

SUPSI

Ambienti Operativi: Espressioni Regolari 2

Amos Brocco, Ricercatore, DTI / ISIN

grep

- grep può filtrare le righe in input in base a un'espressione regolare
 - Se viene trovato almeno una corrispondenza, la linea viene stampata
 - Noi utilizzeremo grep con l'opzione -E (extended regular expression) che supporta i quantificatori `?`, `+`, `{n}`, `{n,m}`, e `{n,}` (senza l'escaping)
- Le espressioni regolari devono, di norma, essere inserite tra apici semplici o doppi:
 - Se vogliamo che Bash esegua l'espansione delle variabili all'interno dell'espressione regolare usiamo i doppi apici (virgolette)
 - Altrimenti solo `' '`

[X] bash

```
utente@host:~$ grep -E "\bmani\b" testo.txt
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

```
utente@host:~$ grep -E "\bl[aei]\b" testo.txt
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
```

```
utente@host:~$ grep "l\{2,3\}" testo.txt
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
```

```
utente@host:~$ grep -E "l{2,3}" testo.txt
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
```

Caratteri speciali, escaping: grep vs. grep -E

grep	grep -E
x\+	x+
x\?	x?
a\b	a b
\(gruppo\)	(gruppo)
\{m\}	{m}
\{m,\}	{m,}
\{,n\}	{,n}
\{m,n\}	{m,n}

Un nuovo problema...

Cerco la parola “mani”...



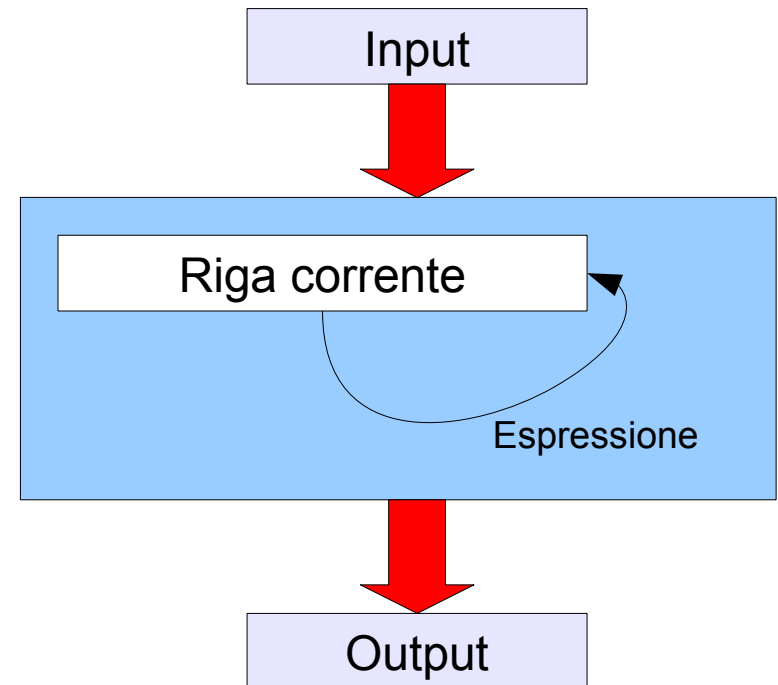
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire. Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi, ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.

...e voglio sostituirla con “piedi”



Come funziona sed

- L'editing avviene attraverso delle espressioni che contengono i comandi per sed
- L'input viene letto riga per riga
 - I comandi sono eseguiti su tutte le righe (se non è indicato un indirizzo specifico)
 - Se sono presenti più comandi, il risultato di ogni espressione viene passato come input al comando successivo
 - L'input non viene modificato



Sintassi di sed

- La sintassi del comando sed è
 - **sed [opzioni] -e 'espressione' inputfile**
 - L'espressione indica il comando di manipolazione del testo da eseguire
 - Posso avere più “espressioni”

Se non viene specificato un file di input, sed legge da stdin

Escaping nelle espressioni regolari in sed

grep	grep -E	sed
x\+	x+	x\+
x\?	x?	x\?
a\b	a b	a\b
\(gruppo\)	(gruppo)	\(gruppo\)
\{m\}	{m}	\{m\}
\{m,\}	{m,}	\{m,\}
\{,n\}	{,n}	\{,n\}
\{m,n\}	{m,n}	\{m,n\}

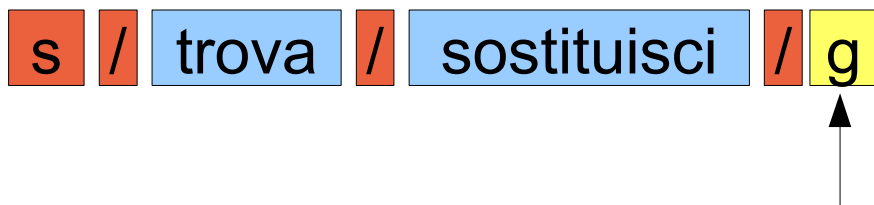
Esempio: sostituzione (s///)

- Il comando **s///** mi permette di sostituire un termine (un'espressione regolare) con un'altro
 - Esempio:

[X] bash

