

# FUN PROJECT BY DESIGN\*

How LibreOffice development can be full of flow?

László Németh

\* The ~ten funniest moments of my recent Numbertext, LibreLogo, Hunspell, LibreOffice and Hungarian spelling dictionary developments



# MEHR LICHT!

Johann Wolfgang von Goethe

# 01 Wikipedia 1-liner

Extend spelling dictionaries: push text of the Wikipedia in a Unix pipeline, ranking their entries by counting their backlinks (~PageRank of Google), and hunspell it.





```
cat wikipedia.xml | grep -Eo '[\w|-|\.]*' | tr -d '[' | sort | uniq -c | sort -rsn | hunspell -l
```



  
**01**

## **cat** wikipedia.xml |

The standard Unix tool *cat* concatenates the given file(s), and pushes the result to the standard output. E.g. only a single 4.6 GB file, text content of the Hungarian Wikipedia. Now extract all "[[wikilink]]s" (linked entries):

**02**

## **grep -Eo '[[\w|-|\.]\*]]'**

The tool *grep* globally searches in file(s) by **regex**. Here **prints --only-matching** text parts, all the single words\* within double brackets, e.g. for "some *[[random]]* *[[text]]* in a *[[text]]* file" gives only "*[[random]]*" and two "*[[text]]*".

**03**

## **| tr -d ']' | sort | uniq -c |**

The tool *tr* **translates** characters, or it can also **--delete** them, as here the opening and closing brackets: "*[[text]]*" → "*text*". The words are sorted by **sort** alphabetically, and **uniq** gives **count** data, removing the duplicates.

**04**

## **sort -rsn | hunspell -l**

Sort words by their frequency using a **reverse stable<sup>†</sup> numeric** sort:

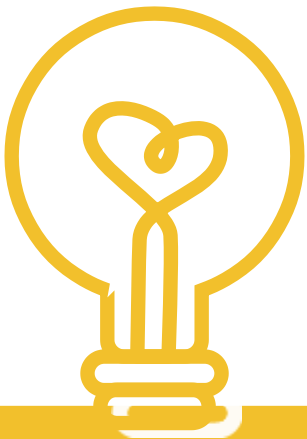
*2 text*

*1 random*

Filter bad words by **hunspell** spell checker.

\* Words can contain letters (\w), dashes and dots ([.]), too.

† It keeps the alphabetical order of the words with the same count.





# Unknown words by frequency

## Examples

- 8044 McLaren (British F1 racing team)
- 7375 LINEAR (Lincoln Near-Earth Asteroid Research)
- 5001 AllMusic (American music database)
- 4803 Rajna-vidék-Pfalz (German state)
- 4822 Billboard (American music chart)
- 4756 Aprilia (Italian motorcycle manufacturer)
- 3588 Cosworth (British car engineering co.)
- 2991 UTC (Coordinated Universal Time)
- 2881 Baden-Württemberg (German state)
- 2835 Kalex (German motorcycle engineering co.)
- 2606 Naucsnij (Crimean Astrophysical Observatory)
- 2504 Schleswig-Holstein (German state)
- 2323 Eutheria (biology: classification of mammals)
- 2265 Palomar (American observatory)
- 2258 Grammy-díj (American music award)
- 2199 Oricon (Japanese media database)
- 2139 Metacritic (American media database)
- 2032 Ducati (Italian motorcycle manufacturer)
- 2002 Index.hu (news site)
- 1900 Brabham (British F1 racing team)
- 1716 Feyenoord (Dutch football club)
- 1715 Laurasiatheria (classification of mammals)
- ...
- 78 mugli (muggle)



## New dictionary words

- 855 anime
- 403 klád (clade)
- 333 24.hu (news site)
- 317 444.hu (news site)
- 272 remake
- 222 brakkvíz (brack water)
- 154 mainstream
- 111 monofiletikus
- 101 kladisztika (cladistics)
- 93 heraldikus
- 91 krill
- 94 parafiletikus
- 74 epifiton
- 73 archeák
- 69 szkúner
- 68 ammoniteszek
- 66 pörgekarúak
- 65 sztratégosz
- 65 szkeleton
- 65 szfalerit
- 64 kaldera
- 63 citokin
- 62 zigomorf
- 52 metamfetamin

