

Woordafbreking op ë en ï

Abstract

LaTeX heeft moeite met het afbreken van woorden die een ë bevatten. Dit verhaal laat zien hoe dat probleem op te lossen: gebruik `\"e` of `\{"e}` of `ë` voor een eindstandige ë en `"e` voor alle andere. Analoog voor de ï.

Keywords

hyphenation diacesis umlaut

Wie in het Nederlands een tekst schrijft waar veel landennamen in voorkomen, heeft een gereede kans om problemen te krijgen met namen die op ë eindigen. De gebruikelijke manier om dit karakter in te voeren is door `\"e` te schrijven of, om het helemaal netjes te doen: `\{"e}`. Echter: woordafbreking vindt voor zo'n ë niet plaats en in de volgende voorbeeldtekst is het gevolg veel te veel wit in de tweede zin:

Itali\"e, ge\"eindigd
nooit afbreking vóór \"e

We gaan naar Indië en niet naar België of Italië want we houden van de tropen.
Maar...uiteindelijk zijn we in Zuid-Frankrijk geëindigd.

Bij gebruik van het BABEL package kan men in het Nederlands de ë ook invoeren met als `"e` en zo'n ë gedraagt zich anders (we zullen straks zien waarom): woorden kunnen ervoor afgebroken worden, zelfs, helaas, als die ë de laatste lettergreep van het woord is; het gevolg is een losse e aan het begin van de volgende zin:

Itali"e, ge"eindigd
afbreking vóór "e zelfs als het de laatste lettergreep is

We gaan naar Indië en niet naar België of Italië want we houden van de tropen.
Maar...uiteindelijk zijn we in Zuid-Frankrijk geëindigd.

Een andere mogelijkheid zou kunnen zijn, een kant-en-klare ë uit een latin1-font te gebruiken, maar die gedraagt zich net als `\"e`: woorden worden daarvoor nooit afgebroken:

Italië, geëindigd
nooit afbreking vóór ë

We gaan naar Indië en niet naar België of Italië want we houden van de tropen.
Maar...uiteindelijk zijn we in Zuid-Frankrijk geëindigd.

De conclusie is daarom: gebruik `\"e` voor een eindstandige en `"e` voor alle andere ë's:

Itali\"e, ge"eindigd
de oplossing:

We gaan naar Indië en niet naar België of Italië want we houden van de tropen.
Maar...uiteindelijk zijn we in Zuid-Frankrijk geëindigd.

De discussie betreffende dit probleem op de `tex-nl` mailing list vermocht Johannes Braams, de auteur van BABEL, te verleiden tot het volgende commentaar:

Inderdaad `"e` is gebaseerd op de dubbel-quote als actief teken. Dit stamt uit het 7-bit tijdperk toen `ë` nog niet mogelijk was als invoer. Bovendien zorgde `\"e` ervoor dat een woord in het geheel niet werd afgebroken vanwege het invoegen van het `\accent` primitief. De constructie `"e` doet dus een aantal dingen:

- er wordt een `\hskip` van 0pt ingevoegd waardoor het woord voor TeX in twee woorden uiteenvalt. Daarbij wordt ervoor gezorgd dat als de regel tussen die twee woorden wordt afgebroken toch een afbreekteken verschijnt.
- er verschijnt een trema op de e.
- die trema is trouwens wel degelijk lager gezet dan wat het accent primitief normaal doet.
- Tot het repertoire behoort overigens ook `"i` (geeft `ï`, vroeger als `\{"i}` in te voeren).

Dat het in dit geval met het afbreken misgaat omdat die `ë` in z'n eentje een woord gaat vormen (zo ziet TeX het althans) is ietwat onfortuinlijk. Italië breekt overigens alleen goed af als je ook de 8-bits afbreekpatronen van Piet Tutelaers gebruikt...

Wybo Dekker
wybo@servalys.nl