USER MANUAL

GALLITO 2.0

64bits

CONTENTS

1. INTRODUCTION

The aim of this document is to explain how Gallito 2.0 works. Gallito is a tool suitable for two domains:

Research:

Makes it possible to process language samples and extract samples for psycholinguistics experiments, such as term entropy, frequency measured through vector length, similarity between terms, lists of semantic neighbors. All of this is very useful to exercise experimental control or even to do research into the effect of this kind of variables on processing.

Technology:

For text categorizers in any domain. In fact, Gallito 2.0 is currently being used as the first module in cloud applications that categorize customer service calls. It can also be used to visualize useful information for marketing departments.

The functionality associated with this application is:

Creation of semantic-vectorial spaces from texts

- Calculation of term significance functions (Entropy or IDF)
- Calculation of vector norm for every term
- Lists of semantic neighbors of a term
- Similarity between terms
- Similarity between existing documents
- Similarity between documents that do not exist in space (pseudodocuments)
- Batch processes (similarities, graphs, neighbors)
- Matrix output in plain text
- Pajek format output for term visualization
- Essay evaluation
- Text Cohesion (sentence-sentence, paragraph-paragraph, etc)
- Dimensions Interpretation

2.INSTALLATION

2.1.Requirements

In order for Gallito 2.0 to work properly the following components must be installed (they are included in the installation CD and can be downloaded from the Microsoft website):

- **Windows 64 bits operating system (Windows 7 or Windows Server)**
- **Microsoft SDK 4**, included in the installation CD
- **Microsoft Visual C++ 2010 Redistributable Package (64bits)**
- **Writing permissions to write in the installation directory**

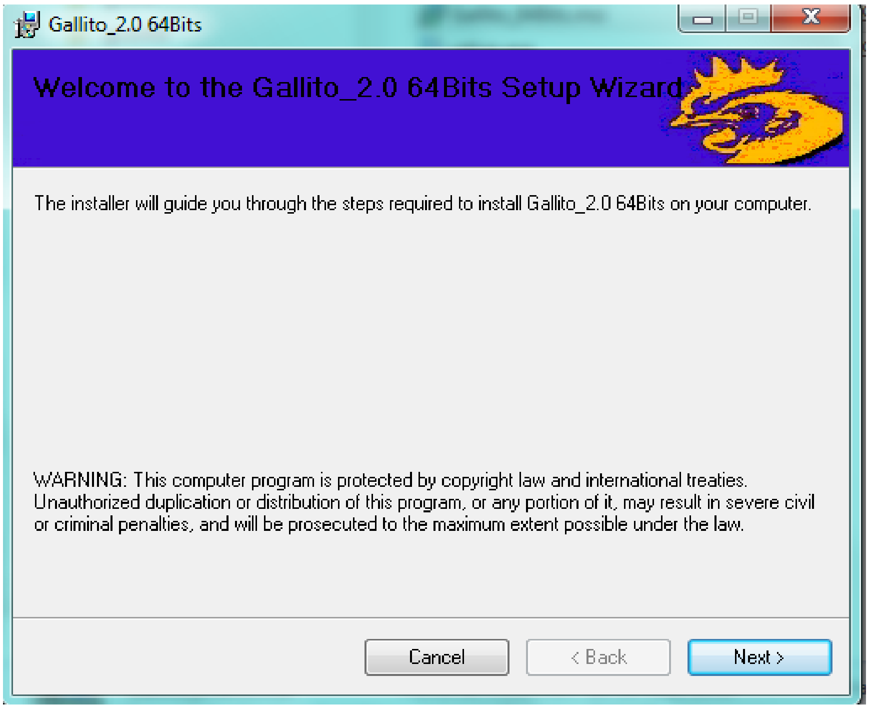
Some knowledge of Latent Semantic Analysis (LSA) and its applications is also required.

