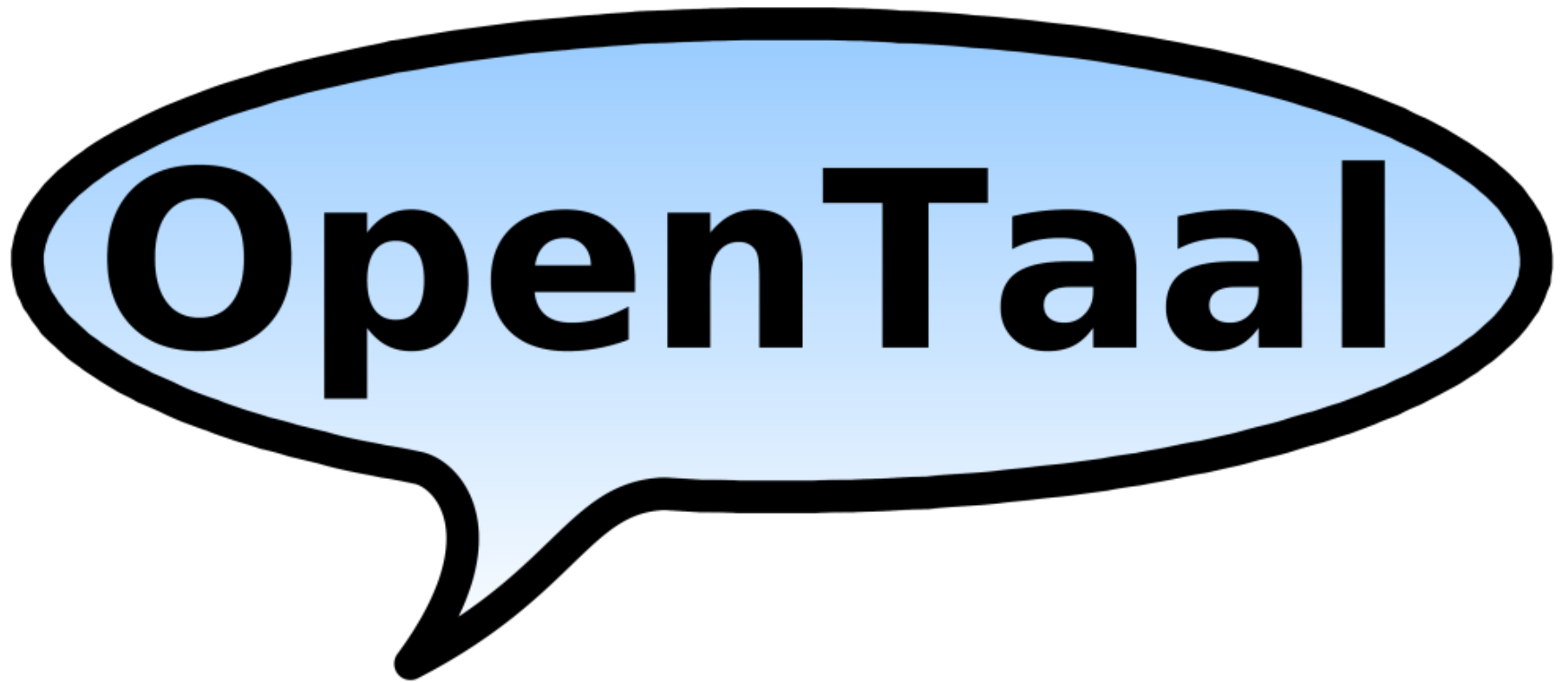


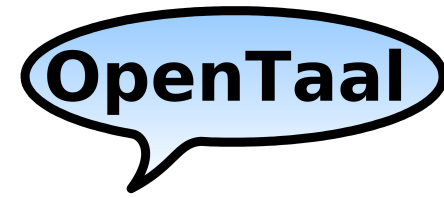
*Taal is van ons allemaal*



Sander van Geloven  
Simon Brouwer  
[www.opentaal.org](http://www.opentaal.org)

OpenTaal  
voorstellen aan de  
Spellingcommissie

# Agenda



- **Inleiding**
  - Achtergrond
  - Doelstelling
  - Licenties
  - Organisatie
- **Projecten en producten**
  - Samenstellen spellingcontrole
  - Woordenlijst
  - Spellingcontrole
  - Voorbeelden gebruik
- **Conclusie**
  - Toekomst
  - Ambities
  - Vragen

# Achtergrond



OpenTaal

Waar komen wij vandaan?

