

Procés de traducció d'un programa informàtic anomenat OpenOffice

Jesús Corrius i Llavina
jesus@softcatala.org

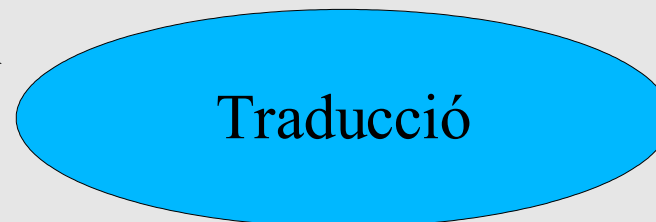


"Baixem, doncs, i confonguem la seva llengua perquè ningú no entengui la parla del seu germà"
(Gen 11:7)

Terminologia

Internacionalització (I18N): Modificació d'un programa informàtic de manera que permeti que aquest pugui gestionar múltiples llenguatges de manera eficaç.

Localització (L10N): Conversió d'un producte (programa, pàgina web, ajuda en línia, documentació) d'una llengua a una altra, tot fent que el producte sigui lingüísticament i culturalment apropiat pels parlants de la llengua de destinació.

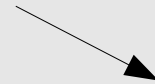
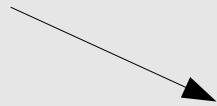


Fases generals del procés

Fase 1: Extracció

Fase 2: Traducció

Fase 3: Integració



Fase 1: Extracció

- **Familiarització amb el producte**
- **Anàlisi de les cadenes originals**
- **Creació d'un glossari**
- **Extracció de les cadenes**
- **Conversió de les cadenes en un format còmode de treballar pels traductors**

Familiartizació amb el producte

- No podem traduir correctament un programa que no coneixem



"Si volem entendre el món, com a mínim n'hem d'estimar una petita part"
(Henrik Ibsen)

Anàlisi de les cadenes originals

Format intern de les cadenes

- Recursos multiplataforma (Windows, GNU/Linux, Solaris)
- Estructures de dades en el codi font en 3 tipus de fitxer diferents.

```
FixedText FT_FORMULA_RESULT
```

```
{  
    Text = "Ergebnis" ;  
    Text [ ENGLISH ] = "Result" ;  
    Text [ italian ] = "Risultato" ;  
}
```

```
StringList [ english_us ] =
```

```
{  
    < "Last Used" ; Default ; > ;  
    < "All" ; Default ; > ;  
    < "Database" ; Default ; > ;  
}
```

"Totes les famílies felices s'assemblen, totes les famílies desgraciades ho són a la seva manera"
(Lev Tolstoi)

Anàlisi de les cadenes originals

Codificació:

- Format de text UNIX

Identificar possibles problemes en la traducció abans que es produixin:

- Llargada de les caixes de diàleg
- Cadenes que no es poden traduir
- Caràcters especials (en Win98 no es veu correctament el punt de la ela gemminada)

"Totes les famílies felices s'assemblen, totes les famílies desgraciades ho són a la seva manera"
(Lev Tolstoi)

Creació d'un glossari

Per què és important crear un glossari?

Perquè és una eina de consulta imprescindible, el desconeixement de la qual impedeix poder ser traductor del producte.

Perquè l'anàlisi terminològic que hem fet el puguin aprofitar altres persones o entitats com a guia.

Perquè ens pot servir per fer la pretraducció de les cadenes

Per coherència en les traduccions. Per exemple com s'ha de traduir la paraula "File" en la versió catalana d'OpenOffice?

Creació d'un glossari

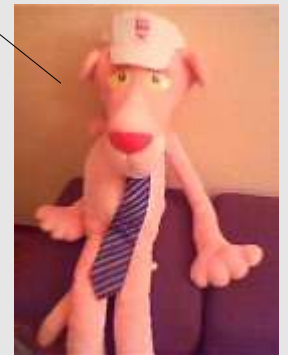
El glossari de l'OpenOffice en català

Conté gairebé 7.000 paraules anglès-català sobre ofimàtica

Es pot consultar des de:

<http://www.softcatala.org/projectes/openoffice/glossari.htm>

Necessita un repàs per actualitzar-lo a les últimes versions i algú que en faci un buidat terminològic. Algú s'hi anima?



Extracció de les cadenes

Format original en el codi font del programa
OpenOffice

```
graph TD; A([Format original en el codi font del programa OpenOffice]) --> B[Format intermig d'enginyeria de producte]; B --> C[Format de traducció (fitxers PO)];
```

Format intermig d'enginyeria de producte

Format de traducció (fitxers PO)

Extracció de les cadenes

Format intermig d'enginyeria?

Necessitat de tenir un format còmode per corregir petits errors, incoherències, etc.

Mantenir les cadenes agrupades.

Característiques

S'ha de poder importar/exportar de manera molt fàcil al producte amb un risc del 0% (o molt petit)

Conté informació tècnica que pot “molestar” els traductors, però imprescindible per l'enginyeria de producte

Format molt estricte.