2.2. Installation procedure

- Access the installation folder and execute the Setup.exe program.
- The following screens will appear to guide you through the installation process:

2.2.1. Step one

Welcome screen to product installation, specifying the product to be installed (Gallito 2.0). Click NEXT to continue installing or CANCEL to exit the installation process.



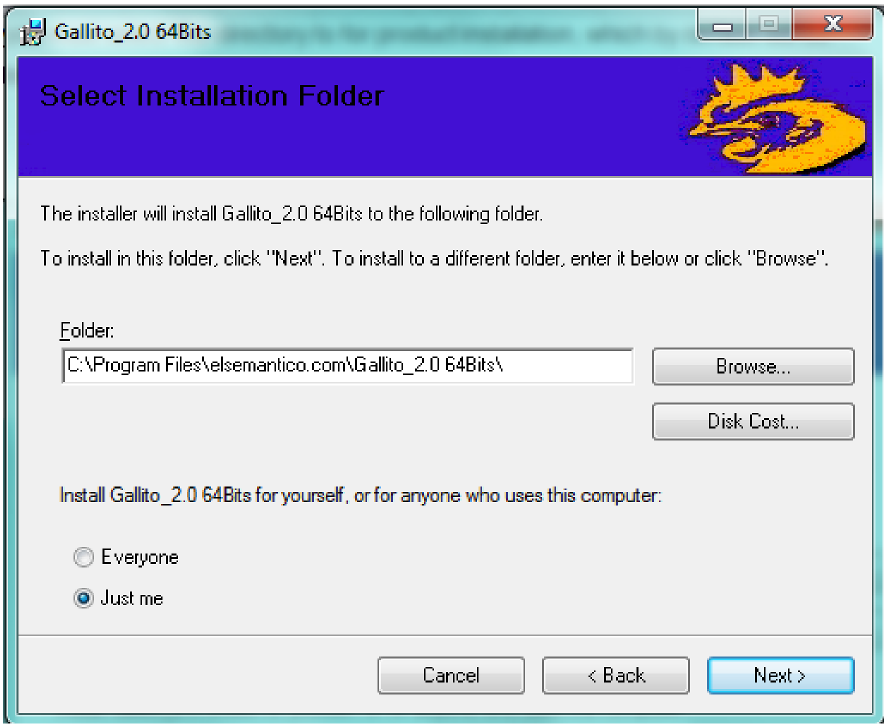
Welcome to the Gallito_2.0 64Bits Setup Wizard

The installer will guide you through the steps required to install Gallito_2.0 64Bits on your computer.

WARNING: This computer program is protected by copyright law and international treaties. Unauthorized duplication or distribution of this program, or any portion of it, may result in severe civil or criminal penalties, and will be prosecuted to the maximum extent possible under the law.

2.2.2. Step two

In this screen you must select the directory to for product installation, which by default will be C:\Program Files\elsemantico.com\Gallito_2.0 64Bits\



Select Installation Folder

The installer will install Gallito_2.0 64Bits to the following folder.

To install in this folder, click "Next". To install to a different folder, enter it below or click "Browse".