```
utente@host:~$ sed -e 's/\bmani\b/piedi/' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le piedi, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le  
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,  
ma domani possiamo esser poveri. E non si misero in cammino a piedi vuote.
```



Opzionale, permette di effettuare una sostituzione di tutte le corrispondenze di 'trova' nella linea corrente (e non solo la prima occorrenza)

Sostituzione globale o no?

[X] bash

```
utente@host:~$ sed -e 's/le/LE/' testo.txt
```

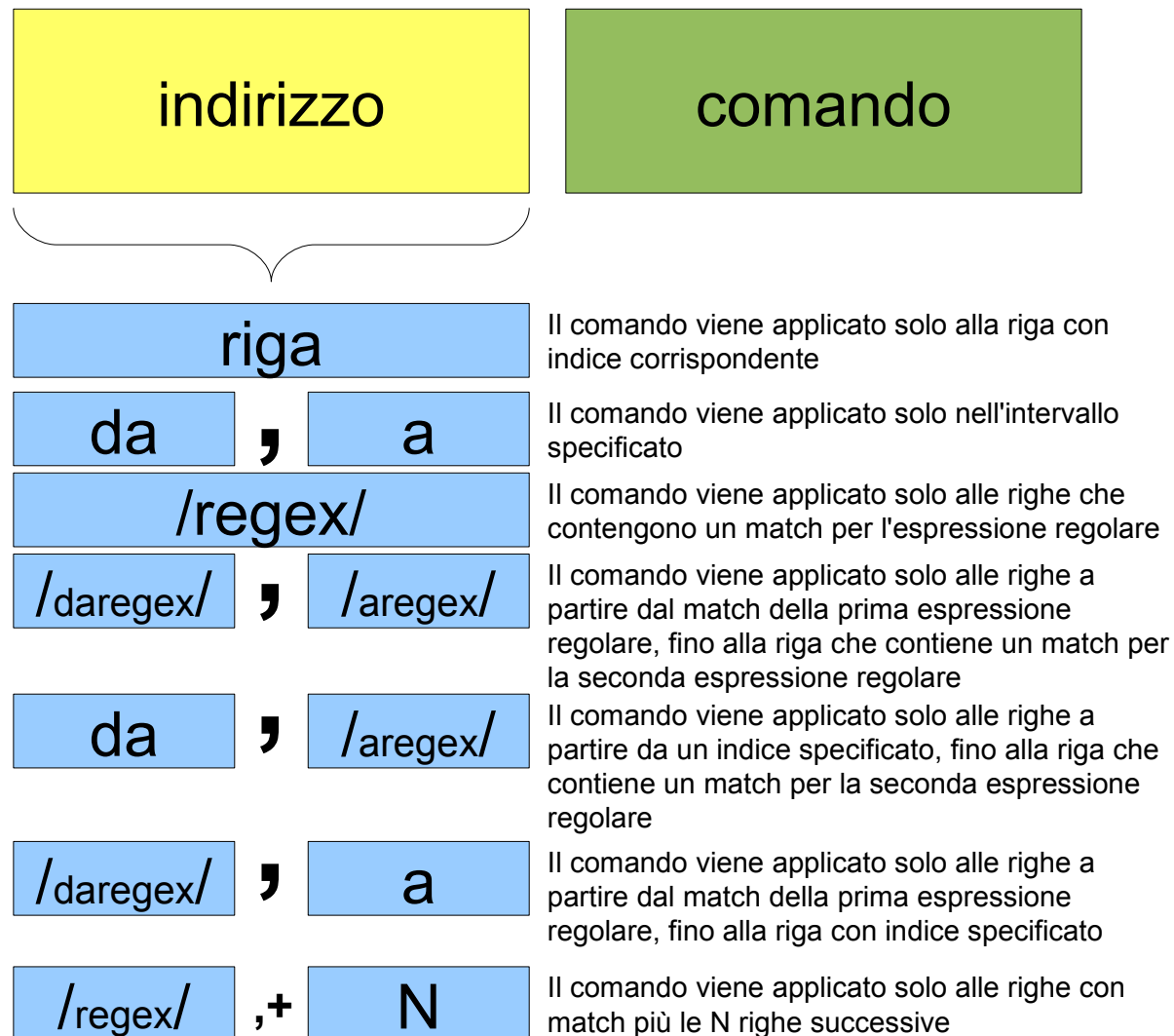
Il 16 febbraio del 1951, sedevo su un panchettino di **LE**gno, ai suoi piedi, puntavo i gomiti sul**LE** sue ginocchia e con la faccia appoggiata tra **le** mani, stavo a sentire. Gli Egiziani trasmisero ai Romani **LE** preparazioni che permettevano di trasformar **le** fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi, ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.

[X] bash

