

Comandes:

Busca si una PARAULA existeix:

```
curl -X POST --data "DecEntradaText=$PARAULA" "https://dlc.iec.cat/Results"
```

També es pot buscar amb el post a la url però falla amb caràcters especials com à,è,ç,í,ü....

```
curl -s "https://dlc.iec.cat/Results?DecEntradaText=$PARAULA"|grep "1 registres"
```

Buscar per ID:

```
https://dlc.iec.cat/Results/PrintAccepcio?id=$ID
```

Descarreguem totes les paraules de <https://dlc.iec.cat>

Primer fem consulta de totes les paraules que comencen per unes lletres.

Aquí falten paraules, sobretot els femenins, com alesiada

Hem tronar alesià que té l'id 0043250

Si busquem aquest id amb la comanda:

```
https://dlc.iec.cat/Results/PrintAccepcio?id=0043250
```



Ens dona:

```
<div id="Definition" class="resultDefinition"><h2
xmlns:fo="http://www.w3.org/1999/XSL/Format"><span class="title">alesiat
</span><span class="title">-ada </span><br></h2><br>
xmlns:fo="http://www.w3.org/1999/XSL/Format"><span class="tagline"
xmlns:fo="http://www.w3.org/1999/XSL/Format">adj.</span><span class="body"
xmlns:fo="http://www.w3.org/1999/XSL/Format"> <span class="tip"
onmouseover="doTooltip(event, ' [LC] ' )" onmouseout="hideTip()"> [LC]
</span><span class="tip" onmouseover="doTooltip(event, ' [MD] ' )"
onmouseout="hideTip()"> [MD] </span> </span><span class="body"
xmlns:fo="http://www.w3.org/1999/XSL/Format">Espatllat, inutilitzat, en
alguna part del cos. </span><span class="body"
xmlns:fo="http://www.w3.org/1999/XSL/Format"><span class="italic">Estar
nafrat o alesià d'algun dels seus membres.</span></span></div>
```

Ens fixem en aquesta part:

```
<span class="title">alesiat </span><span class="title">-ada </span>
```

Al web surt així

alesiat

alesiat -ada

adj. [LC] [MD] Espatllat, inutilitzat, en alguna part del cos. Estar nafrat o alesiat d'algun dels seus membres.

Amb aquesta consulta treiem tots els resultats:

```
grep 'id="Definition"' 0* |awk -F '<|>' '{for(i=1;i<=NF-2;i++) if ( $i ~ /title/ ) printf $i" - "$(i+1)"\n"}'
grep 'id="Definition"' 0* |awk -F '<|>' '{for(i=1;i<=NF-2;i++) if ( $i ~ /title/ ) print $(i+1)}'
```

Explicació:

1. amb `awk -F '<|>'` treiem el que hi ha entre `< i >`:

```
<div id="Definition" class="resultDefinition">
<h2 xmlns:fo="http://www.w3.org/1999/XSL/Format">
<span class="title">
silf
</span><span class="title">
síl fide
</span>
```

Qeuda així:

```
div id="Definition" class="resultDefinition"

h2 xmlns:fo="http://www.w3.org/1999/XSL/Format"

span class="title"
silf
/span
```

El que fem es buscar quan un camp es **span class="title"** imprimim el següent

amb aquesta consulta ens dona la segona part de la línia, la de "-ada"

```
grep 'id="Definition"' 0043250|sed -e 's/^. *title">\([^ ]*\).*$/\1/'

-ada
```

Amb això treiem la primera i la segona definició:

```
grep 'id="Definition"' 0043250|awk -F'class="title">|<' {'print $5'}
alesiat
grep 'id="Definition"' 0043250|awk -F'class="title">|<' {'print $8'}
```

alesiada

Busquem amb awk que el paràmetre \$8 comenci per - i escrivim la paraula, que es el paràmetre 5 i el sufix que es el 8

```
grep 'id="Definition"' *|awk -F'class="title">|<' '{if ( $8 ~ /^-/ ) printf $5$8"\n"}
```