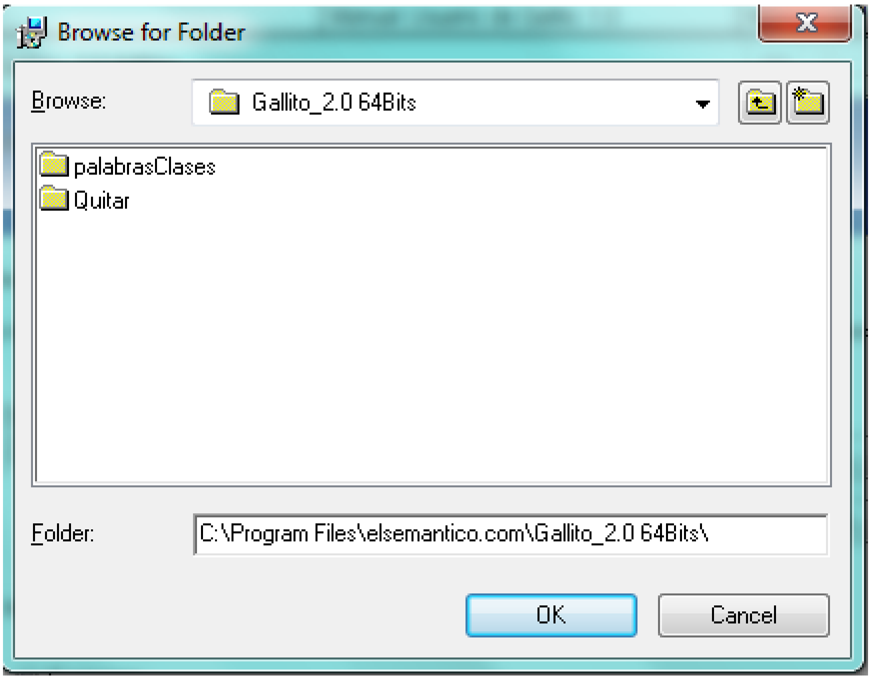
Folder: C:\Program Files\elsemantico.com\Gallito_2.0 64Bits\

Install Gallito_2.0 64Bits for yourself, or for anyone who uses this computer:

○ Everyone

● Just me

If you want to install the product in a directory other than the default directory, select it by clicking BROWSE in the installation window. Clicking this button you will access to a standard Windows directory browser, through which you will be able to select the directory where you want to install Gallito 2.0 (see following image).



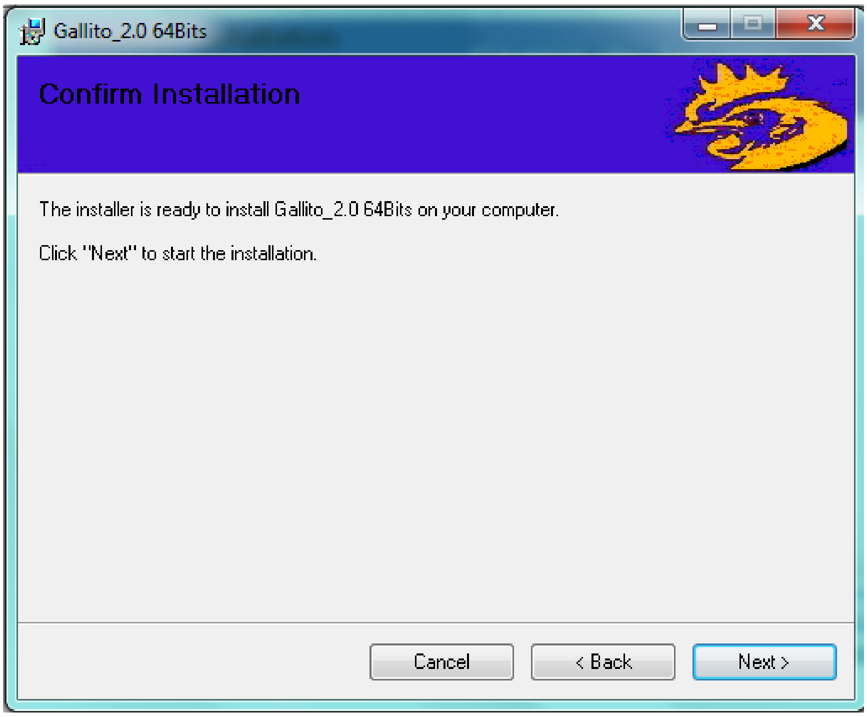
Browse: Gallito_2.0 64Bits

Folder: C:\Program Files\elsemantico.com\Gallito_2.0 64Bits\

Once you select the installation directory, click NEXT to continue the installation process.