```
utente@host:~$ sed -e 's/le/LE/g' testo.txt
```

Il 16 febbraio del 1951, sedevo su un panchettino di **LE**gno, ai suoi piedi, puntavo i gomiti sul**LE** sue ginocchia e con la faccia appoggiata tra **LE** mani, stavo a sentire. Gli Egiziani trasmisero ai Romani **LE** preparazioni che permettevano di trasformar **LE** fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi, ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.

Limitare l'esecuzione di un comando solo a certe righe



Esempio: stampa di una linea con 'p'

- Il comando **p** mi permette di stampare una riga (o più righe)

Stampa le righe da 1 a 3

[X] bash

```
utente@host:~$ sed -n -e '1,3p' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
```

Con l'opzione -n evitiamo che sed stampi le righe, se non esplicitamente richiesto (con il comando 'p')

[X] bash

```
utente@host:~$ sed -n -e '/faccia/p' testo.txt
```

```
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
```

[X] bash

```
utente@host:~$ sed -n -e '1,/Romani/p' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
```

[X] bash

```
utente@host:~$ sed -n -e '/Romani/,+1p' testo.txt
```

```
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le  
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
```

Esempio: cancellazione di una linea con 'd'

- Il comando **d** mi permette di cancellare una riga (o più righe)
 - Le righe non vengono cancellate dal file di input, semplicemente non appariranno nell'output

[X] bash

```
utente@host:~$ sed -e '/faccia/d' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le  
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,  
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

[X] bash

```
utente@host:~$ sed -e '4,5d' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
```

[X] bash

```
utente@host:~$ sed -e '/Egiziani/,+1d' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.  
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

Sostituire solo in alcune righe

[X] bash

```
utente@host:~$ sed -e '2s/le/LE/g' testo.txt
```

Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i gomiti sulLE sue ginocchia e con la faccia appoggiata tra LE mani, stavo a sentire. Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi, ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.

Sostituisce 'le' con 'LE' solo nella seconda riga

[X] bash

```
utente@host:~$ sed -e '/\bc\w\w\b/s/a/e/g' testo.txt
```

Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i gomiti sulle sue ginocchie e con le feccie eppoggiate tre le meni, stevo e sentire. Gli Egizieni tresmisero ei Romeni le preperezioni che permettevano di tresformer le fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi, ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.

Sostituisce 'a' con 'e' solo nelle righe in cui ci sono parole di tre lettere che cominciano con 'c'

Sostituzione tra un gruppo di caratteri

- Con il comando **y** è possibile effettuare sostituzioni tra due gruppi di caratteri (di lunghezza uguale), carattere per carattere

[X] bash

```
utente@host:~$ sed -e 'y/sleo/5130/' testo.txt
I1 16 f3bbrai0 d31 1951, 53d3v0 5u un panch3ttin0 di 13gn0, ai 5u0i pi3di, puntav0 i
g0miti 5u113 5u3 gin0cchia 3 c0n 1a faccia app0ggiata tra 13 mani, 5tav0 a 53ntir3.
G1i Egiziani tra5mi53r0 ai R0mani 13 pr3parazi0ni ch3 p3rm3tt3van0 di tra5f0rmar 13
fibr3 v3g3ta1i d31 papir0 in 5up3rfici pu1it3, bianch3, pi3gh3v01i. Oggi 5iam0 ricchi,
ma d0mani p055iam0 3553r p0v3ri. E n0n 5i mi53r0 in cammin0 a mani vu0t3.
```

y / s l e o / 5 1 3 0 /

I caratteri vengono sostituiti
uniformemente

Più espressioni

- Posso eseguire più comandi specificando più parametri “expression” (-e)
 - Per ogni riga (riga corrente), i comandi verranno eseguiti nell'ordine indicato

[X] bash

