Conversion table for **DOS Code Page 850 (DOS850)** to URL percent encoding (UTF-8).

Character	DOS850 hex	Unicode	UTF-8 hex	URL encoding
ä	84	U+00E4	C3 A4	%C3%A4
å	86	U+00E5	C3 A5	%C3%A5
ö	94	U+00F6	C3 B6	%C3%B6
Ä	8E	U+00C4	C3 84	%C3%84
Å	8F	U+00C5	C3 85	%C3%85
Ö	99	U+00D6	C3 96	%C3%96
é	82	U+00E9	C3 A9	%C3%A9
è	8A	U+00E8	C3 A8	%C3%A8
ü	81	U+00FC	C3 BC	%C3%BC
Ü	9A	U+00DC	C3 9C	%C3%9C
á	A0	U+00E1	C3 A1	%C3%A1
à	85	U+00E0	C3 A0	%C3%A0
ñ	A4	U+00F1	C3 B1	%C3%B1
ß	E1	U+00DF	C3 9F	%C3%9F

The six Nordic characters to remember

Character	DOS850	URL encoding
å	\x86	%C3%A5
ä	\x84	%C3%A4
ö	\x94	%C3%B6
Å	\x8F	%C3%85
Ä	\x8E	%C3%84
Ö	\x99	%C3%96

Important distinction

- \x84 = hexadecimal DOS850 byte notation.
- %C3%A4 = URL encoding of the UTF-8 representation of ä.
- %84 = URL encoding of the raw DOS850 byte, **not** the UTF-8 encoding.

Thus, when converting DOS850 filenames to UTF-8 URL encoding, the conversion is:
DOS850 → Unicode → UTF-8 → URL encoding
This is the distinction that matters when preparing URLs for your HTML pages.

s/\x84/%%C3%%A4/g
Tämä on **säännöllisen lausekkeen (regex) korvauskomento**, joka muuttaa CP437- tai CP850-merkistön **ä-kirjaimen** URL-koodattuun UTF-8-muotoon. Komento etsii kaikki tavun 0x84 esiintymät ja korvaa ne merkkijonolla %%C3%%A4.
Komennon osat

- **s/** aloittaa korvausoperaation (substitute).
- **\x84** määrittelee etsittävän merkin heksadesimaaliarvona, joka vastaa pientä **ä-kirjainta** vanhoissa DOS-merkistöissä (kuten CP437 ja CP850).
- **%%C3%%A4** on korvaava merkkijono, jossa C3 A4 on pienen **ä-kirjaimen** heksadesimaalinen esitys UTF-8-merkistössä. Kaksinkertaiset prosenttimerkit (%%) viittaavat siihen, että komentoa ajetaan Windowsin komentokehotteessa (.bat-tiedostossa), jossa yksittäinen prosenttimerkki täytyy suojata toisella prosenttimerkillä.
- **/g** (global) määrää, että korvaus tehdään rivin jokaiselle osumalle eikä vain ensimmäiselle.

Käyttötarkoitus
Tätä komentoa käytetään tyypillisesti työkaluissa kuten sed tai Perl, kun vanhaa DOS-pohjaista tekstiä muunnetaan verkkoselaimille tai API-rajapinnoille sopivaan URL-koodattuun (percent-encoded) muotoon.
Yleiset skandimerkit samassa muodossa
Jos muunnat suomenkielistä tekstiä samalla menetelmällä, tässä ovat vastaavat arvot muille skandinaavisille merkeille:

Merkki	DOS-heksa (CP437/850)	URL-koodattu UTF-8
ä	\x84	%C3%A4
Ä	\x8E	%C3%84
ö	\x94	%C3%B6
Ö	\x99	%C3%96
å	\x86	%C3%A5
Å	\x8F	%C3%85

Haluatko, että kirjoitan valmiin **komentoriviskriptin** tai **säännöllisen lausekkeen**, joka muuntaa kaikki suomen kielen erikoismerkit kerralla?
Tekoälyn antamat vastaukset voivat sisältää virheitä. Lue lisää

Tässä on **laajennettu muunnostaulukko**, joka kattaa suomenkielisissä teksteissä yleisimmin esiintyvät erikoismerkit, lainausmerkit ja välimerkit. Taulukko vertaa vanhaa DOS-merkistöä (CP437/CP850) ja yleistä Windows-merkistöä (Windows-1252 / ISO-8859-1) URL-koodattuun UTF-8-muotoon.