Extracció de les cadenes

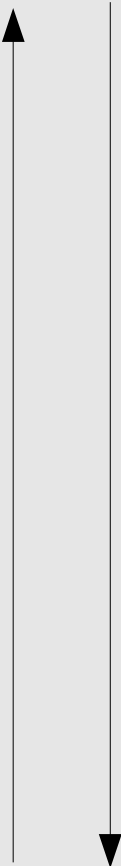
Format intermig

basctl source\basicide\basidesh.src 20020426 20:25:25	0	string	RID_STR_FILTER_ALLFILES	0	37	<Tot>
basctl source\basicide\basidesh.src 20020426 20:25:25	0	string	RID_STR_NOMODULE	0	37	<Cap mòdul>
basctl source\basicide\basidesh.src incorrecta.	0	string	RID_STR_WRONGPASSWORD	0	37	Contrasenya
basctl source\basicide\basidesh.src 20020426 20:25:25	0	string	RID_STR_OPEN	0	37	Carrega
basctl source\basicide\basidesh.src 20020426 20:25:25	0	string	RID_STR_SAVE	0	37	Desa
basctl source\basicide\basidesh.src gran i no es pot compilar ni desar.\n Suprimiu alguns	0	string	RID_STR_SOURCETOBIG	0	37	El text font és massa
basctl source\basicide\basidesh.src obrir el fitxer	0	string	RID_STR_ERROROPENSTORAGE	0	37	S'ha produït un error en
basctl source\basicide\basidesh.src carregar la biblioteca	0	string	RID_STR_ERROROPENLIB	0	37	S'ha produït un error en
basctl source\basicide\basidesh.src biblioteca del BASIC	0	string	RID_STR_NOLIBINSTORAGE	0	37	El fitxer no conté cap
basctl source\basicide\basidesh.src 20020426 20:25:25	0	string	RID_STR_BADSBXNAME	0	37	El nom no és vàlid
basctl source\basicide\basidesh.src pot tenir fins a 30 caràcters.	0	string	RID_STR_LIBNAMETOLONG	0	37	Un nom de biblioteca
basctl source\basicide\basidesh.src d'altres documents no són accessibles.	0	string	RID_STR_ERRORCHOOSEMACRO	0	37	Les macros
basctl source\basicide\basidesh.src només lectura.	0	string	RID_STR_LIBISREADONLY	0	37	Aquesta llibreria és de
basctl source\basicide\basidesh.src 20020426 20:25:25	0	string	RID_STR_REPLACELIB	0	37	No es pot reemplaçar 'XX'.

Format de treball dels traductors

Extracció de les cadenes

Format intermig

- 
- A vertical line with an upward-pointing arrow at the top and a downward-pointing arrow at the bottom, indicating a bidirectional relationship or flow between the 'Format de treball dels traductors' and the 'Format intermig'.
- Convertir la codificació de caràcters
 - Indexació de les cadenes
 - Divisió de les cadenes en fitxers petits
 - Formatar les cadenes
 - Pretraducció amb glossari
 - Pretraducció amb memòries de traducció
 - Afegir-hi suggerències, etc.

Format de treball dels traductors

Extracció de les cadenes

Format intermig

Format de treball dels traductors

source\basicide\basidesh.src, RID_STR_OPEN, string

Suggerència de traducció: Carrega

“Load”

Carrega

source\basicide\basidesh.src, RID_STR_RUNTIMEERROR, string

“Runtime error”

Error en temps d'execució

source\ui\wrtsh\wrtsh.src, STR_INSERT_UNDO, string

Suggerència de traducció: Número

“Number”

Número

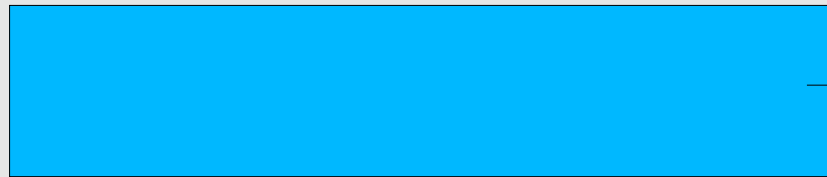
Fase 2: Traducció

- **Pretraducció**
- **Recuperació de les cadenes**
- **El format de fitxer PO**
- **El nostre format**
- **Eines de traducció**

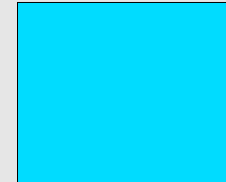
Pretraducció

- El glossari que hem generat serveix per traduir parts del programa
- Normalment hi ha moltes cadenes repetides.
- Realimentació de traduccions
- *Utilitzar altres glossaris per fer pretraduccions.*

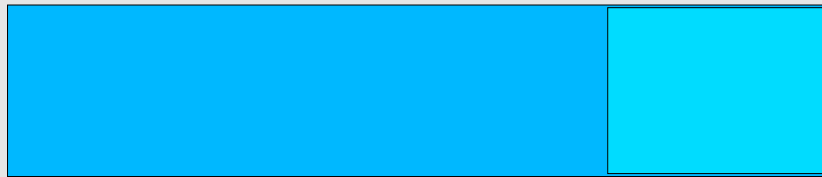
Recuperació de les cadenes



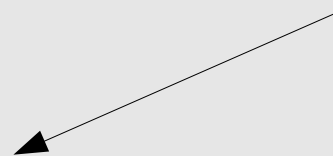
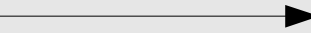
Versió actual del programa



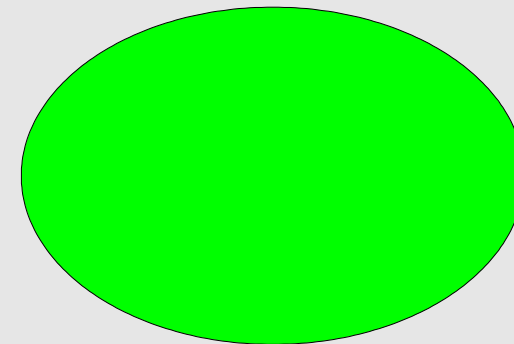
Traduccions



Versió traduïda



?



