

NAAM

dos2unix – DOS/Mac naar Unix en vice versa tekstbestand formaat omzetter

OVERZICHT

```
dos2unix [opties] [BESTAND ...] [-n INVOERBESTAND UITVOERBESTAND ...]  
unix2dos [opties] [BESTAND ...] [-n INVOERBESTAND UITVOERBESTAND ...]
```

BESCHRIJVING

Het Dos2unix pakket bevat toepassingen `dos2unix` en `unix2dos` om platte tekstbestanden in DOS of Mac formaat naar Unix om te zetten en vice versa.

In DOS/Windows tekstbestanden bestaat een regeleinde uit een combinatie van twee tekens: een 'Carriage Return' (CR) gevolgd door een 'Line Feed' (LF). In Unix tekstbestanden bestaat een regeleinde uit een enkel 'Newline' teken die gelijk is aan een DOS 'Line Feed' (LF) teken. In Mac tekst bestanden, voor Mac OS X, bestaan regeleindes uit een enkel 'Carriage Return' teken. Mac OS X is Unix gebaseerd en heeft dezelfde regeleindes als Unix.

Binaire bestanden worden automatisch overgeslagen, behalve als de omzetting geforceerd wordt.

Niet-reguliere bestanden, zoals mappen en FIFOs, worden automatisch overgeslagen.

Symbolische koppelingen en hun doelen blijven standaard onaangeroerd. Optioneel kunnen symbolische koppelingen worden vervangen, of de uitvoer kan naar het doel van de symbolische koppeling worden geschreven. Symbolische koppelingen op Windows worden niet ondersteund. Windows symbolische koppelingen worden altijd vervangen, de doelen onaangeroerd gelaten.

Dos2unix is gemodelleerd naar `dos2unix` onder SunOS/Solaris en heeft gelijke conversiemodi.

OPTIES

- Behandel alle volgende opties als bestandsnamen. Gebruik deze optie als je bestanden wil converteren wiens naam met een streepje begint. Bijvoorbeeld, om een bestand genaamd “-foo” om te zetten, gebruik je de volgende opdracht:

```
dos2unix -- -foo
```

Of in nieuw-bestand-modus:

```
dos2unix -n -- -foo uit.txt
```

-ascii

Converteer alleen regeleindes. Dit is de standaard conversiemodus.

-iso

Conversie tussen DOS en ISO-8859-1 lettertekenreeks. Zie ook paragraaf CONVERSIEMODI.

-1252

Gebruik Windows codetabel 1252 (West-Europees).

-437

Gebruik DOS codetabel 437 (VS). Dit is de standaard codetabel die gebruikt wordt bij ISO conversie.

-850

Gebruik DOS codetabel 850 (West-Europees).

-860

Gebruik DOS codetabel 860 (Portugees).

-863

Gebruik DOS codetabel 863 (Canadees Frans).

-865

Gebruik DOS codetabel 865 (Scandinavisch).

- 7 Converteer 8 bit lettertekens naar 7 bit spatie.

-c, --convmode CONVMODUS

Stel conversiemodus in. Waarbij CONVMODUS een is van: *ascii*, *7bit*, *iso*, *mac* met *ascii* als standaard instelling.

-f, --force

Forceer omzetting van binaire bestanden.

-h, --help

Laat help tekst zien.

-k, --keepdate

Behoud de datum van het invoerbestand in het uitvoerbestand.

-L, --license

Laat de software licentie zien.

-l, --newline

Voeg een extra regeleinde toe.

dos2unix: Alleen DOS regeleindes worden omgezet naar twee Unix regeleindes. In Mac modus worden alleen Mac regeleindes omgezet naar twee Unix regeleindes.

unix2dos: Alleen Unix regeleindes worden omgezet naar twee DOS regeleindes. In Mac modus worden Unix regeleindes omgezet naar twee Mac regeleindes.