- 1996: groep vrijwilligers NTG maakt een vrije Nederlandstalige spellingwoordenlijst voor gebruik in TeX
- 2000: OpenOffice.org verschijnt. Simon past hierop de bestanden aan
- 2005: Spellinghervorming; bestanden moeten worden aangepast
  - OSOSS (Bart) zoekt contact met NTU.
  - NTU: "Groene boekje niet integraal vrij beschikbaar, wel keurmerk mogelijk"
  - Oprichting vrijwilligersproject OpenTaal: Fabrice Mous (OSOSS, KDE.nl), Bart Knubben (OSOSS), Hans Hagen (NTG), Simon Brouwer (nl.OpenOffice.org), Piet van Oostrum (NTG).
- 2009: oprichting Stichting OpenTaal

# Doelstelling



- Doel: “OpenTaal maakt vrije Nederlandstalige taalhulpbestanden voor gebruik in opensourceprojecten.”
- Uitgangspunt: De Nederlandse taal is dé standaard voor gegevensuitwisseling in Nederland, België en Suriname. Deze standaard zou open moeten zijn.
- Motto: *Taal is van ons allemaal*

# Licenties



- Twee licenties (keuze eindgebruiker):
  - BSD-licentie
  - CC-BY-licentie



- gratis kopiëren, gebruiken, distribueren en afleiden
- naamsvermelding, geen garantie en zelfde licentie

# Organisatie



## Wie is OpenTaal?

- meer dan 50 individuen:
  - mensen uit opensourcegemeenschappen
  - taalwetenschappers
  - taalpuristen en -liefhebbers
- 5 à 10 echte actievelingen:
  - bestuursleden
  - software-ontwikkelaars
- allen zijn vrijwilligers
- crowdsourcing

# Projecten en producten

The logo for 'OpenTaal' is a blue speech bubble with a black outline, containing the text 'OpenTaal' in a bold, black, sans-serif font. The bubble has a tail pointing downwards and to the left.

**OpenTaal**

- Woordenlijst en spellingcontrole
- Grammaticacontrole
- Woordafbreking
- Synoniemenlijst
- Thesaurus

# Samenstellen spellingcontrole



- oogsten
  - alinea's
  - woorden
- beoordelen
  - geogste woorden
- eindredactie
  - beoordeelde woorden
  - gedoneerde woorden
- basiswoorden laten keuren (levert woordenlijsten op)
- spellingcontrolebestand maken





# Woordenlijst

OpenTaal

Tekstbestanden versie 2.10 september 2011

- 164.313 **basiswoorden** (NTU Keurmerk Spelling) bijvoorbeeld werkgroep, lezing, Amsterdam\*
- 157.000 **flexievormen** (ongekeurd) bijvoorbeeld werkgroepje, lezingen, Amsterdamse
- 23.773 **ongekeurde basiswoorden** bijvoorbeeld NTU, Balkenende, Amsterdam-Noord
- 153 **verwarrende woorden** (basis gekeurd) bijvoorbeeld bei, fijt, heef, peg, zijl, verassend

Verwachting versie 2.20, eind 2012:

- 200.000 basis, 200.000 flexievormen, 25.000 ongekeurde basiswoorden, 250 verwarrende woorden en 50.000 toponiemen\*