Noves versions del programa

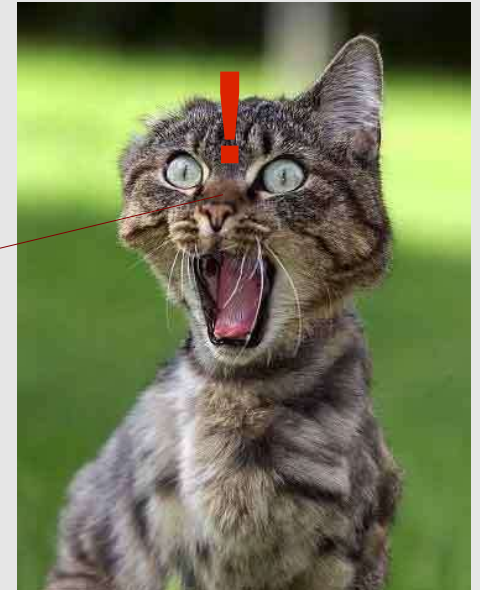
Recuperació de les cadenes

Recuperació de les cadenes del programa

- Cada cadena té el seu identificador
- La majoria de programes proporcionen eines pròpies o estàndard

Recuperació de les cadenes d'ajuda

- La majoria de programes no tenen en compte l'actualització de l'ajuda
- Hem d'indexar el seu contingut per paràgrafs o frases



El format de fitxer PO

Què és un fitxer PO?

- PO - Portable Executable / MO – Machine Object
- POT - Fitxers base
- Un fitxer PO per a cada idioma
- GNU Gettext ofereix diverses utilitats per manipular les cadenes

El format de fitxer PO

Format genèric

espai-en-blanc

comentaris-del-traductor

#. comentaris-automàtics

#: referència

#, indicador...

msgid cadena-original

msgstr cadena traduïda

Revisat. Jc 12-03-2003

#. Glossari OpenOffice

#: SC_09743;String

#, fuzzy, no-c-format

msgid "Size"

msgstr "Mida"

El format de fitxer PO

Particularitats del format

Traduccions amb formes plurals

msgid cadena-original-singular
msgid_plural cadena-original-plural
msgstr[0] cadena-traduïda-cas-0
...
msgstr[N] cadena-traduïda-cas-n

Traduccions de més d'una línia

msgid “
“Achilles' wrath, to Greece the direful spring\n”
“Of woes unnumber'd, heav'nly Goddess, sing!\n”
msgstr “
“Canta, oh dea, la còlera d'Aquil·les, fill de Peleu\n”
“que ocasionà infinits mals als aqueus, etc.\n”

El format de fitxer PO

Indicadors

- Indicador *fuzzy*:

El pot generar automàticament les eines o un traductor. Indica que la cadena msgstr pot no ser una traducció correcta. Només una persona pot decidir si és correcte o no. Una vegada decidit s'ha d'eliminar manualment.

- Indicador *no-c-format*:
- Indicador *c-format*:

Són generats automàticament per les eines de traducció per indicar el format de traducció de les cadenes. Amb el c-format les eines han de fer comprovació automàtica de la correcció del format de les cadenes.

El format de fitxer PO

Caràcters d'escapament (c-format)

\\	barra invertida	\\n	nova línia
\\a	alerta	\\r	principi de línia
\\b	enrera	\\t	tabulació horitzontal
\\f	pàgina lògica següent	\\v	tabulació vertical

“Sóc home; \\ni penso\\tque res d'humà\\n no m'és aliè”

↓

Sóc home;
i penso que res
d'humà
no m'és aliè

El format de fitxer PO

Variables de substitució (c-format)

%%	tant per cent	%c	un sol caràcter
%o	valor en octal	%s	una cadena de text
%x	valor en hexadecimal	%f	valor en coma flotant
%d	valor en decimal	%g	coma flotant de notació exp.

“S'ha trobat %d fitxers”

“El meu nom és: %s”

“Ha escollit la lletra %c”

“S'ha tobat 8 fitxers”

“El meu nom és: Jesús”

“Ha escollit la lletra a”

“You've selected component:\t%s%f %dtimes.\n”

“Heu seleccionat %d vegades el component:\t%s %f.\n”

Heu seleccionat 10 vegades el component: OpenOffice 1.0.

El nostre format

- Implementació molt semblant al format PO, però sense indicadors de control (**Error**)
- Indicació en la referència del tipus de recurs: menú, pantalla d'avís, opcions de menú desplegable, botó, etc.
- Recomanació de traduccions segons el glossari. Sense traducció automàtica.
- Divisió dels fitxers en unes mides raonables i còmodes de treballar

El nostre format

Traducció de l'OpenOffice avui

- Es poden extreure totes les cadenes amb una eina
- Hi ha scripts per convertir les cadenes en format PO
- Existeix un framework especial per fer builds d'OpenOffice en diferents llengües fàcilment
- Hi ha eines per traduir l'ajuda de manera més o menys automatitzada.