-m, --add-bom

Schrijf een UTF-8 Byte Order Mark in het uitvoerbestand. Gebruik deze optie nooit als de codering het van uitvoerbestand geen UTF-8 is. Zie ook paragraaf UNICODE.

-n, --newfile INVOERBESTAND UITVOERBESTAND ...

Nieuw-bestand-modus. Converteer bestand INVOERBESTAND en schrijf naar bestand UITVOERBESTAND. Bestandsnamen moeten opgegeven worden in paren. Jokertekens moeten NIET gebruikt worden, anders verlies je bestanden.

De persoon die de conversie start in nieuw-bestand (gepaarde) modus wordt de eigenaar van het geconverteerde bestand. De lees/schrijf permissies van het nieuwe bestand worden de permissies van het originele bestand minus de *umask* (1) van de persoon die de conversie draait.

-o, --oldfile BESTAND ...

Oud-bestand-modus. Converteer bestand BESTAND en overschrijf het. Dit is de standaard modus. Jokertekens kunnen gebruikt worden.

In oud-bestand (plaatsvervangende) modus krijgt het geconverteerde bestand dezelfde eigenaar, groep en lees/schrijf rechten als het originele bestand. Ook wanneer het bestand wordt omgezet door een andere gebruiker die schrijfrechten heeft op het bestand (b.v. gebruiker root). De omzetting wordt afgebroken wanneer het niet mogelijk is de originele waardes te behouden. Verandering van eigenaar kan betekenen dat de originele eigenaar het bestand niet meer kan lezen. Verandering van groep zou een veiligheidsrisico kunnen zijn, het bestand zou leesbaar kunnen worden voor personen voor wie het niet bestemd is. Behoud van eigenaar, groep en lees/schrijf rechten is alleen ondersteund op Unix.

-q, --quiet

Stille werking. Onderdruk alle waarschuwingen. De teruggave waarde is nul. Behalve wanneer verkeerde opties worden gegeven.

-s, --safe

Sla binaire bestanden over (standaard).

-F, --follow-symlink

Volg symbolische koppelingen en converteer de doelen.

-R, --replace-symlink

Vervang symbolische koppelingen door geconverteerde bestanden (originele doelbestanden blijven ongewijzigd).

-S, --skip-symlink

Laat symbolische koppelingen en doelen ongewijzigd (standaard).

-V, --version

Laat versie informatie zien.

MAC MODUS

In normale modus worden DOS regeleindes naar Unix omgezet en vice versa. Mac regeleindes worden niet omgezet.

In Mac modus worden Mac regeleindes naar Unix omgezet en vice versa. DOS regeleindes blijven ongewijzigd.

Om in Mac modus te draaien gebruik de schakeloptie `-c mac` of gebruik de opdrachten `mac2unix` of `unix2mac`.

CONVERSIEMODI

Conversie-modi *ascii*, *7bit*, en *iso* zijn gelijk aan die van dos2unix/unix2dos onder SunOS/Solaris.

ascii

In modus *ascii* worden alleen regeleindes omgezet. Dit is de standaard modus.

Hoewel de naam van deze modus ASCII is, wat een 7 bit standaard is, is de werkelijke modus 8 bit. Gebruik altijd deze modus wanneer u Unicode UTF-8 bestanden omzet.

7bit

Alle 8 bit niet-ASCII lettertekens (met waardes van 128 t/m 255) worden omgezet naar een 7 bit spatie.

iso Lettertekens worden omgezet tussen een DOS lettertekenreeks (codetabel) en ISO lettertekenreeks ISO-8859-1 op Unix. DOS lettertekens zonder gelijkwaardig ISO-8859-1 teken, waarvoor geen omzetting mogelijk is, worden omgezet in een punt. Het zelfde geldt voor ISO-8859-1 tekens zonder DOS tegenhanger.