2.2.3. Step three

In this step the user is warned that clicking NEXT will launch product installation. If you click BACK, you will be able to re-enter the installation information.



The installer is ready to install Gallito_2.0 64Bits on your computer.

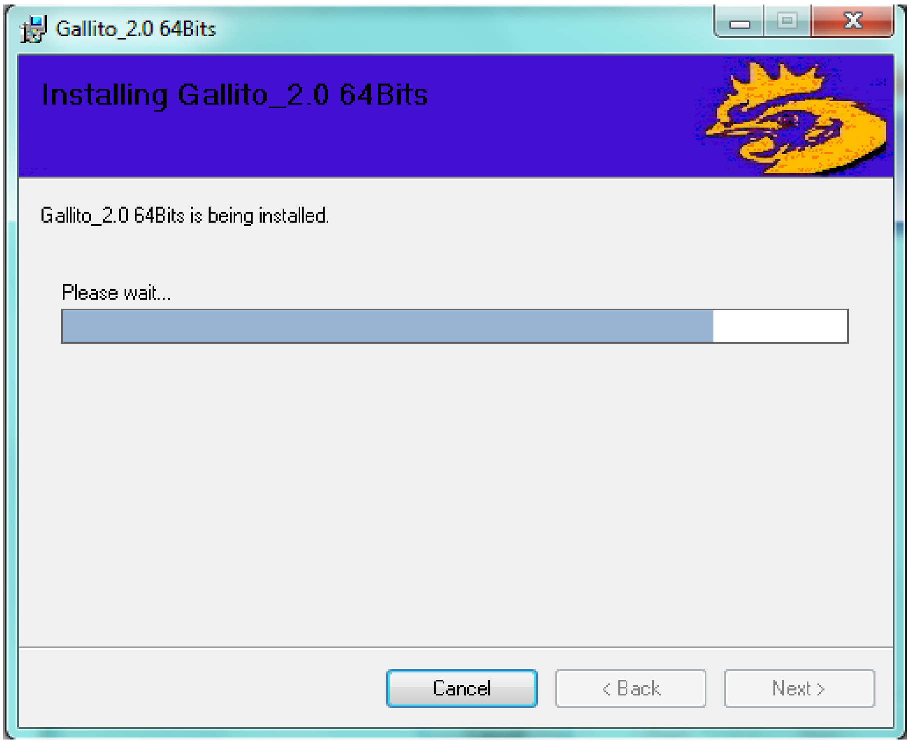
Click "Next" to start the installation.

2.2.4. Step four

The window shows installation progress as a percentage and the files being copied to the machine on which the product is being installed.

Once the installation process is 100% complete, you will access the next step (see next section) after being informed that the necessary information for the application to work correctly has been entered.

Clicking CANCEL in the progress window halts the installation process.



