всех.

# ESearch

Обязательные параметры: db, term.

Необязательные параметры: usehistory, WebEnv, retstart, rettype...

`https://eutils.ncbi.nlm.nih.gov/entrez/eutils/ +`

`esearch.fcgi?db=nucleotide&term=gspt1`

`esearch.fcgi?db=nucleotide&term=gspt1&+`

`"Homo%20sapiens"[porgn]`

`esearch.fcgi?db=nucleotide&term=gspt1&usehistory=y`

# ESummary

Обязательные параметры: db, id\* или query\_key + WebEnv.

\* Это UID (unique record integer) для конкретной базы.

```
https://eutils.ncbi.nlm.nih.gov/entrez/eutils/  
esummary.fcgi?db=protein&id=6678417,  
9507199,28558982,28558984,28558988,28558990
```

# EFetch

Обязательные параметры: db, id\* или query\_key + WebEnv.

\* Это UID (unique record integer) для конкретной базы.  
Дополнительные параметры: rettype, retmode (список допустимых значений).

```
https://eutils.ncbi.nlm.nih.gov/entrez/eutils/  
efetch.fcgi?db=protein&id=6678417,  
9507199,28558982,28558984,28558988,28558990&  
rettype=fasta&retmode=text
```

# Примеры для EDirect

```
esearch -email email@example.com -tool mytool \  
-db nucleotide -query "gspt1 AND human[orgn]" \  
>human_gspt.xml
```

```
esearch -email email@example.com -tool mytool \  
-db protein -query "gspt1 AND \  
human[orgn]" | efetch -format fasta -mode text \  
>human_gspt.fa
```

```
esearch -email email@example.com -tool mytool \  
-db nucleotide -query "gspt1 AND human[orgn]" | \  
esummary -mode text -format docsum | xtract -pattern \  
DocumentSummary -element Id Caption Slen >hum_gspt1.txt
```

```
elink -id 12890024 -db pubmed -target nuccore | \  
efetch -format fasta -mode text
```

Представляйтесь системе (email).

# Примеры для Biopython::Entrez

```
from Bio import Entrez
Entrez.email = "A.N.Other@example.com"
handle = Entrez.esearch(db = "nucleotide", term = \
                        "gspt AND human[orgn]")
record = Entrez.read(handle)
mylst = []
for rec in record["IdList"]:
    temphandle = Entrez.esummary(db = "nucleotide", \
                                id=rec, retmode="text")
    mylst.append(temphandle)

fhandle = Entrez.efetch(db = "nucleotide", \
                        id=record["IdList"], rettype="fasta", retmode="text")

lhandle = Entrez.elink(dbfrom="pubmed", \
                        dbto="nucleotide", id=12890024)
lrecord = Entrez.read(lhandle)

for link in lrecord[0]["LinkSetDb"][0]["Link"]:
    print(link["Id"])
```

## Примеры для reutils

```
library(reutils)
options(reutils.email = "email@example.com")
s <- esearch(db = "nucleotide", term = "gspt1 AND \
                                         human[orgn]")
su <- esummary(s); cosu <- content(su, "parsed")
as.data.frame(cosu[,c("Id", "Caption", "Slen")])

f <- efetch(uid = s[1:10], db = "nucleotide", \
            rettype = "fasta", retmode = "text")
write(content(f), "gspt1_human_1_9.fa")
l <- elink(s[1:9], dbFrom = "nucleotide", \
           dbTo = "protein")
l1 <- linkset(l)
f2 <- efetch(l1, db = "protein", \
            rettype = "fasta", retmode = "text")
write(content(f2), "gspt1_human_1_9.faa")
l11 <- elink("12890024", dbFrom = "pubmed", \
            dbTo = "nucleotide")
l4 <- linkset(l11)
efetch(l4[,], db="nuccore", rettype = "fasta", \
       retmode = "text")
```



# Полезные ссылки

Entrez Programming Utilities Help  
Entrez Direct Book

## Вебинары NCBI:

- ✓ Короткое введение (3');
- ✓ Introduction to NCBI's E-utilities API (50');
- ✓ EDirect (~90');
- ✓ Все материалы со всех вебинаров.

# NCBI toolkit & fastq-dump

- ✓ Не только fastq!
- ✓ "NCBI's fastq-dump has to be one of the worst-documented programs available online."

[https:](https://edwards.flinders.edu.au/fastq-dump/)

[//edwards.flinders.edu.au/fastq-dump/](https://edwards.flinders.edu.au/fastq-dump/) <=

Рекомендую!

- ✓ <https://rnnh.github.io/bioinfo-notebook/docs/fasterq-dump.html>