Ens dona una llista com aquesta:

```
romandalusí -ina
reconegut -uda
-nervi -nèrvia
boreoalpí -ina
galileà -ana
glotalitzat -ada
este -a
morenc -a
```

A partir d'aquí busquem patrons, per exemple:

```
este -a
morenc -a
```

Seria este/esta i morenc/morena en un trec una lletra i en un altre no. Poter es si la paraula acaba amb vocal

Treiem tots els sufixos:

```
grep 'id="Definition"' *|awk -F'class="title">|<' '{if ( $8 ~ /^-/ ) printf $8"\n"}'|sort | uniq
```

Treiem sufixos de sufixos, es a dir:

```
-adelf -adelfa
-agog -agoga
-àlgic -àlgica
-àmbul -àmbula
-andre -andra
```

Treiem els que comencin amb - dient: \$5 !~ /^-/

```
grep 'id="Definition"' *|awk -F'class="title">|<' '{if ( $8 ~ /^-/ && $5 !~ /^-/ ) printf $5 $8"\n"}'|wc -l
```

Busquem sufix -a i que la paraula acabi amb e:

```
grep 'id="Definition"' *|awk -F'class="title">|<' '{if ( $8 ~ /^-a / && $5 ~ /e $/ ) printf $5$8"\n"}'
```

O els que acaben amb vocal:

```
grep 'id="Definition"' *|awk -F'class="title">|<' '{if ( $8 ~ /^-a / && $5 ~
/[aeiou] $/ ) printf $5$8"\n"}'
```

Buscar definicions de sufixes:

```
#!/bin/bash
>paraules_amb_sufix
while read linea
do
    paraula=`echo $linea |awk {'print $1'}`
    sufix=`echo $linea |awk {'print $2'}| sed 's/[- ]//g'`
    if [ $sufix == "a" ] || [ $sufix == "ana" ]
    then
        #si acaba amb consonant, no trec res i si acaba amb consonant trec 1
        caràcter
        echo ${paraula: -1}
        echo ${paraula: -1} | grep [aeiou]
        if [[ "`echo ${paraula: -1} | grep [aeiouàèìòùáéíóüäëïöüâêîôû]`" == ""
    ]]
    then
        echo "edípic-a"
        let resta=0
    else
        let resta=1
    fi
    elif [ $sufix == "ina" ] || [ $sufix == "una" ] || [ $sufix == "ona" ] ||
[ $sufix == "ena" ]
    then
        let resta=1
    elif [ $sufix == "ea" ] || [ $sufix == "issa" ] || [ $sufix == "essa" ] ||
[ $sufix == "ossa" ]
    then
        let resta=2
    else
        let resta=$((#sufix)-1)
    fi
    if [ $resta -eq 0 ]
    then
        echo ${paraula}$sufix" "${paraula}"-"$sufix >> paraules_amb_sufix
    else
        echo ${paraula::-${resta}}$sufix" "${paraula}"-"$sufix >>
paraules_amb_sufix
    fi
# echo
done < <( grep 'id="Definition"' *0 |awk -F'class="title">|<' '{if ( $8 ~
/^-/ && $5 !~ /^-/ ) printf $5 $8"\n"}'')
```

```
while read linea
do
    #abans treia les tildes perquè feia el POST a la mateixa URL, ara faig amb
-X POST
    #paraula=`echo $linea|sed 's/[àáä]/a/g' | sed 's/[èéë]/e/g' | sed
's/[íïî]/i/g' | sed 's/[òóö]/o/g' | sed 's/[ùúü]/u/g' | sed 's/·//g'`

    echo $linea
    until torify curl -s -o definicions_sufix/$linea -X POST --data
"DecEntradaText=${linea}" "https://dlc.iec.cat/Results"
    do
        echo "Reiniciando"
        sudo /etc/init.d/tor restart
    done
done < <(cat paraules_amb_sufix |awk {'print $1'})
```

From:

<http://wiki.legido.com/> - **Legido Wiki**

Permanent link:

<http://wiki.legido.com/doku.php?id=informatica:linux:diccionarios:catala>



Last update: **2022/02/28 21:40**