Wanneer alleen optie `-iso` gebruikt wordt, zal dos2unix proberen de actieve codetabel te gebruiken. Als dat niet mogelijk is wordt codetabel CP437 gebruikt, die met name in de VS gebruikt word. Om een bepaalde codetabel te forceren gebruik opties `-850` (West-Europees), `-860` (Portugees), `-863` (Canadees Frans) of `-865` (Scandinavisch). Windows codetabel CP1252 (West-Europees) wordt ook ondersteund met optie `-1252`. Gebruik voor andere codetabellen dos2unix in combinatie met *iconv* (1). Iconv kan omzetten tussen een lange lijst letterteken-coderingen.

Gebruik ISO conversie nooit op Unicode tekst bestanden. Het zal UTF-8 gecodeerde bestanden beschadigen.

Enkele voorbeelden:

Omzetten van DOS standaard codetabel naar Unix Latin-1

```
dos2unix -iso -n in.txt uit.txt
```

Omzetten van DOS CP850 naar Unix Latin-1

```
dos2unix -850 -n in.txt uit.txt
```

Omzetten van Windows CP1252 naar Unix Latin-1

```
dos2unix -1252 -n in.txt uit.txt
```

Omzetten van Windows CP1252 naar Unix UTF-8 (Unicode)

```
iconv -f CP1252 -t UTF-8 in.txt | dos2unix > uit.txt
```

Omzetten van Unix Latin-1 naar DOS standaard code tabel

```
unix2dos -iso -n in.txt uit.txt
```

Omzetten van Unix Latin-1 naar DOS CP850

```
unix2dos -850 -n in.txt uit.txt
```

Omzetten van Unix Latin-1 naar Windows CP1252

```
unix2dos -1252 -n in.txt uit.txt
```

Omzetten van Unix UTF-8 (Unicode) naar Windows CP1252

```
unix2dos < in.txt | iconv -f UTF-8 -t CP1252 > uit.txt
```

Zie ook <http://czyborra.com/charsets/codepages.html> en <http://czyborra.com/charsets/iso8859.html>.

UNICODE

Coderingen

Er bestaan verschillende Unicode coderingen. Op Unix en Linux zijn Unicode bestanden typisch gecodeerd in UTF-8. Op Windows kunnen Unicode tekst bestanden gecodeerd zijn in UTF-8, UTF-16 of UTF-16 big endian, maar meestal zijn ze gecodeerd in UTF-16 formaat.

Conversie

Unicode tekst bestanden kunnen DOS, Unix of Mac geregeleindes hebben, net als reguliere tekst bestanden.

Alle versies van dos2unix en unix2dos kunnen UTF-8 gecodeerde bestanden omzetten, want UTF-8 is ontworpen op compatibiliteit met ASCII.

Dos2unix en unix2dos met Unicode UTF-16 ondersteuning, kunnen little en big endian UTF-16 gecodeerde tekst bestanden lezen. Om er achter te komen of dos2unix gebouwd is met UTF-16 ondersteuning type `dos2unix -V`.

De Windows versies van dos2unix en unix2dos converteren UTF-16 gecodeerde bestanden altijd naar UTF-8 geconcodeerde bestanden. Unix versies van dos2unix/unix2dos zetten UTF-16 gecodeerde bestanden om naar de lokale karakter codering als die gelijk is aan UTF-8. Gebruik de opdracht *locale* (1) om uit te vinden wat de lokale karakter codering is.

Omdat UTF-8 geformateerde tekstbestanden zowel op Windows en Unix goed ondersteund worden, hebben dos2unix en unix2dos geen optie om UTF-16 bestanden te schrijven. Alle UTF-16 karakters kunnen worden omgezet naar UTF-8. De omzetting van UTF-16 naar UTF-8 is verliesvrij. Op Unix worden UTF-16 bestanden overgeslagen als de locale karakter codering ongelijk is aan UTF-8, om te voorkomen dat er per ongeluk tekst verloren gaat. Wanneer een UTF-16 naar UTF-8 conversiefout optreedt, bijvoorbeeld wanneer het UTF-16 invoerbestand een fout bevat, wordt het bestand overgeslagen.

ISO en 7-bit conversie werkt niet op UTF-16 bestanden.