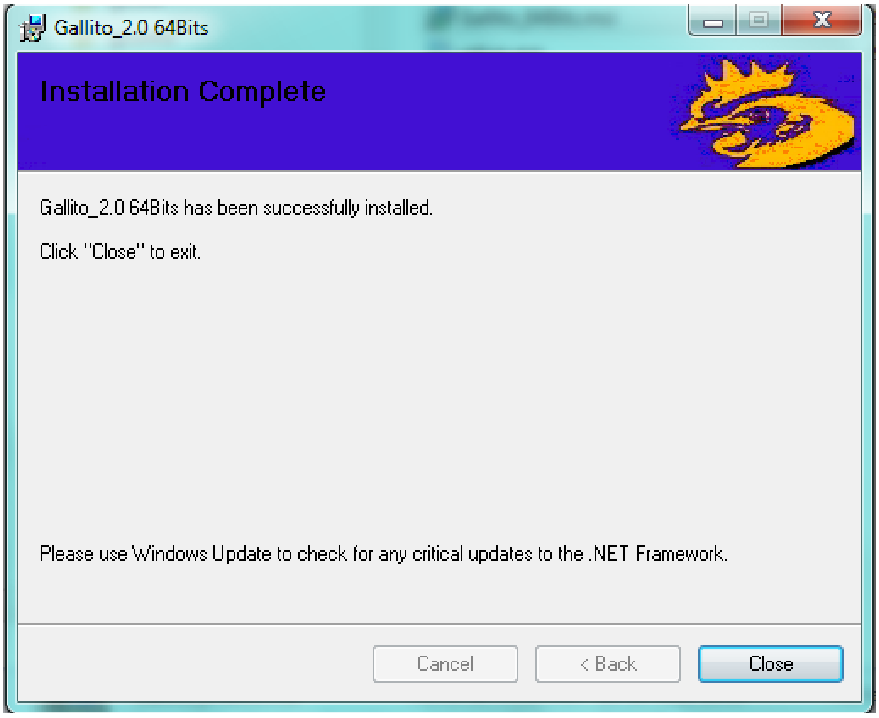
Installing Gallito_2.0 64Bits

Gallito_2.0 64Bits is being installed.

Please wait...

2.2.5. Step five

End of the installation. This screen lets the user know that product installation is complete. By clicking on FINISH you will access the last screen in the installation process.



Installation Complete

Gallito_2.0 64Bits has been successfully installed.

Click "Close" to exit.

Please use Windows Update to check for any critical updates to the .NET Framework.

3. SUMMARY OF POSSIBLE ACTIONS

The application makes it possible:

- To generate semantic spaces under various parameters.
- To find the similarity between terms, documents, and pseudodocuments.
- To generate lists of neighbors under various parameters (with the possibility of exporting them to Microsoft Excel).
- To load and save semantic spaces in different formats.
- To generate .txt files from significant matrices.
- To batch process (neighbors, similarity by pairs, similarity matrices)

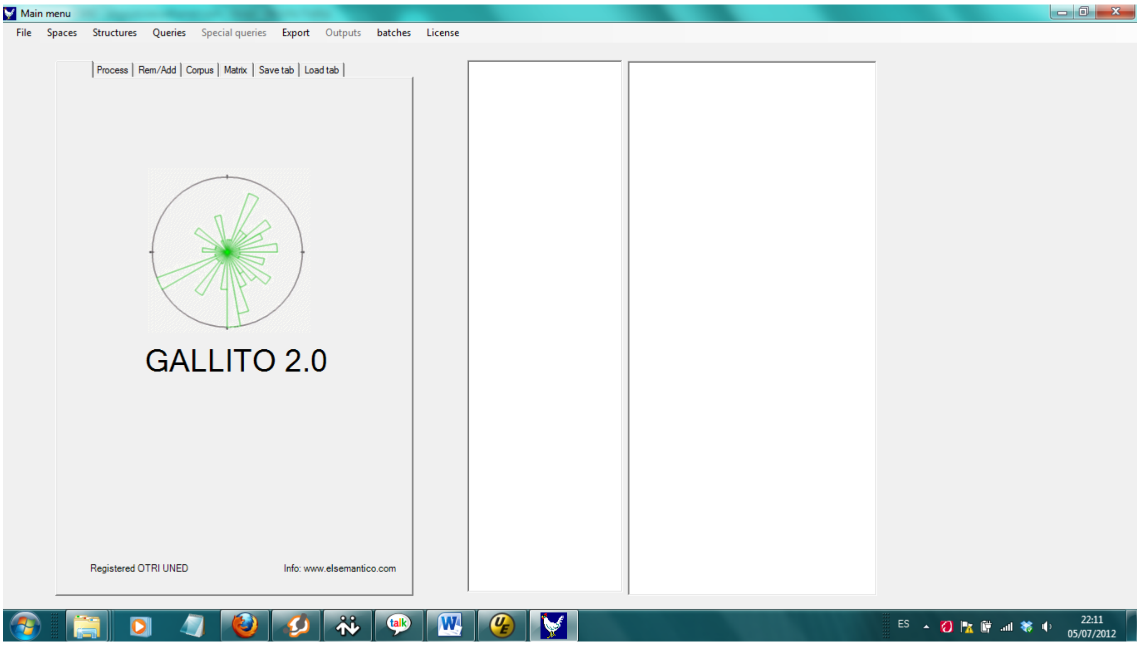
To do this, the application has a single form or control panel with tabs and drop lists. The application has two basic functionalities:

- Generation of a semantic space.
- Operations on a semantic space (Calculate Similarities, Save Space, Load Space, etc.)

4. APPLICATIONS

When the application is launched, a small presentation is deployed, followed by the control screen. From then on, you can create a new semantic space or load an existing space that has been saved in the hard disc.

The screen that appears is the following:



You can start to perform operations using the tabs and drop lists.

4.1 Creating a semantic space

To create a semantic space, a corpus is needed. This corpus will be called “reference corpus” and will be a text plain file (.txt extension). The contextual window can be separated by a character or simply by natural-language sentences. In the former case, the separating character (usually “**#**”) have to be between two documents:

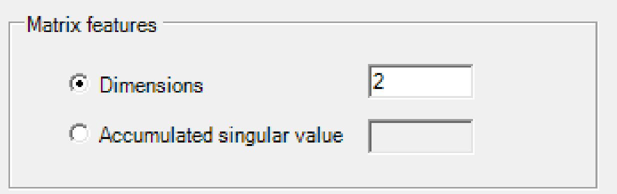
Los archivos planos constituyen la forma más básica de una base de datos
#
Los archivos planos incluyen un campo por cada uno de los elementos que se desean contemplar
#
La redundancia de elementos es una característica
de estos archivos
#
La base de datos relacional soluciona la redundancia en los datos
#
Son frutos largos y con sabor
#
La recogida será buena si ha tenido una buena base como semillero
#
Los frutos son de color verde
#
En la recogida es parecida a los demás frutos largos
#

In the second case, a well-established sentence dot separation is mandatory but a special format is not needed.

To create a semantic space, the following parameters are required:

- **No. of dimensions or Accumulated singular value:** The number of dimensions must not surpass the total number of documents. The accumulated singular value is expressed as a percentage (with no “%” sign). This percentage reflects the saved dimensionality, that is to say, the dimensionality percentage that will be maintained. In this way, 40% refers to the number of dimensions corresponding to said dimensionality. In extremely large corpora, it will not be possible to calculate this percentage, and so 300 dimensions will be used.

These data are entered in the matrix tab in the central panel.

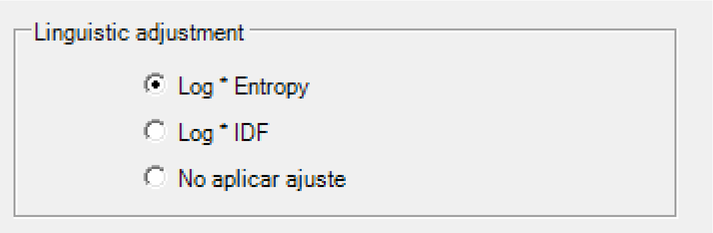


Matrix features

● Dimensions 2

○ Accumulated singular value

- **Linguistic adjustment:** This option refers to the significance calculation for every corpus term. Log-Entropy or log-IDF can be selected, as well as the absence of these calculations. This information is entered in the matrix tab in the central panel.