```
utente@host:~$ sed -e '/\bc\w\w\b/s/a/e/g' -e '3,5s/o/u/' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchie e con le feccie eppoggiate tre le meni, stevo e sentire.  
Gli Egizieni tresmiseru ei Romeni le preperezioni che permettevono di tresformer le  
fibre vegetali del papiru in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,  
ma dumani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

`/\bc\w\w\b/s/a/e/g`

`3,5s/o/u/`

Input

Riga corrente

Riga corrente

Riga corrente

Output

Definire le espressioni in un file separato

- È possibile definire le espressioni da utilizzare in un file separato
 - Utile se dobbiamo usare tante trasformazioni
- L'opzione **-f** mi permette di definire il file con le espressioni da utilizzare

[X] bash

```
utente@host:~$ sed -f trasformazioni.sed testo.txt
```


Più comandi per gli stessi indirizzi

- È possibile raggruppare più comandi da eseguire sugli stessi indirizzi (nell'ordine indicato) utilizzando le parentesi graffe **{ }**

[X] bash

```
utente@host:~$ sed -f trasformazioni.sed testo.txt
Il cifra:1cifra:6 febbraio del cifra:1cifra:9cifra:5cifra:1, sedevo su un penchettino di legno, ei
suoi piedi, puntevo i
gomiti sulle sue ginocchie e con le feccie eppoggiate tre le meni, stevo e sentire.
Gli Egizieni tresmisero ei Romeni le preperezioni che permettevono di tresformer le
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

- Dove 'trasformazioni.sed' contiene:

```
1,3 {
    s/a/e/g
    s/[0-9]/cifra:&/g
}
```

Inserire una riga prima di ogni riga

- Il comando `i\` mi permette di inserire una riga (o più righe) prima di ogni riga
 - Ogni riga aggiuntiva termina con `\` (backslash)

[X] bash

```
utente@host:~$ sed -f trasformazioni.sed testo.txt
    Riga aggiunta
    Altra riga
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i
    Riga aggiunta
    Altra riga
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
    Riga aggiunta
    Altra riga
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

- Dove 'trasformazioni.sed' contiene:
1,3 {
 i\
 Riga aggiunta \
 Altra riga
}

Inserire una riga dopo di ogni riga

- Il comando **a** mi permette di aggiungere una riga (o più righe) dopo di ogni riga
 - Ogni riga aggiuntiva termina con **** (backslash)

[X] bash

```
utente@host:~$ sed -f trasformazioni.sed testo.txt
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i
    Riga aggiunta
    Altra riga
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
    Riga aggiunta
    Altra riga
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
    Riga aggiunta
    Altra riga
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

- Dove 'trasformazioni.sed' contiene:
1,3 {
 a\
 Riga aggiunta \
 Altra riga
}

Riprendere la corrispondenza dell'espressione regolare

- Con **&** posso riprendere il valore dell'espressione regolare utilizzata nel campo '*trova*' del comando di sostituzione 's'

[X] bash

```
utente@host:~$ sed -e 's/./Linea: &/g' testo.txt
Linea: Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i
Linea: gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
Linea: Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
Linea: fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
Linea: ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

[X] bash

```
utente@host:~$ sed -e 's/./ -- & -- /g' testo.txt
I -- l  -- 16 febbraio de -- l  -- 1951, sedevo su un panchettino di  -- le -- gno, ai suoi piedi,
puntavo i
gomiti su -- ll -- e sue ginocchia e con  -- la -- faccia appoggiata tra  -- le -- mani, stavo a
sentire.
G -- li -- Egiziani trasmisero ai Romani  -- le -- preparazioni che permettevano di trasformar  --
le --
fibre vegeta -- li -- de -- l  -- papiro in superfici pu -- li -- te, bianche, pieghevo -- li -- .
Oggi siamo ricchi,
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

Back-reference

- Posso referenziare i gruppi utilizzati nel campo '*trova*' del comando di sostituzione 's' nel campo '*sostituisci*' con la notazione ***\N***
 - dove N è il gruppo da referenziare

[X] bash