Byte Order Mark

Op Windows hebben Unicode tekstbestanden typisch een Byte Order Mark (BOM), omdat veel Windows programma's (inclusief Kladblok) standaard een BOM toevoegen. Zie ook http://en.wikipedia.org/wiki/Byte_order_mark.

Op Unix hebben Unicode tekstbestanden typisch geen BOM. Er wordt aangenomen dat de codering van tekstbestanden gelijk is aan de lokale karakter codering.

Dos2unix kan alleen detecteren of een bestand in UTF-16 formaat is als het bestand een BOM heeft. Wanneer een UTF-16 bestand geen BOM heeft, ziet dos2unix het bestand als een binair bestand.

Gebruik dos2unix in combinatie met *iconv* (1) om een UTF-16 bestand zonder BOM om te zetten.

Dos2unix schrijft nooit een BOM in het uitvoerbestand, tenzij optie *-m* gebruikt wordt.

Unix2dos schrijft een BOM in het uitvoerbestand wanneer het invoerbestand een BOM heeft, of wanneer optie *-m* gebruikt is.

Unicode voorbeelden

Omzetten van Windows UTF-16 (met BOM) naar Unix UTF-8

```
dos2unix -n in.txt uit.txt
```

Omzetten van Windows UTF-16 (zonder BOM) naar Unix UTF-8

```
iconv -f UTF-16 -t UTF-8 in.txt | dos2unix > uit.txt
```

Omzetten van Unix UTF-8 naar Windows UTF-8 met BOM

```
unix2dos -m -n in.txt uit.txt
```

Omzetten van Unix UTF-8 naar Windows UTF-16

```
unix2dos < in.txt | iconv -f UTF-8 -t UTF-16 > uit.txt
```

VOORBEELDEN

Lees invoer van 'stdin' en schrijf uitvoer naar 'stdout'.

```
dos2unix
dos2unix -l -c mac
```

Omzetten en vervangen a.txt. Omzetten en vervangen b.txt.

```
dos2unix a.txt b.txt
dos2unix -o a.txt b.txt
```

Omzetten en vervangen a.txt in ascii conversiemodus.

```
dos2unix a.txt
```

Omzetten en vervangen a.txt in ascii conversiemodus. Omzetten en vervangen b.txt in 7bit conversiemodus.

```
dos2unix a.txt -c 7bit b.txt
dos2unix -c ascii a.txt -c 7bit b.txt
dos2unix -ascii a.txt -7 b.txt
```

Omzetten a.txt van Mac naar Unix format.

```
dos2unix -c mac a.txt
mac2unix a.txt
```

Omzetten a.txt van Unix naar Mac format.

```
unix2dos -c mac a.txt
unix2mac a.txt
```

Omzetten en vervangen a.txt met behoud van originele datum.

```
dos2unix -k a.txt
dos2unix -k -o a.txt
```

Omzetten a.txt en schrijf naar e.txt.

```
dos2unix -n a.txt e.txt
```

Omzetten a.txt en schrijf naar e.txt, behoud datum e.txt gelijk aan a.txt.

```
dos2unix -k -n a.txt e.txt
```

Omzetten en vervangen a.txt. Omzetten b.txt en schrijf naar e.txt.

```
dos2unix a.txt -n b.txt e.txt
dos2unix -o a.txt -n b.txt e.txt
```

Omzetten c.txt en schrijf naar e.txt. Omzetten en vervangen a.txt. Omzetten en vervangen b.txt. Omzetten d.txt en schrijf naar f.txt.

```
dos2unix -n c.txt e.txt -o a.txt b.txt -n d.txt f.txt
```

RECURSIEVE CONVERSIE

Gebruik dos2unix in combinatie met de *find*(1) en *xargs*(1) opdrachten om tekstbestanden in een directoryboomstructuur recursief om te zetten. Bijvoorbeeld om alle .txt bestanden in de directoryboom onder de huidige map te converteren type:

```
find . -name *.txt |xargs dos2unix
```

