



# Spellingcontrole in OpenOffice.org



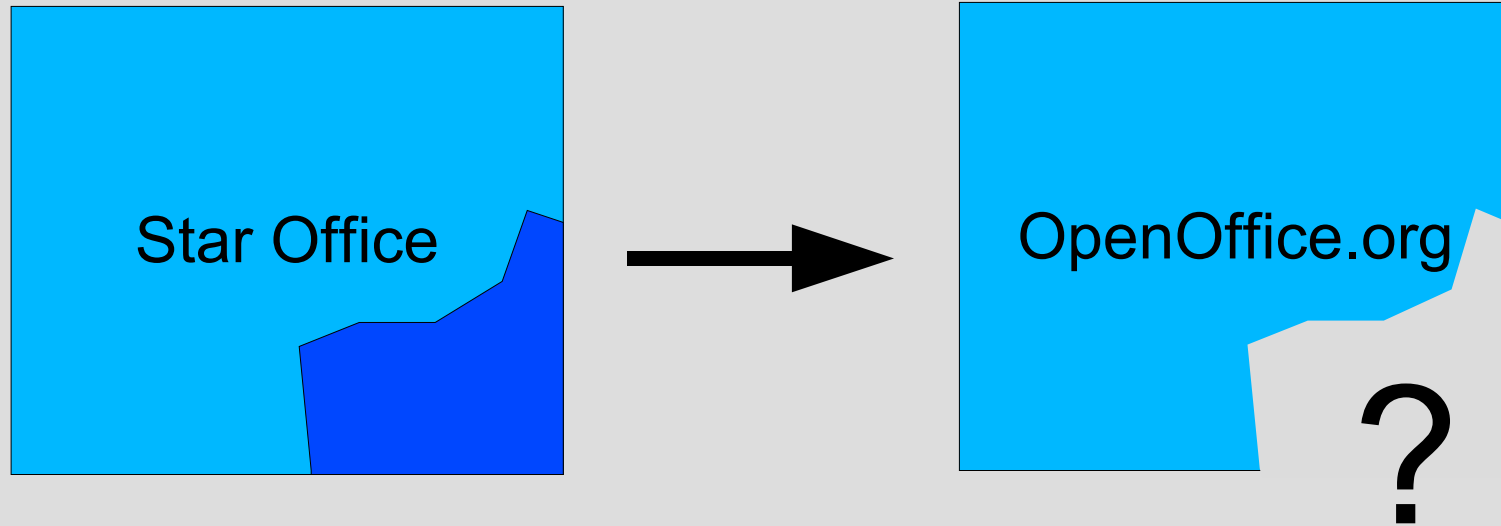
ir. S. Brouwer  
[nl.openoffice.org](http://nl.openoffice.org)



# Overzicht

- Ontstaan van de OpenOffice.org S/W
- Ontwikkeling spellingcontrole
- Nederlandstalige woordenlijst
- Licenties...
- Oprichting [nl.openoffice.org](http://nl.openoffice.org)
- Ontwikkelingen en toekomstperspectief
- Meehelpen
- Tenslotte

# Ontstaan OpenOffice.org



- Spellingcontrole (!)
- Afdrukken onder Unix/Linux (!)
  - Synoniemenlijst
  - Sjablonen, clip-art



# Ontwikkeling spellingcontrole

Kevin B. Hendricks: Myspell ← Ispell

## Affix-compressie (.dic en .aff bestand)

voorbeeld: **spel/DWMIN** :

speld, spelde; spelt, spelde, spelden; spellen; speling;  
spelingen

... **spelling/N** etc.

**Suggesties** 1 letter weghalen, 1 letter toevoegen (TRY list),

1 letter wijzigen, 2 letters omwisselen

Voorbeeld: sugggestie, sugestie, suggesrie, suggetsie ->  
suggestie

Beperking: meerdere letters fout... ?

Voorbeeld: buro, kompuuter (...)



# Nederlandstalige woordenlijst

Gebaseerd op:

Ispell woordenlijst

Nederlandstalige TeX gebruikersgroep

- Affixbestand aanpassen
- Discussie over licenties...
- Oplossing: apart downloaden

# Licenties...

OpenOffice.org: LGPL/SSSL (JCA)

Spellingwoordenlijst: GPL

Niet samenvoegen, wel combineren

- apart downloaden, dictionary.lst aanpassen
- DictInstall (Windows)
- DicOOo (multi platform)
- ?

```
dict nl nl nl
DICT nl NL nl_NL
DICT nl BE nl_NL
DICT nl NL nl mod
```





# Oprichting Lang/NL

Naar aanleiding van de NL spellingcontrole...

Hi,

I am constructing a web site about OpenOffice.org intended to make the project and the product more accessible to the Dutch speaking community. It will contain information in Dutch, links to mirrors that carry Dutch localized builds, etc.

My question is, could I get permission to use the OpenOffice logo on this web site?

[nl.openoffice.org](http://nl.openoffice.org):

Informatie (webpagina's), NL builds,  
mirrors (...),  
vertaling online Help, mailinglists...