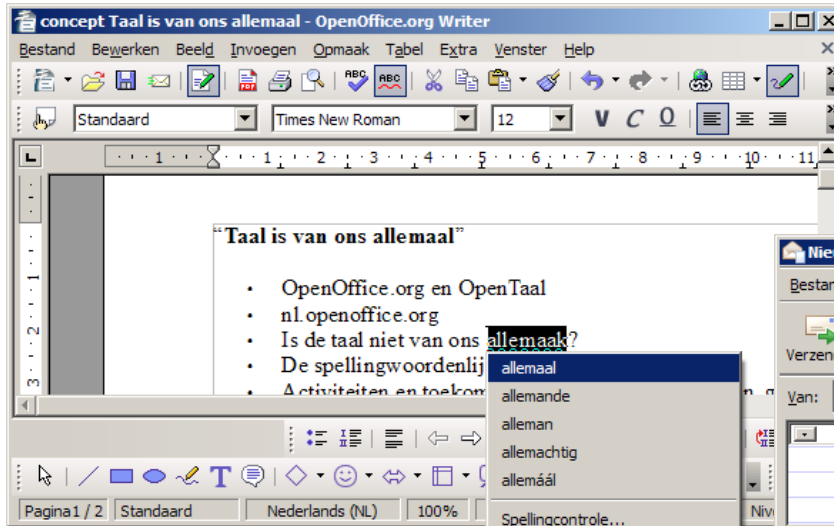
# Spellingcontrole



## Hunspell FOSS-spellingcontrole

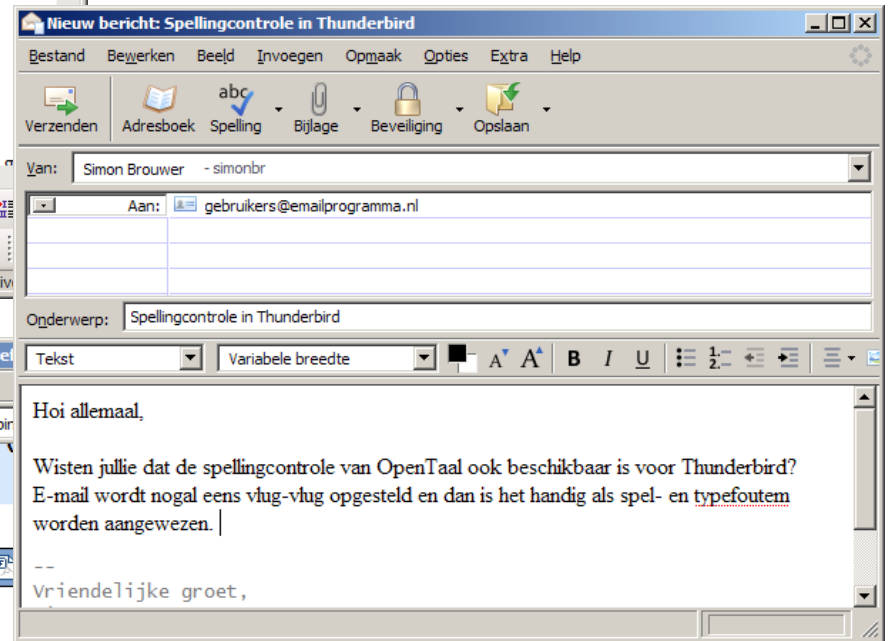
- **grote gebruikersgroep:** LibreOffice, OpenOffice.org, Mozilla Firefox, Mozilla Thunderbird, InDesign, Google Chrome, Google Chromium, Opera, Wiktionary, TinyMCE, LyX, Eclipse, Apache Solr, et cetera
- ondersteunt **samenstellingen**
- maatwerksuggesties voor correcties: ceedees»cd's, enigsinds» enigszins, snietsel»schnitzel, lazanja»lasagne, disleksie»dyslexie, et cetera
- helaas geen ondersteuning spaties in woorden → grammaticaconcontrole
- eenvoudig in te bouwen in software
- subsidie NTU verbeterde ondersteuning Nederlands

# Voorbeelden gebruik



Tekstverwerker

E-mail



Webbrowser

Voorstellen OpenTaal aan Spellingcommissie

# Toekomst



2012

- start partnerprogramma en convenant NTU
- woordenlijst en spellingcontrole 2.20

2013 / 2014

- interne processen en software verbeteren voor spellingcontrole 3.0 en toegankelijker crowdsourcing
- meer samenwerking via internationale standaarden
- afbreekpatronen:
  - eigen internationale standaard IETF RFC
- woordclassificatie en -relaties:
  - FoLiA, ISO 12620 DCR (Data Category Registry), ISO 24613 LMF (Lexical Markup Framework), POS-tagging (Part of Speech), ISOcat

•

# Ambities



Verbeteren kwaliteit en kwantiteit van al onze producten d.m.v. informatieuitwisseling met partners voor meer crowdsourcing en bijdragen in bulk.

- overheid zoals NTU en INL
- wetenschap en onderwijs zoals Radboud Universiteit Nijmegen en Tilburg University
- commercie, bijvoorbeeld via mobiele toepassingen
- onderdak bieden aan open taalinitiatieven



Bedankt voor uw aandacht.



Stel uw vragen en zie ook  
**<http://opentaal.org>**  
of neem contact op via  
**[info@opentaal.org](mailto:info@opentaal.org)**