Eines de traducció

- GNU Emacs: <http://www.gnu.org/software/emacs/emacs.html>
- KBabel: <http://i18n.kde.org/tools/kbabel/>
- Gtranslator: <http://www.gtranslator.org/>
- PoEdit: <http://poedit.sourceforge.net/>
- Microsoft Word:
<http://www.microsoft.com/office/word/default.asp>

Eines de traducció

GNU Emacs: <http://www.gnu.org/software/emacs/emacs.html>

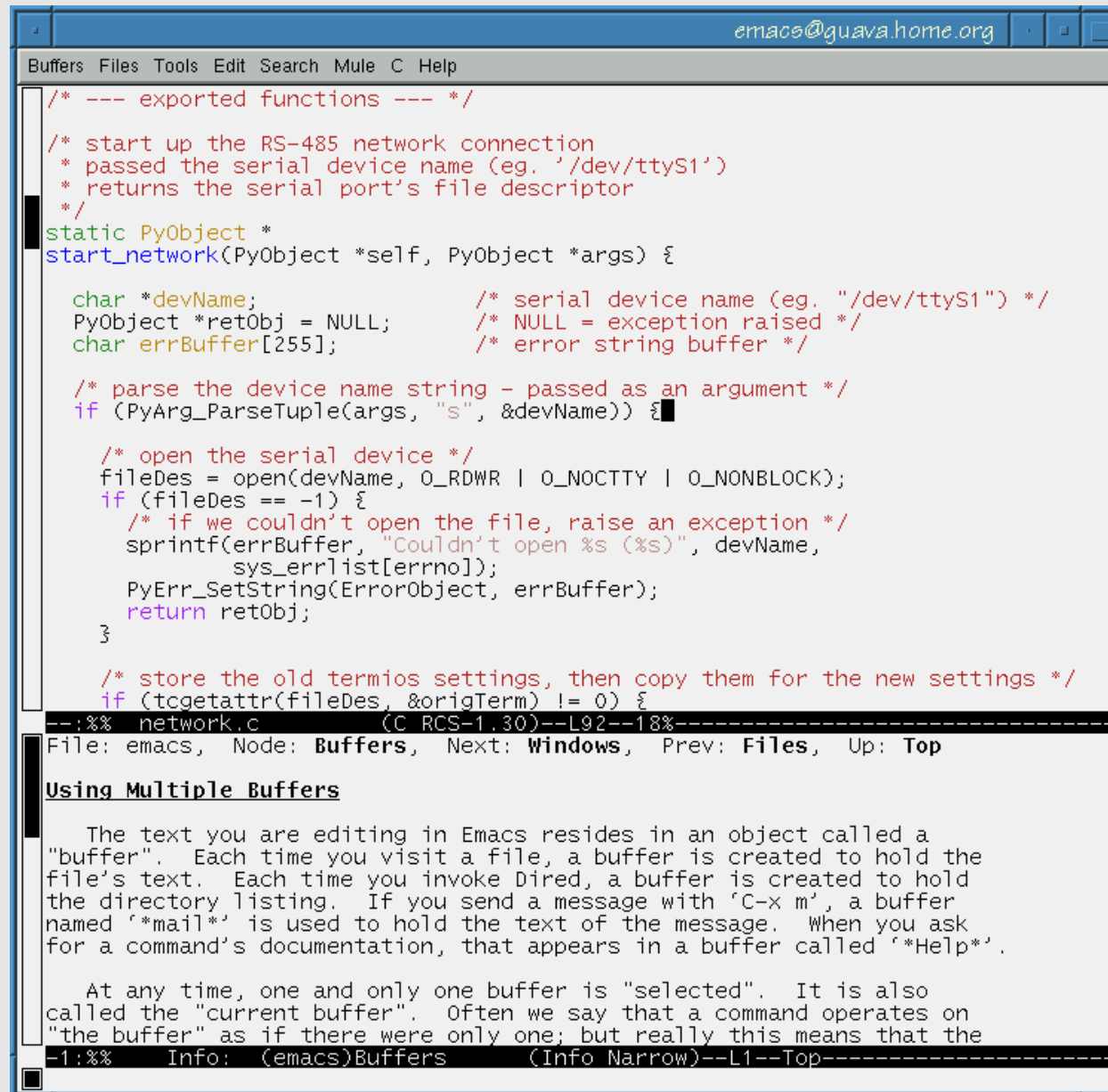
Pros:

- Possiblement és el millor editor de textos que existeix (i ja té més de 30 anys)
- És altament configurable
- Té un mòdul especial per treballar amb fitxers PO

Contres:

- És un programa molt difícil d'utilitzar pels que no són programadors. Necessites un curs per aprendre'n.
- No té una interfície d'usuari moderna.
- És molt gran, més de 70 Mb.