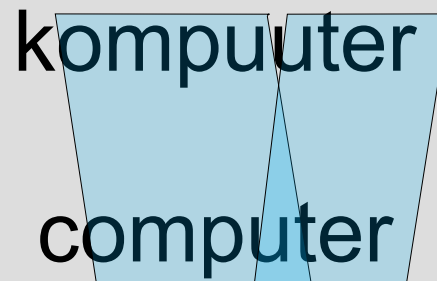
# Verdere ontwikkeling

tot nu toe o.a.:

- uitbreiding/correctie woordenlijst
- betere suggesties met REP

voorbeeld: buro (REP o eau) -> bureau

- betere suggesties met “n-gram scoring”:



komputer  
computer





# Toekomstperspectief

Meest geuite klacht:

**“OOo kent geen samengestelde woorden”**

Voorbeelden:

Cocaïnesmokkelaar -  
Parkmoord -

Korenwolfbumpersticker  
Routenavigatiesysteem  
Spinazievriesdroogfabriek

Machtsoverdracht -  
Rechtshulpverzoek -  
Wederopbouwcontracten -  
Toestemmingswet -  
Telecomwaakhond -  
Internetapotheek -  
Kernwapenprogramma -  
Massavernietigingswapen +

# Toekomstperspectief

Meest geuite klacht:

**“OOo kent geen samengestelde woorden”**  
voorbeeld: “kantoorpakket”  
*suggestie*: “kantoor pakket”.



Oplossing:

- Voor suggestie géén woorden splitsen (OOo 1.1.1)
- samenstellingen herkennen (kantoor+pakket)

Uitdaging: zo goed mogelijk onjuiste samenstellingen  
verwerpen:

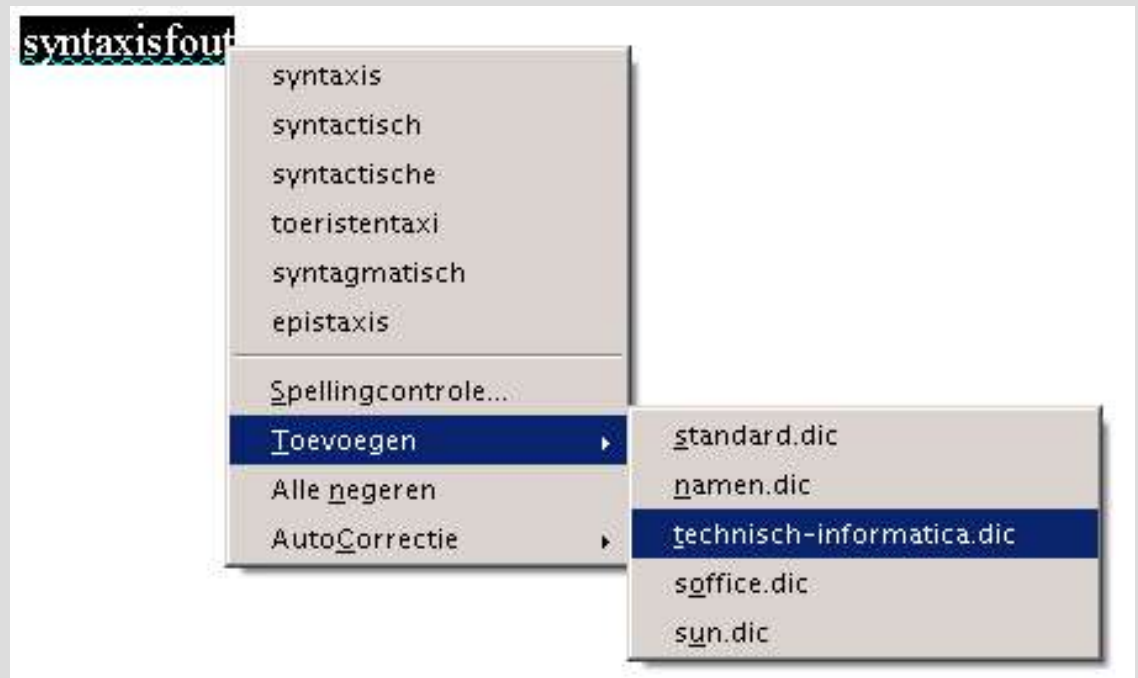
“Kruitkoek”, “Bruidengom”, “Rodenkool”, “Ezelbruggetje”

Op langere termijn: grammatica-controle

# Meehelpen

Gemiddeld krantenbericht: > 99% van de correcte woorden wordt herkend.

Woorden toevoegen aan gebruikerswoordenboek:



Stuur de bestanden op!

<http://nl.openoffice.org/about-helping.html>

Specialistische woordenboeken: Medisch, Technisch, Juridisch, Eigennamen, ...?



# Tenslotte

“ Hei reet op zijn paart over de hij ”

“ Speling controle is niet zalig makend! ”

OpenOffice.org is volop in ontwikkeling

Hierna: Reginald Riquier

Bedankt voor de aandacht!