LOKALISATIE

LANG

The primaire taal wordt geselecteerd met de omgevingsvariabele LANG. De LANG variabele bestaat uit verschillende onderdelen. Het eerste deel is in kleine letters de taalcode. Het tweede deel is optioneel en is de landcode in hoofdletters, voorafgegaan door een laag streepje. Er is ook een optioneel derde deel: lettertekenreeks-codering, voorafgegaan door een punt. Enkele voorbeelden voor een POSIX standaard shell:

export LANG=nl	Nederlands
export LANG=nl_NL	Nederlands, Nederland
export LANG=nl_BE	Nederlands, België
export LANG=es_ES	Spaans, Spanje
export LANG=es_MX	Spaans, Mexico
export LANG=en_US.iso88591	Engels, VS, Latin-1 codering
export LANG=en_GB.UTF-8	Engels, GB, UTF-8 codering

Voor een complete lijst van taal- en landcodes zie de gettext handleiding:
<http://www.gnu.org/software/gettext/manual/gettext.html#Language-Codes>
<http://www.gnu.org/software/gettext/manual/gettext.html#Language-Codes>

Op Unix systemen kunt u de opdracht *locale* (1) gebruiken om specifieke lokalisatie informatie te verkrijgen.

LANGUAGE

Met de omgevingsvariabele LANGUAGE kunt u een prioriteitenlijst specificeren van talen, gescheiden door dubbele punten. Dos2unix geeft voorkeur aan LANGUAGE boven LANG. Bijvoorbeeld, eerst Nederlands en dan Duits: LANGUAGE=nl:de. U moet eerst lokalisatie in werking stellen, met het instellen van LANG (of LC_ALL) in een waarde ongelijk aan "C", voordat u een talen prioriteitenlijst kunt gebruiken via de LANGUAGE variabele. Zie ook de gettext handleiding:
<http://www.gnu.org/software/gettext/manual/gettext.html#The-LANGUAGE-variable>
<http://www.gnu.org/software/gettext/manual/gettext.html#The-LANGUAGE-variable>

Als u een taal kiest die niet beschikbaar is worden de standaard Engelse berichten gebruikt.

DOS2UNIX_LOCALEDIR

Met de omgevingsvariabele DOS2UNIX_LOCALEDIR kan de LOCALEDIR gebruikt tijdens compilatie worden overstemd. LOCALEDIR wordt gebruikt om de taalbestanden te vinden. De GNU standaard waarde is /usr/local/share/locale. De optie "-V" laat de gebruikte LOCALEDIR zien.

Voorbeeld (POSIX shell):

```
export DOS2UNIX_LOCALEDIR=$HOME/share/locale
```

TERUGGAVE WAARDE

Bij succes wordt nul terug gegeven. Wanneer een systeemfout optreedt wordt het laatste systeemfoutnummer terug gegeven. Bij andere fouten wordt 1 terug gegeven.

De teruggave is altijd nul in de modus stille werking, behalve wanneer verkeerde opties worden gegeven.

STANDAARDEN

<<http://nl.wikipedia.org/wiki/Tekstbestand>>
 <http://nl.wikipedia.org/wiki/Carriage_Return>
 <<http://nl.wikipedia.org/wiki/Linefeed>>
 <<http://nl.wikipedia.org/wiki/Unicode>>

AUTEURS

Benjamin Lin – <blin@soecs.uts.edu.au>, Bernd Johannes Wuebben (mac2unix modus) – <wuebben@kde.org>, Christian Wurl (voeg extra regeleinde toe) – <wurl@ira.uka.de>, Erwin

Waterlander – <waterlan@xs4all.nl>

Project pagina: <<http://waterlan.home.xs4all.nl/dos2unix.html>>

SourceForge pagina: <<http://sourceforge.net/projects/dos2unix/>>

Freecode: <<http://freecode.com/projects/dos2unix>>

ZIE OOK

file (1) find (1) iconv (1) locale (1) xargs (1)