Eines de traducció



```
emacs@guava.home.org
Buffers Files Tools Edit Search Mule C Help
/* --- exported functions --- */
/* start up the RS-485 network connection
 * passed the serial device name (eg. '/dev/ttyS1')
 * returns the serial port's file descriptor
 */
static PyObject *
start_network(PyObject *self, PyObject *args) {
    char *devName;          /* serial device name (eg. "/dev/ttyS1") */
    PyObject *retObj = NULL; /* NULL = exception raised */
    char errBuffer[255];    /* error string buffer */

    /* parse the device name string - passed as an argument */
    if (PyArg_ParseTuple(args, "s", &devName)) {

        /* open the serial device */
        fileDes = open(devName, O_RDWR | O_NOCTTY | O_NONBLOCK);
        if (fileDes == -1) {
            /* if we couldn't open the file, raise an exception */
            sprintf(errBuffer, "Couldn't open %s (%s)", devName,
                    sys_errlist[errno]);
            PyErr_SetString(ErrorObject, errBuffer);
            return retObj;
        }

        /* store the old termios settings, then copy them for the new settings */
        if (tcgetattr(fileDes, &origTerm) != 0) {
            --:%% network.c (C RCS-1.30)--L92--18%-----
        }
    }
}
File: emacs, Node: Buffers, Next: Windows, Prev: Files, Up: Top
Using Multiple Buffers
The text you are editing in Emacs resides in an object called a
"buffer". Each time you visit a file, a buffer is created to hold the
file's text. Each time you invoke Dired, a buffer is created to hold
the directory listing. If you send a message with 'C-x m', a buffer
named '*mail*' is used to hold the text of the message. When you ask
for a command's documentation, that appears in a buffer called '*Help*'.

At any time, one and only one buffer is "selected". It is also
called the "current buffer". Often we say that a command operates on
"the buffer" as if there were only one; but really this means that the
-1:%% Info: (emacs)Buffers (Info Narrow)--L1--Top-----
```

Eines de traducció

- KBabel: <http://i18n.kde.org/tools/kbabel/>

Pros:

- Permet mantenir catàlegs de projectes
- Correcció ortogràfica automàtica.
- És molt fàcil d'utilitzar i molt còmode.

Contres:

- Només funciona sota la plataforma GNU/Linux
- Està especialment dissenyat per treballar en l'entorn KDE.

Eines de traducció

The screenshot shows the KBabel application window. The title bar reads "file:/usr/local/src/kde/kde-i18n/de/messages/kdesdk/kbabel.po [modified] - KBabel". The menu bar includes "File", "Edit", "Go", "Dictionaries", "Tools", "Settings", and "Help". The toolbar contains various icons for file operations and translation. The main window is divided into several sections:

- Original String (msgid):** Contains the text: `Result of the translation:\n\nEdited entries: %1\nExact translations: %2 (%3)\nNon-exact translations: %4 (%5)\nNothing found: %6 (%7)`
- Translated String (msgstr):** Contains the text: `Resultat : \n\nBearbeitete Einträge: %1\nExakte Übersetzungen: %2 (%3)\nNicht-exakte Übersetzungen: %4 (%5)\nKeine Übersetzung: %6 (%7)`
- Comment:** Contains: `#: kbabel/roughtransdlg.cpp:537\n#, fuzzy`
- Summary:** Shows "Total: 1" and "Found in: kdcop.po". "Current: 1" and "Translator: Thomas Diehl <thd@kde.org>" are also visible. A "Date:" field and a "More" button are present.
- Table:** A table with columns "Score", "Original", "Translation", and "Location". The first row shows a score of 100, the original text "Result", the translation "Ergebnis", and the location "kdcop.po".
- Navigation:** Includes "< Previous" and "Next >" buttons.
- Search:** Includes "Search" and "Context" buttons.
- Status Bar:** Shows "Current: 481", "Total: 565", "Fuzzy: 1", "Untranslated: 0", "INS", "RW", "Line: 1", "Col: 10".

Eines de traducció

- Gtranslator: <http://www.gtranslator.org/>

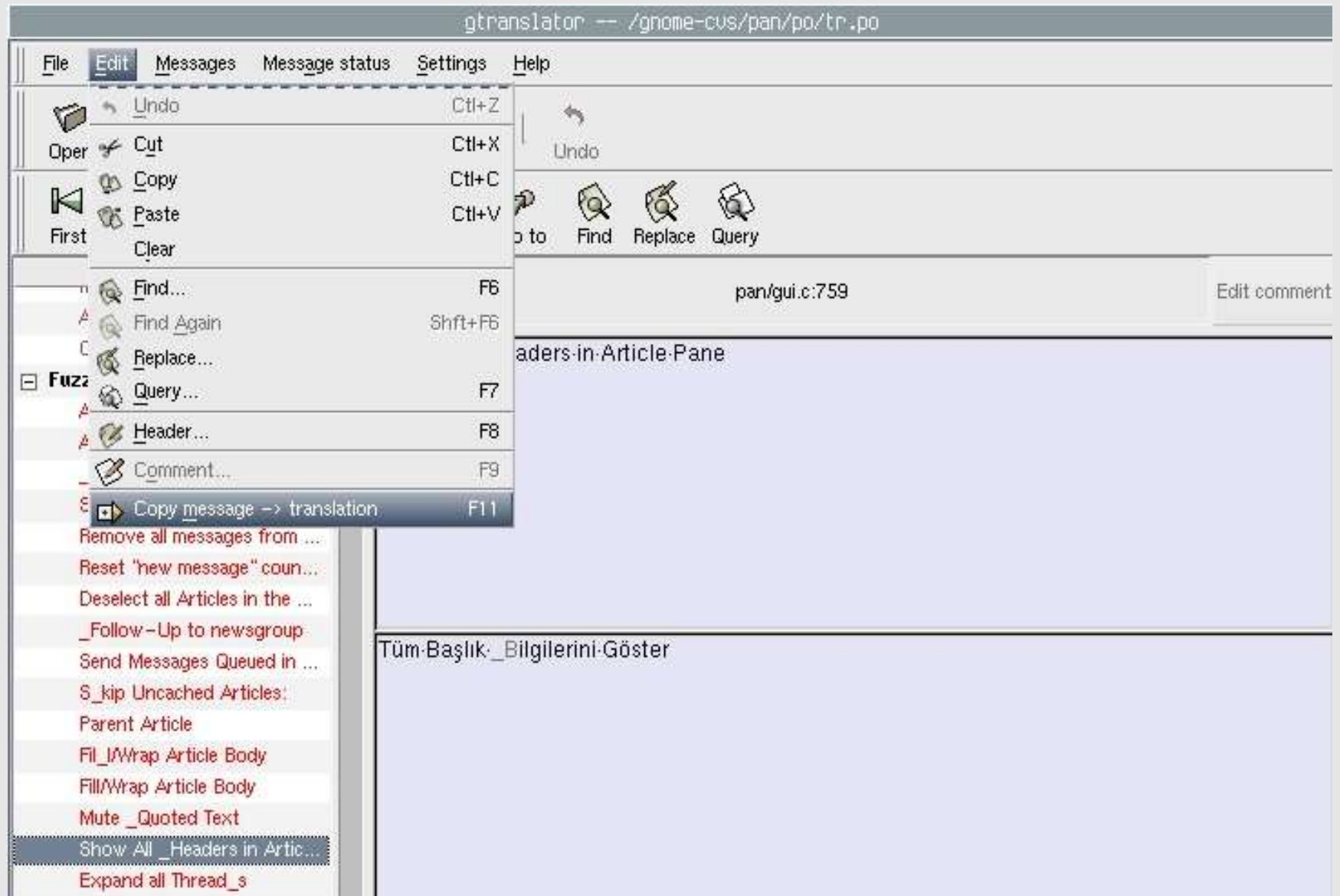
Pros:

- És molt fàcil d'utilitzar

Contres:

- Només funciona sobre GNU/Linux.
- El projecte està bastant abandonat.
- Està especialment preparat per treballar en l'entorn GNOME.

Eines de traducció



Eines de traducció

- PoEdit: <http://poedit.sourceforge.net/>

Pros:

- Té les mateixes característiques que el kBabel o gtranslator.
- Funciona sobre Windows.

Contres:

- És bastant inestable sobre Windows.
- El disseny de l'aplicació no és molt modern.

Eines de traducció

The screenshot shows the poEdit application window titled "poEdit : /home/vasek/softlabs/poedit/locales/cs.po (modified)". The menu bar includes "File", "Edit", "Catalog", "View", and "Help". Below the menu bar is a toolbar with icons for file operations and help. The main area is a table with two columns: "Original string" and "Translation".

Original string	Translation
(%i new, %i obsolete)	(%i nových, %i neplatných)
Original string	Originální řetězec
Translation	Překlad
● poEdit	poEdit
● &File	&Soubor
● (none of these)	(žádný)
● Select catalog's language	Vyberte jazyk katalogu
Please select language code:	Vyberte prosím kód jazyka:
Catalog modified. Do you want to save changes?	Katalog byl modifikován. Chcete uložit změny?
● Save changes	Uložit změny
Open catalog	
✎ File	
doesn't exist!	
Save as...	
Automatically translating...	
Automatically translated %u strings	
No references to this string found.	
Please choose the reference you want to show:	
References	
Error loading message catalog file "	

An "Edit comment" dialog box is open over the table. It has a title bar "Edit comment" and a close button. The dialog contains a text area with the text: "Any entry in po catalog may have a comment attached to it..". At the bottom of the dialog are three buttons: "OK", "Cancel", and "Clear".

At the bottom of the poEdit window, there are two text boxes: the first contains "File " and the second contains "Soubor ". The status bar at the very bottom shows "66 strings (4 fuzzy, 0 not translated)".