### PHYSICS

- 85 Higgs (boson)
- 81 Planck (constant)
- 69 Boltzmann (constant)

### BIOLOGY

- 122 Triceratops
- 35 földimalac (aardvark)



# 02 Soros language

Soros programs translate numbers and money amounts, e.g. 1 → *een* in Dutch, EUR 2 → *deux euros* in French. Complete implementation and LO integration.



# - History of Soros language



2007

Qiang He suggested NUMBERTEXT() instead of BAHTTEXT() (cited by David A. Wheeler on open-formula list)

"...defining a general purpose NUMBERTEXT would be a nightmare" (Eike Rathke)



2009-03

- 1-liner in awk
- Calc pyUNO add-on: NUMBER/MONEYTEXT
- en, th, de, hu, it lang.
- 花碼: 一〇二三  
千
- spin-off of grammar checker development funded by FSF.hu



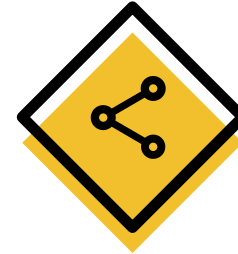
2009-04

- JavaScript port, IDE
- revised spec: \1 → \$1, equivalent of \$(\1)
- 11 new languages: proof of the concept
- funded by NLnet



2018

C++ port; conditional text; prefix macro; l10n lines; tutorial; LibO integration: chapter numbering; NatNum12 format code (Mike Kaganski, Eike Rathke); 37 languages; (FSF.hu)



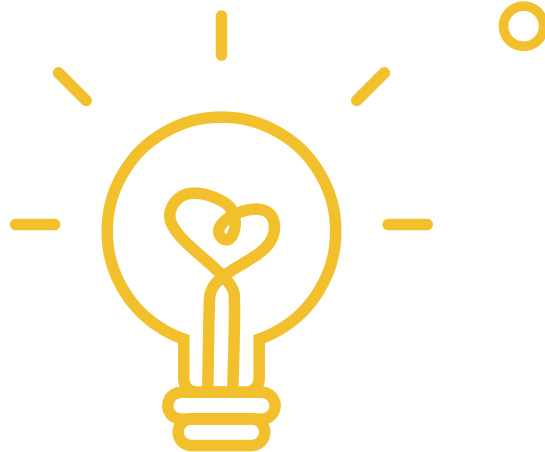
2022

Full implementation of conditional text with boundary inheritance; complete LibreOffice integration of numbertext library: **"spell out" number formats** in cell & field format dialog windows (FSF.hu). Other usage: grammar options; affixation of date fields; articles of page references; hu\_Hung RTL transliteration





`"(?:\uE00A?\uE00A)+\uE00A|?)"`, // optional nested calls



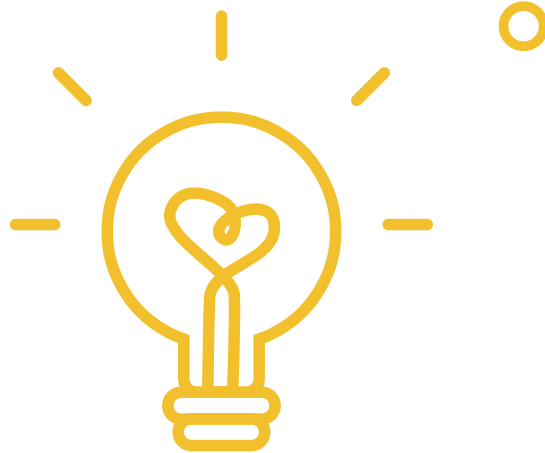
# 03 LibreLogo

Planned localization update and minor fixes resulted much more: fix SVG ghost lines, compiler regression, performance, add font features and portable SVG