```
utente@host:~$ sed -e 's/\(w\)\(w\)/\2\1/g' testo.txt
lI 61 efbbaroi edl 9115, esedov us nu apcnehttnio id elngo, ia usio ipdei, uptnvaio i
ogimit uslle use igoncciha e ocn al afccai paopggaiat rta el amin, tsvao a estnrie.
lGi gEziaiin rtsaimesor ia oRamin el rpperazaoiin hce epmrteetavon id rtsaofmrria el
ifrbe evegatil edl apipor ni usepfrcii upilet, ibnahce, ipgeehovil. gOig ismao irccih,
am odamin opssaioim seesr opevir. E onn is imesor ni acmmnio a amin uvtoe.
```

L'ordine di tutte le sequenze di due caratteri alfanumerici è invertito

Stampare i numeri di linea

- Il comando **=** mi permette di stampare i numeri di linea

[X] bash

```
utente@host:~$ sed -e '=' testo.txt
```

```
1
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i
```

```
2
```

```
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
```

```
3
```

```
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le
```

```
4
```

```
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
```

```
5
```

```
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

Concatenare più righe

[X] bash

```
utente@host:~$ sed -e 's/le fibre/LE FIBRE/g' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar le  
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,  
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

Non funziona perché sed lavora su una riga per volta!

Concatenare più righe

- Con il comando **N** posso concatenare due righe consecutive
 - Attenzione! La riga corrente conterrà il carattere **\n** in corrispondenza della fine della prima riga
 - Con un'espressione regolare appropriata posso individuare il pattern desiderato

[X] bash

```
utente@host:~$ sed -e 'N' -e 's/le\s*fibre/LE FIBRE/g' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginocchia e con la faccia appoggiata tra le mani, stavo a sentire.
```

```
Gli Egiziani trasmisero ai Romani le preparazioni che permettevano di trasformar LE FIBRE vegetali  
del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,  
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```


Indirizzo inverso

- Se aggiungo **!** all'indirizzo posso invertire le corrispondenze

```
[X] bash
```

```
utente@host:~$ sed -n -e '/le/!p' testo.txt
fibre vegetali del papiro in superfici pulite, bianche, pieghevoli. Oggi siamo ricchi,
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

Stampa le righe che non contengono 'le'

```
[X] bash
```

```
utente@host:~$ sed -n -e '/gomiti/,/bianche/!p' testo.txt
Il 16 febbraio del 1951, sedevo su un panchettino di legno, ai suoi piedi, puntavo i
ma domani possiamo esser poveri. E non si misero in cammino a mani vuote.
```

Stampa le righe che non sono nell'intervallo specificato

Delimitatori "custom"

- Invece di / è possibile utilizzare, nelle sostituzioni, un qualsiasi carattere delimitatore
 - utile per non dover fare l'escape ogni volta se il delimitatore appare nell'espressione

[X] bash

```
utente@host:~$ sed -e 's/c/a/g' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panahettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginoaahia e aon la faaaia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni ahe permettevano di trasformar le  
fibre vegetali del papiro in superfiai pulite, bianahe, pieghevoli. Oggi siamo riaahi,  
ma domani possiamo esser poveri. E non si misero in aammino a mani vuote.
```

[X] bash

```
utente@host:~$ sed -e 's@c@a@g' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panahettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginoaahia e aon la faaaia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni ahe permettevano di trasformar le  
fibre vegetali del papiro in superfiai pulite, bianahe, pieghevoli. Oggi siamo riaahi,  
ma domani possiamo esser poveri. E non si misero in aammino a mani vuote.
```

[X] bash

```
utente@host:~$ sed -e 'siciaig' testo.txt
```

```
Il 16 febbraio del 1951, sedevo su un panahettino di legno, ai suoi piedi, puntavo i  
gomiti sulle sue ginoaahia e aon la faaaia appoggiata tra le mani, stavo a sentire.  
Gli Egiziani trasmisero ai Romani le preparazioni ahe permettevano di trasformar le  
fibre vegetali del papiro in superfiai pulite, bianahe, pieghevoli. Oggi siamo riaahi,  
ma domani possiamo esser poveri. E non si misero in aammino a mani vuote.
```