Laajennettu merkistötaulukko

Merkki	Kuvaus	DOS (CP437/850)	Windows-1252 / ISO-8859-1	URL-koodattu UTF-8
ä	Pieni ä	\x84	\xE4	%C3%A4
Ä	Iso Ä	\x8E	\xC4	%C3%84
ö	Pieni ö	\x94	\xF6	%C3%B6
Ö	Iso Ö	\x99	\xD6	%C3%96
å	Pieni å	\x86	\xE5	%C3%A5
Å	Iso Å	\x8F	\xC5	%C3%85
ü	Saksalainen ü	\x81	\xFC	%C3%BC
Ü	Saksalainen Ü	\x9A	\xDC	%C3%9C
é	E-akuutti	\x82	\xE9	%C3%A9
É	Iso E-akuutti	\x90	\xC9	%C3%89
§	Pykälämerkki	\x15 (tai \xF5 CP850)	\xA7	%C2%A7
½	Puolikas	\xAB	\xBD	%C2%BD
°	Astemerkki	\xF8	\xB0	%C2%B0
€	Euro	– *(ei tuettu)*	\x80	%E2%82%AC
–	Ajatusviiva	– *(ei tuettu)*	\x96	%E2%80%93
”	Lainausmerkki	– *(ei tuettu)*	\x94	%E2%80%9D

Käyttövinkit

- **Ympäristön huomiointi:** Jos käytät komentoja Windowsin .bat-tiedostossa, korvaa jokainen yksittäinen %-merkki kahdella prosenttimerkillä (%%). Esimerkiksi muoto %C3%A4 kirjoitetaan silloin muodossa %%C3%%A4.
- **Merkistön tunnistus:** Jos tekstissäsi näkyy ä-kirjaimen tilalla merkki „, kyseessä on DOS-merkistö (\x84). Jos tilalla näkyy kysymysmerkki tai tyhjä ruutu ja tavu on E4, kyseessä on Windows-1252- tai ISO-8859-1-merkistö.

Haluatko, että kokoan näistä valmiin **monirivisen komennon** valitsemallesi työkalulle (esimerkiksi sed, Perl tai PowerShell)?
Tekoälyn antamat vastaukset voivat sisältää virheitä. Lue lisää
Laajennettu

Tässä on **kattava laajennettu muunnostaulukko**, joka sisältää lähes kaikki länsieurooppalaisissa kielissä, teknisissä dokumenteissa ja tietokantasiirroissa esiintyvät erikoismerkit.
Taulukko kattaa suomen, ruotsin, saksan, ranskan, espanjan ja norjan merkit sekä yleisimmät erikoissymbolit. CP850 on yleisin länsieurooppalainen DOS-merkistö (esim. vanha suomenkielinen MS-DOS), ja CP437 on alkuperäinen yhdysvaltalainen DOS-merkistö.
Kattava merkistötaulukko (DOS, Windows-1252 ja URL-koodattu UTF-8)

1. Pohjoismaiset ja saksalaiset merkit

Merkki	Kuvaus	DOS (CP850)	DOS (CP437)	Windows-1252 / ISO-8859-1	URL-koodattu UTF-8
ä	Pieni ä	\x84	\x84	\xE4	%C3%A4
Ä	Iso Ä	\x8E	\x8E	\xC4	%C3%84
ö	Pieni ö	\x94	\x94	\xF6	%C3%B6
Ö	Iso Ö	\x99	\x99	\xD6	%C3%96
å	Pieni å	\x86	\x86	\xE5	%C3%A5
Å	Iso Å	\x8F	\x8F	\xC5	%C3%85
ü	Pieni ü	\x81	\x81	\xFC	%C3%BC
Ü	Iso Ü	\x9A	\x9A	\xDC	%C3%9C
ß	Saksalainen kaksois-s	\xE1	\xE1	\xDF	%C3%9F
ø	Tanskalainen/norjalainen ø	\x9B	–	\xF8	%C3%B8
Ø	Iso Ø	\x9D	–	\xD8	%C3%98
æ	Tanskalainen/norjalainen æ	\x91	\x91	\xE6	%C3%A6
Æ	Iso Æ	\x92	\x92	\xC6	%C3%86

2. Ranskan, espanjan ja muiden kielten aksenttimerkit

Merkki	Kuvaus	DOS (CP850)	DOS (CP437)	Windows-1252 / ISO-8859-1	URL-koodattu UTF-8
é	E-akuutti	\x82	\x82	\xE9	%C3%A9
É	Iso E-akuutti	\x90	\x90	\xC9	%C3%89
è	E-gravis	\x8A	\x8A	\xE8	%C3%A8
È	Iso E-gravis	\xD4	–	\xC8	%C3%88
ê	E-sirkumfleksi	\x88	\x88	\xEA	%C3%AA
à	A-gravis	\x85	\x85	\xE0	%C3%A0
á	A-akuutti	\xA0	\xA0	\xE1	%C3%A1
í	I-akuutti	\xA1	\xA1	\xED	%C3%AD
ó	O-akuutti	\xA2	\xA2	\xF3	%C3%B3
ú	U-akuutti	\xA3	\xA3	\xFA	%C3%BA
ñ	Espanjalainen ñ	\xA4	\xA4	\xF1	%C3%B1
Ñ	Iso Ñ	\xA5	\xA5	\xD1	%C3%91
ç	Sedilji-c	\x87	\x87	\xE7	%C3%A7
Ç	Iso Ç	\x80	\x80	\xC7	%C3%87