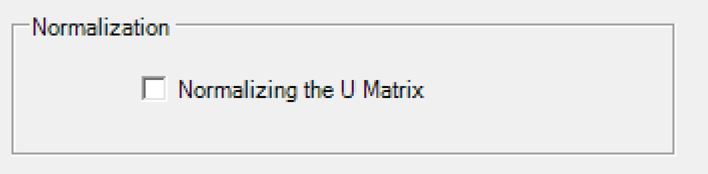
Linguistic adjustment

● Log * Entropy

○ Log * IDF

○ No aplicar ajuste

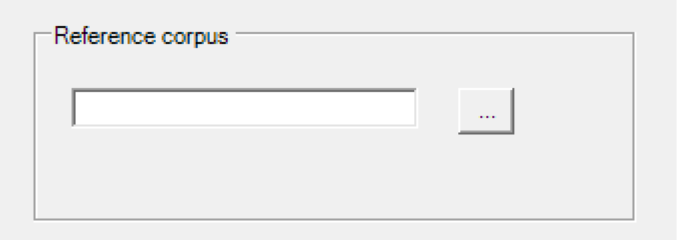
- **Normalizing the U Matrix:** The U matrix (of terms) extracted in the SVDF process is normalized before weighting it applying the weight of each dimension. In this way the effect of term frequency is countered. This information is entered in the matrix tab in the central panel.



Normalization

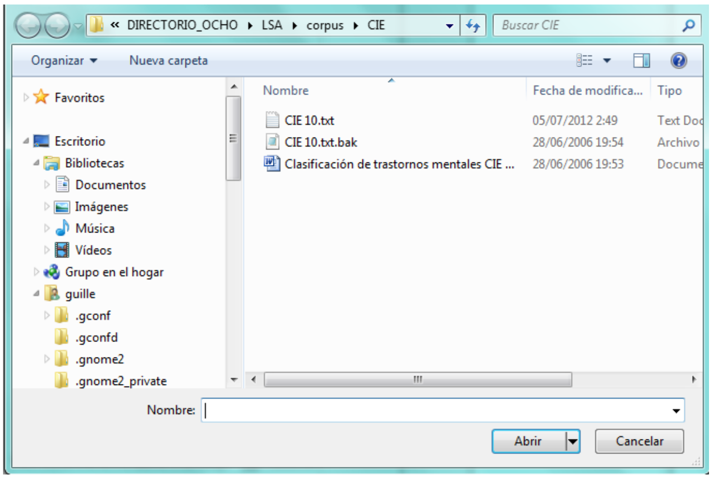
□ Normalizing the U Matrix

- **Reference corpus:** the path of the text file where the linguistic corpus is stored (in a legitimate format). Click on the button to browse through the directories. This information is entered in the Corpus tab in the central panel.

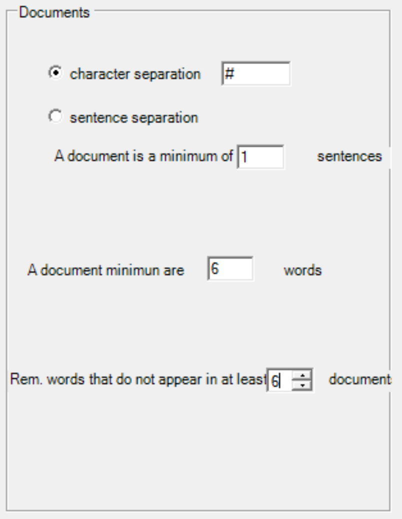


Reference corpus

The chosen corpus can be selected in the standard Windows browser.



- **Document separation:** The documents can be separated by a character or simply by natural-language sentences. In the former case, by the separating character (usually "#"). In the second case, the number of sentences constituting a document (usually 1) should be established. This information is entered in the Corpus tab in the central panel.



