

# UTF-8 & Longpress

What makes Brocade tick 

Alain Descamps & Nelia Van Briel

 a net.be

## Encoding?



Ã¼jÃ³Ã°iÃ¶Ã° = ljslands þjóðlöð

○○○

masquÃ© = masqué

 a net.be

## UTF-8?

- encoding
- karakters -> bytes voor opslag

<u>Teken</u>	<u>Byte</u>
A	65
ë	195, 171
т	224, 175, 177

- encodings: ISO 8859(-1), ASCII, UTF-16, ...

	Karakter nr (Unicode)	ASCII (1 byte)	ISO-Latin 1 (1 byte)	UTF-8 (+1 bytes)
	0 = NULL	0	0	0
...	65 = Capital A	65	65	65
...	127 = DEL	127	127	127

	Karakter nr (Unicode)	ASCII (1 byte)	ISO-Latin 1 (1 byte)	UTF-8 (+1 bytes)
	0 = NULL	0	0	0
...	65 = Capital A	65	65	65
...	127 = DEL	127	127	127
...	255 = Latin Small y with diaeresis	-	255	195 191

	Karakter nr (Unicode)	ASCII (1 byte)	ISO-Latin 1 (1 byte)	UTF-8 (+1 bytes)
	0 = NULL	0	0	0
...	65 = Capital A	65	65	65
...	127 = DEL	127	127	127
...	255 = Latin Small y with diaeresis	-	255	195 191
...	256 = Latin Capital A with macron	-	-	196 128
...	8.734 = Infinity	-	-	226 136 158
...	(1.114.111)			

Karakternummer	Voorstelling	UTF-8
197	Å	195,133
9792	♀	226,153,128
9794	♂	226,153,130



Karakternummer	Voorstelling	UTF-8
197	Å	195,133
9792	♀	226,153,128
9794	♂	226,153,130
9793	ð	226,153,129



Karakternummer	Voorstelling	UTF-8
197	Å	195,133
9792	♀	226,153,128
9794	♂	226,153,130
9793	♂	226,153,129
8550	VII	226,133,166
9814	♖	226,153,150
128561	🙄	240,159,152,177



## Brocade 2000 karakterset - vroeger

- gebaseerd op ISO Latin (8859-1)
  - vb: a, b, c, ë, à, ... (#256)
- andere «03B8»
- 20 jaar in gebruik

## Grieks record

**TITEL(s)**

**Titel**  **Titel**

```
«0395»«03BA»«03BA»«03BB»«03B7»«03C3»«03B9»«03B1»
«03B5»«03C0»«03B9»«03C3»«03B7»«03BC»«03BF»«03BD»
«03B4»«03B5»«03BB»«03C4»«03B9»«03BF»«03BD» «03C4»«03B7»«03C3»
«03B5»«03BA»«03BA»«03BB»«03B7»«03C3»«03B9»«03B1»«03C3»
«03C4»«03B7»«03C3» «0395»«03BB»«03BB»«03B1»«03B4»«03BF»«03C3»
```