3. Valuutat, välimerkit ja tekniset symbolit

Merkki	Kuvaus	DOS (CP850)	DOS (CP437)	Windows-1252	URL-koodattu UTF-8
€	Euro	–	–	\x80	%E2%82%AC
£	Punta	\x9C	\x9C	\xA3	%C2%A3
¥	Jeni	\xBE	\x9D	\xA5	%C2%A5
§	Pykälämerkki	\xF5	\x15	\xA7	%C2%A7
°	Astemerkki	\xF8	\xF8	\xB0	%C2%B0
±	Plus-miinus	\xF1	\xF1	\xB1	%C2%B1
²	Potenssiin kaksi	\xFD	\xFD	\xB2	%C2%B2
³	Potenssiin kolme	\xFC	–	\xB3	%C2%B3
µ	Mikro-merkki	\xE6	\xE6	\xB5	%C2%B5
½	Puolikas	\xAB	\xAB	\xBD	%C2%BD
¼	Neljäsosa	\xAC	\xAC	\xBC	%C2%BC
¾	Kolme neljäsosaa	\xF3	–	\xBE	%C2%BE
©	Tekijänoikeus (Copyright)	\xB8	–	\xA9	%C2%A9
®	Rekisteröity tavaramerkki	\xA9	–	\xAE	%C2%AE
™	Tavaramerkki (Trademark)	–	–	\x99	%E2%84%A2
–	Ajatusviiva (En-dash)	–	–	\x96	%E2%80%93
—	Pitkä ajatusviiva (Em-dash)	–	–	\x97	%E2%80%94
”	Oikea lainausmerkki	–	–	\x94	%E2%80%9D
“	Vasen lainausmerkki	–	–	\x93	%E2%80%9C
’	Typografinen heittomerkki	–	–	\x92	%E2%80%99
«	Kulmalainausmerkki vasen	\xAE	\xAE	\xAB	%C2%AB
»	Kulmalainausmerkki oikea	\xAF	\xAF	\xBB	%C2%BB
¿	Ylösalainen kysymysmerkki	\xA8	\xA8	\xBF	%C2%BF
¡	Ylösalainen huutomerkki	\xAD	\xAD	\xA1	%C2%A1

4. URL-rakenteen kannalta kriittiset merkit (ASCII)
Jos merkkijono sisältää tavallisia välimerkkejä, ne on usein myös koodattava URL-osoitteita varten (nämä ovat samat kaikissa merkistöissä):

Merkki	Kuvaus	Heksa (ASCII / DOS / Win)	URL-koodattu UTF-8
(välilyönti)	Välilyönti	\x20	%20 (tai + lomakedatassa)
%	Prosenttimerkki	\x25	%25 *(ajettava ensimmäisenä!)*
&	Et-merkki	\x26	%26
+	Plusmerkki	\x2B	%2B
/	Kauttaviiva	\x2F	%2F
?	Kysymysmerkki	\x3F	%3F
=	Yhtäsuuruusmerkki	\x3D	%3D

Tärkeitä huomioita korvauskomentojen ajamiseen

- **Suoritusjärjestys:** Jos korvaat myös prosenttimerkkejä (\x25 -> %25), aja se **aina ensimmäisenä** ennen muita korvauksia. Muuten komento rikkoo jo luodut %C3%A4-koodaukset muotoon %25C3%25A4.
- **Windows Batch (.bat):** Muista tuplata prosenttimerkit korvaavassa osassa (%C3%A4 %%C3%%A4).
- **Automaattinen vaihtoehto:** Jos käytössäsi on työkaluja kuten iconv tai Python, voit muuntaa koko tiedoston kerralla ilman yksittäisiä regex-sääntöjä (esim. iconv -f CP850 -t UTF-8).

Haluatko, että muodostan tästä valmiin **.bat-skriptin** tai **säännöllisten lausekkeiden komentolistan** tietylle merkistölle (esim. CP850 tai Windows-1252)?
Tekoälyn antamat vastaukset voivat sisältää virheitä. Lue lisää