Eines de traducció

Microsoft Word: <http://www.microsoft.com/office/word/default.asp>

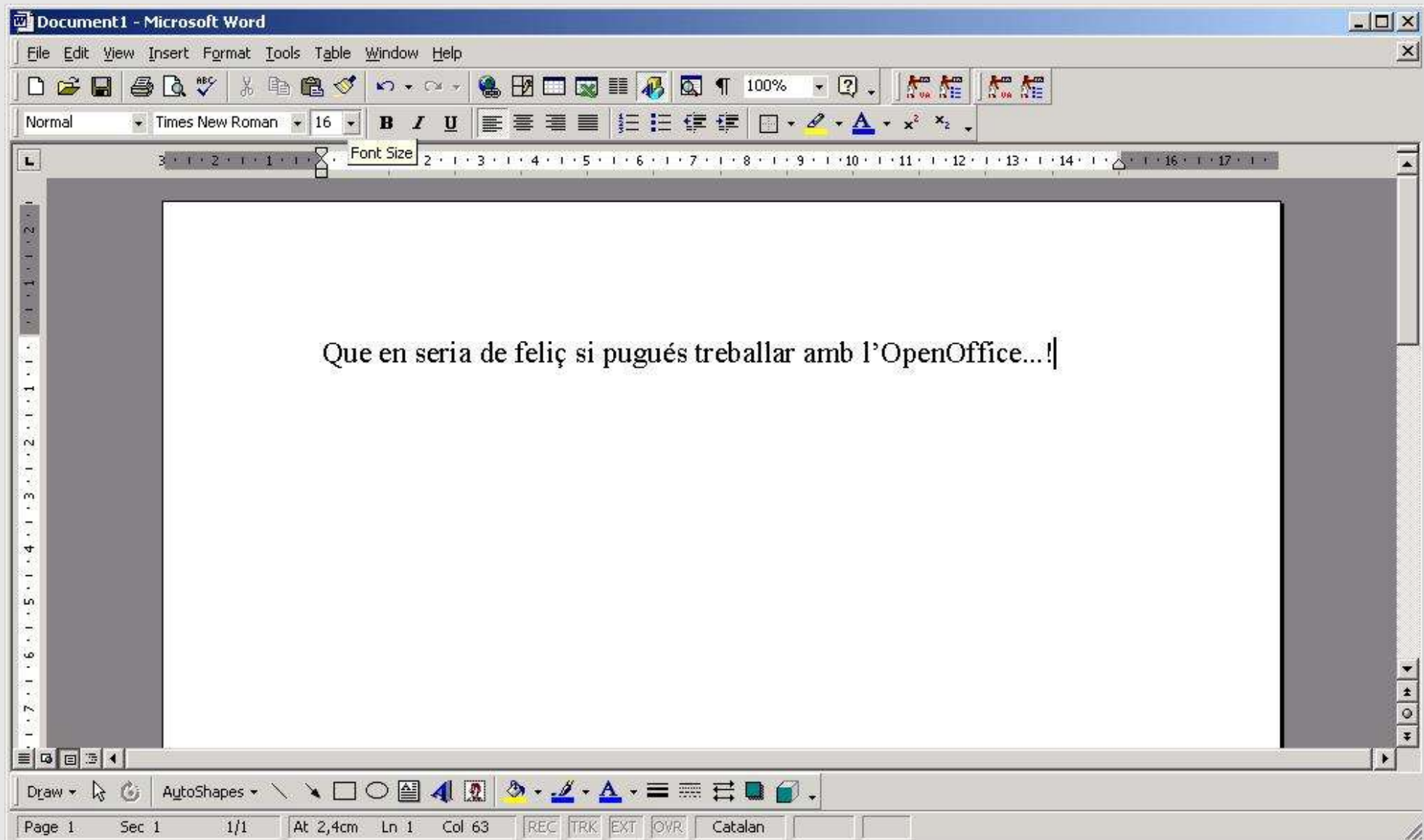
Pros:

- La majoria de traductors es troben còmodes treballant amb aquest producte

Contres:

- No té el més mínim respecte pels formats, codificació dels caràcters
- No és gratuït. Val 150-250 Euros
- El format DOC és secret

Eines de traducció



Fase 3: Integració

- Correcció ortogràfica i gramatical de les cadenes
- Adaptació terminològica del producte
- Adaptació del producte a la nostra realitat cultural
- Integració de components
- El procés de proves
- Producte final

Correcció ortogràfica i gramatical

La figura del corrector

- No ha de tenir coneixements tècnics
- S'encarrega de validar les cadenes a la recerca d'errors ortogràfics, gramaticals, barbarismes, etc.
- Comprova que el vocabulari utilitzat sigui el propi del glossari del propi producte o del recull de termes de Softcatalà.
- Vigila que s'hagi utilitzat les normes i guies de traducció de la guia d'estil de Softcatalà.

Adaptació terminològica

Requisits

- Coneixement elevat del producte
- Coneixements tècnics a nivell de programador
- Familiaritat amb l'estructura del codi font del programa
- Capacitat de prendre decisions sobre terminologia

Per regla general és el cap del projecte

Adaptació a la realitat cultural

- Adaptar els formats del calendari, hora, moneda, etc.
- Traducció de les plantilles
- Traducció/adaptació de les fórmules matemàtiques de la fulla de càlcul i les macros
- Permetre l'ús de la coma com a separador decimal
- Permetre l'ús de la coma geminada en la interfície gràfica
- Diferents adaptacions del corrector ortogràfic, separador de síl·labes, etc.

Integració de components

Creació d'un paquet d'OpenOffice adaptat als usuaris catalano-parlants:

- Plantilles
- Correctors ortogràfics: català, castellà, anglès i francès
- Divisor sil·làbic
- Eines auxiliars