Extensie

Type

**IMPRESSUM**

*Impressum 1*

**Uitgever 1** **Uitgever**

Bron

**Plaats 1** **Plaats**

Bron

## Na de release

**TITEL(s)**  Extra titel(s)

**Titel**  Titel

**Extensie**

**Type**

**IMPRESSUM**


**Impressum 1**

**Uitgever 1** **Uitgever**

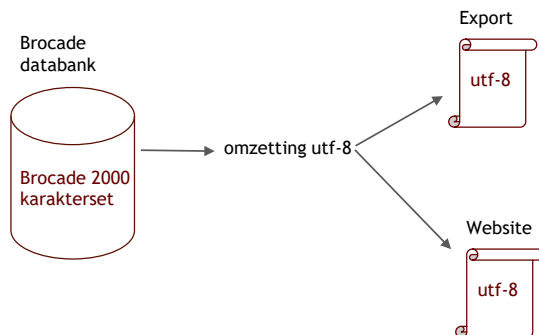
**Bron**

**Plaats 1** **Plaats**

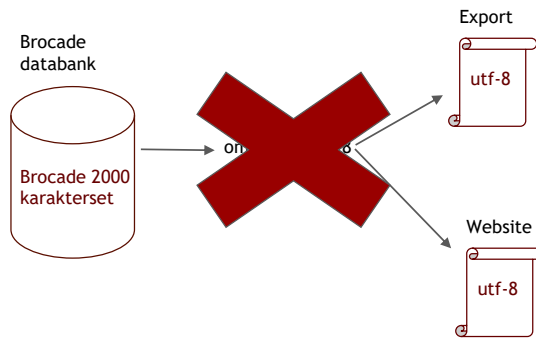
**Bron**



## Brocade 2000 karakterset

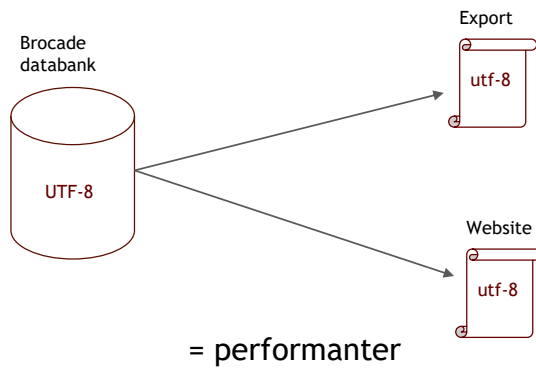


## Brocade 2000 karakterset



a net.be

## UTF-8



a net.be



## Voordelen UTF-8

- open source: 94% websites gebruikt het + browserondersteuning
- beschikbare tools
- uitwisselen data gaat efficiënter  
(import + export + presentatie)
- vlottere ontwikkeling
- what you see is what you get  
θ voordien <<03B8>>  
(- copy paste gaat vlotter ;) )



## Een paar cijfers

- 131.201 UTF-8 karakters in Brocade
- +10% van de code is gewijzigd door 6 ontwikkelaars  
> 49.000 lijnen
- Conversie data  
op productie server  
OF  
scan - scan van wijzigingen (realtime) - verwerken tijdens release  
ipv 2 dagen -> 1u



## Ingave speciale karakters Å

Verskillende manieren:

<<00C5>>

<<Aring>> -> werkt niet meer!

&Aring; = entiteit

&#197; = decimal

&x00C5; of &x212B; of &x0041;&x030A; = hex

\AA

Via toetsenbord

## Normalisatie

Å

=> UTF-8 heeft 3 opties:

U+00C5 ( Å ) latin capital letter A with ring above

U+212B ( Å ) angstrom sign

U+0041 ( A ) latin capital letter A + U+030A ( ° ) combining ring above

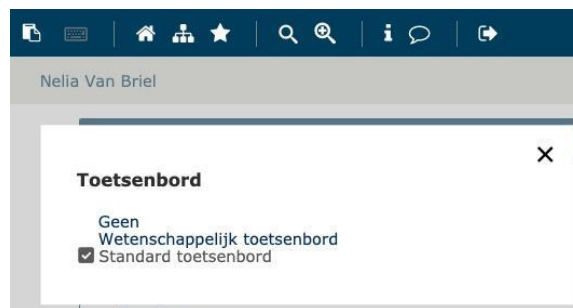
Normalisatie (NFC) = U+00C5

## Toetsenbord / Longpress



a.net.be

## Toetsenbord / Longpress



a.net.be

## Toetsenbord / Longpress

**Titel(s)**  Extra titel(s)

Titel  Titel

Extensie

Type



## Toetsenbord / Longpress

**Titel(s)**  Extra titel(s)

Titel  Titel

Extensie

Type