`__dispatcher__(".uno:ChangeBezier", (), draw)`

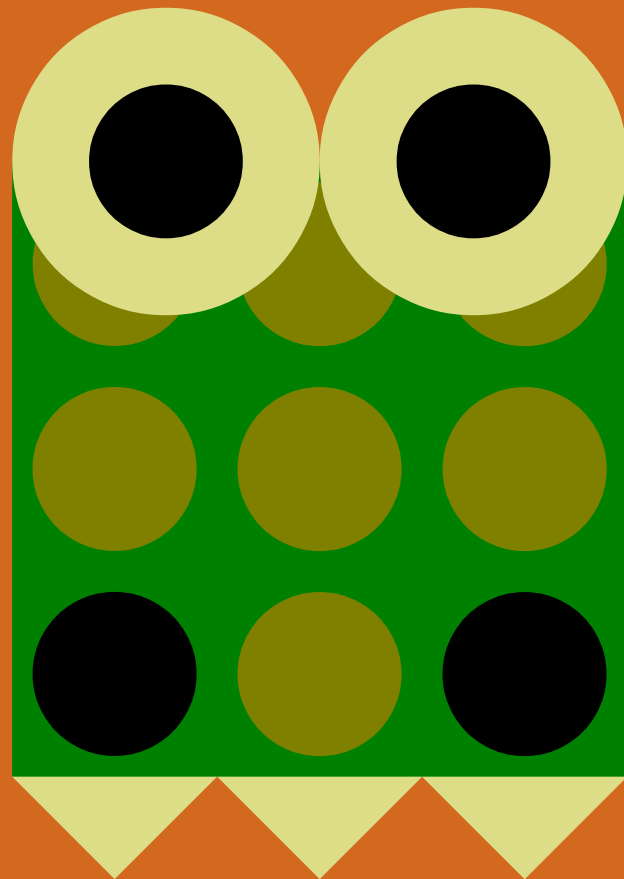


# 04 Books

Improving typesetting of LibreOffice (like line breaking with hyphenation zone), translations of Writer DTP and LibreLogo manual, creating LibreLogo picture books



LOG (mascot of LibreLogo) made  
from 13 circles and 4 squares



**400** bugs fixed

by me for the Hungarian public administration.  
7.5/month in the area of MSO interoperability.





# Commits of 2023

#396

tdf#45949 sw: **add** spell checking to hyperlinked text

Recognize plain words within hyperlinks and check their spelling.

#397

tdf#152964 sw: fix **crash** of Undo of tracked deletion of empty table rows

In Hide Changes mode, Undo of insertion of CH\_TXT\_TRACKED\_DUMMY\_CHAR (workaround of the deletion of empty table rows), resulted a crash because of inconsistency of table and redline nodes. As a workaround, skip that insertion from the Undo.

#400

tdf#152738 sw: fix lost SwPosition in spelling dialog, which resulted **broken** spell checking.

#399

tdf#152493 ucb WebDAV: **fix** upload using HTTP 1.0 fallback

Fix broken libcurl upload to Vibe 4.0.6 WebDAV server using HTTP 1.0 fallback.

#395

tdf#146248 sw: fix **crash** at Undo of hiding page header/footer

When Same content on first page is enabled, Undo could result crashing after hiding header/footer with its local menu.

#398

f#152541 sw: don't show hidden `__RefHeading__` bookmarks in visible bookmark mode (**regression**).

#393

tdf#131203 DOCX import: fix **lost** table when w:tblGrid is missing

Load tables in case of incomplete text table definition, i.e. missing w:tblGrid and its w:gridCol elements, like MSO does.

#394

tdf#152245 sw change tracking: fix copy from Calc to Writer tables

Pasting Calc tables into Writer tables uses temporary table deletion, which never occurred if change tracking was enabled, resulting **freezing**.



- Pest-Buda, 1847



Petőfi Sándor

## Szabadság, szerelem

Szabadság, szerelem!  
E kettő kell nekem.  
Szerelmemért föláldozom  
Az életet,  
Szabadságért föláldozom  
Szerelmemet.