Procés de proves

Versions alpha

Versions internes de desenvolupament

Versions beta

Versions en traducció. Alliberades a la llista

Release
Candidates



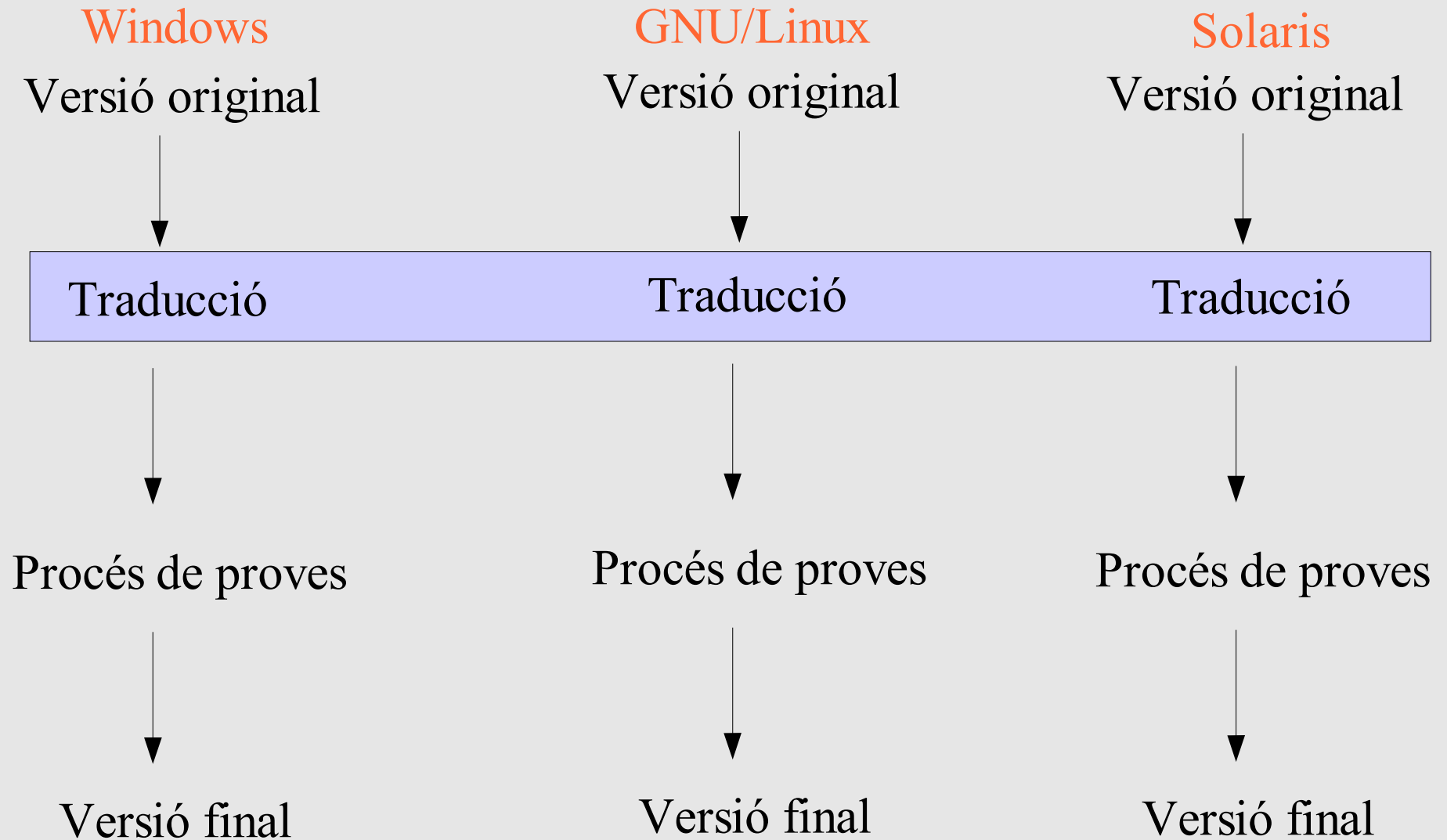
Traducció completa



Hi ha un error!

Versió final

Procés de proves



Producte final

The screenshot shows the OpenOffice.org 1.0.1 interface. The main window displays a document titled "Sense nom1". The text in the document is partially obscured by two dialog boxes. The "Correcció ortogràfica (Català)" dialog is open, showing the word "d'informatica" with suggestions: "d'informàtica", "d'informativa", and "d'informat ica". The "Estils de paràgraf" dialog is also open, showing a list of styles including "Cos del text", "Encapçalament 1" through "10", "Per defecte", "Sagnat de la llista", "Sagnat del cos del text", "Sagnat de primera línia", "Sagnat negatiu", "Signatura", "Tancament complementari", and "Estils personalitzats".

El departament d'informatica de la Universitat de Nova York es troba dins el Warren Weaver Hall, un edifici que sembla una fortalesa dos carrers a l'est del Washington Square Park. Els ventiladors de sortida de l'aire condicionat de potència industrial creen un fossat d'aire calent que allunya els vagons i els trens. És un fenomen formidable: la guàrdia dels trens de Nova York és una feina molt perillosa. Els trens de Nova York són molt antics i els conductors són molt veterans. La música és molt bona i els guàrdies són molt bonics. Diria que els procediments de seguretat de l'edifici. La persona és en Richard M. Stallman, fundador del Projecte GNU, primer president de la Free Software Foundation (Fundació pel programari lliure), guanyador del la beca MacArthur al 1990, guanyador del premi Grace Murray Hopper de l'Association of Computing Machinery (Associació de maquinari informàtic), coguanyador del premi Takeda de la Fundació Takeda al 2001, i antic hacker de l'AI Lab. Tal com s'anuncia a un munt de llocs d'internet relacionats amb els hackers, i també en el del Projecte GNU <http://www.gnu.org>, Stallman està a Manhattan, la seva antiga ciutat, per fer un molt esperat discurs en resposta a la recent campanya de

GRÀCIES

Vicky Alins, Marc Belzunces, Judith Bosch,
Òscar Del Pozo, Miquel Escarrà, Vicent Fornés,
Adolf Fuertes, Pere Gualler, Ester Haba,
Adolf M. Iglesias, Amaia Iturraspe, Jordi Jover,
Quico Llach, Marcel López, Jordi Mas,
Josep Molés, Joan Moratinos, Jesús Moreno,
Míriam Pérez, Pere Pasqual Pérez, Toni Rosado,
Begonya Soler, Josep M. Suelves, Laura Vilanova.

A tot l'equip d'OpenOffice.org i a rms per haver escrit el GCC