Sándor Petőfi

## Liberty and Love

Liberty and love!  
These two I must have.  
For my love I'll sacrifice  
My life  
For liberty I'll sacrifice  
My love







## Budapest, 1882

A young man, who was equally proud of his Serb origin and his Croat homeland, recites a poem in the city park walking with his friend and trainer

“Sie rückt und weich, der Tag ist überlebt,  
Dort eilt sie hin und fördert neues Leben,  
Oh, das kein Flügel mich vom Boden hebt  
Ihr nach und immer nach zu streben!

Ach, zu des Geistes Flügeln wird so leicht  
Kein körperlicher Flügel sich gesellen!”



“The glow retreats, done is the day of toil;  
It yonder hastes, new fields of life exploring;  
Ah, that no wing can lift me from the soil,  
Upon its track to follow, follow soaring!

Alas! the wings that lift the mind no aid  
Of wings to lift the body can bequeath me.”  
(excerpt from Goethe's *Faust*)



„As I spoke the last words, plunged in thought and marveling at the power of the poet, the idea came like a lightning flash. In an instant I saw it all, and I drew with a stick on the sand the diagrams which were illustrated in my fundamental patents of May, 1888 [*A New System of Alternate-Current Motors and Transformers*], and which Szigety understood perfectly.”

(Nikola Tesla: *Some personal recollections*, Scientific American, 1915)



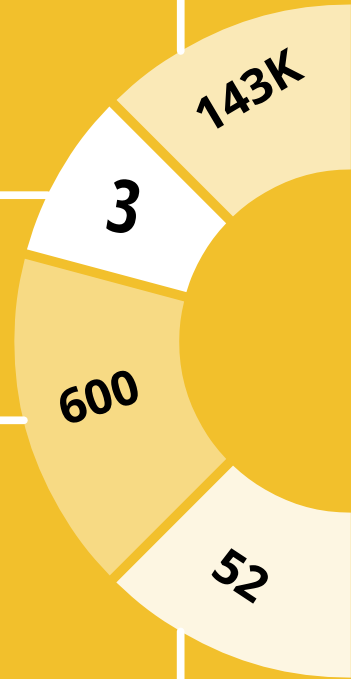
# LibreOffice Technology in numbers

Installations of Numbertext extension of Google Sheets

Popular operating systems shipped with Hunspell  
(Linux, macOS, Windows)

Italian students studying LibreLogo per year  
(University of Florence and University of Suor Orsola  
Benincasa in Naples by Andreas Formiconi)

Numbertext language modules written in Soros





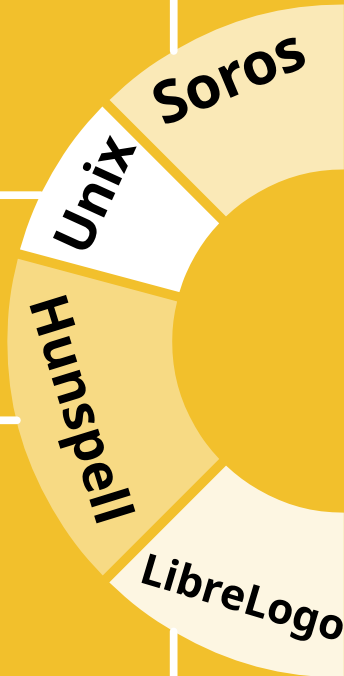
# More information

Soros tutorial and IDE on  
<http://www.numbertext.org>

Kernighan & Pike  
The Unix programming environment

man 5 hunspell  
(<https://github.com/hunspell/hunspell/tree/master/man>)

LibreLogo in LibreOffice Help  
(<https://help.libreoffice.org/latest/en-US/text/swriter/librelogo/LibreLogo.html>)





# THANK YOU

Yellow Idea Template for LibreOffice  
Impress.

Credit by : @ealita.id, 2020.