Documents

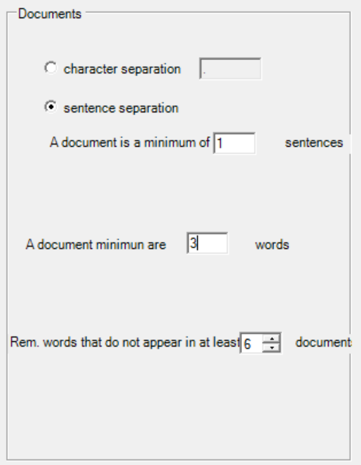
● character separation #

○ sentence separation

A document is a minimum of 1 sentences

A document minimun are 6 words

Rem. words that do not appear in at least 6 document



Documents

○ character separation

● sentence separation

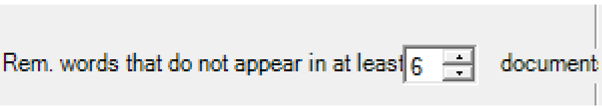
A document is a minimum of 1 sentences

A document minimun are 3 words

Rem. words that do not appear in at least 6 document

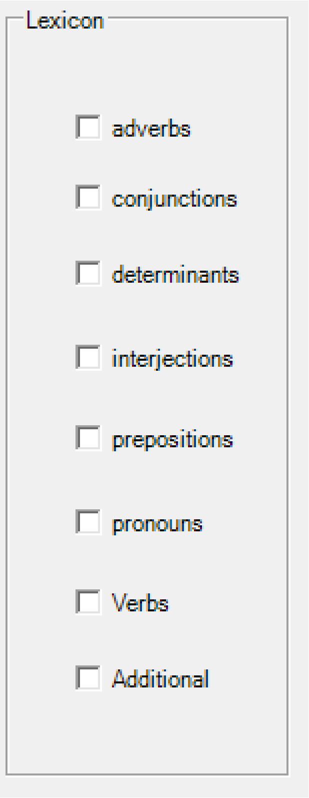
- **"A document is a minimum of ... words":** This is the minimum number of terms for a document to be analyzed. This information is entered in the Corpus tab in the central panel.

- **"Remove words that do not appear in at least … documents":** The minimum number of documents in which a specific term must appear to be included in the analysis. This information is entered in the Corpus tab in the central panel.



Rem. words that do not appear in at least 6 document

- **Remove:** Literal occurrences of each of the structures presented are removed. This information is entered in the Remove tab in the central panel.


Lexicon

□ adverbs

□ conjunctions

□ determinants

□ interjections

□ prepositions

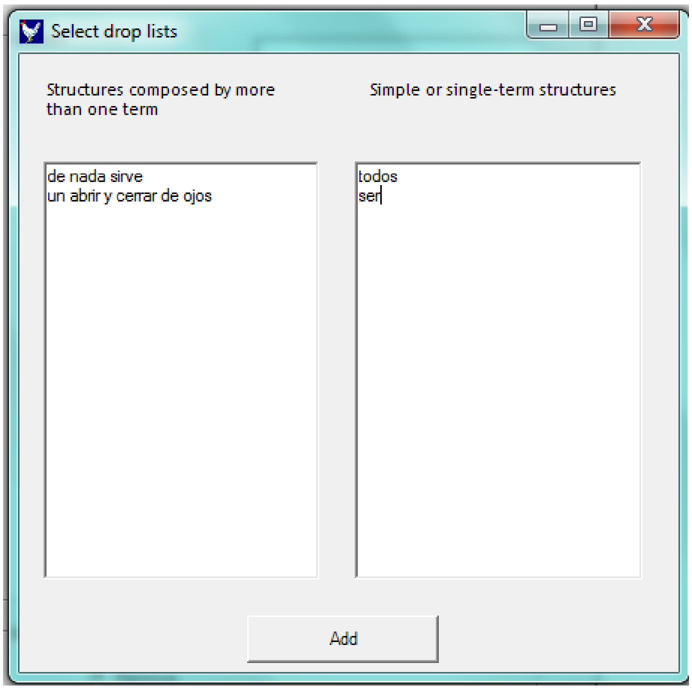
□ pronouns

□ Verbs

□ Additional

- **Generating a “stop list”:**

  1) In the Structures > Select drop lists, select and enter the structures that you wish to remove. In the left-hand side, you will find the structures composed by more than one term. In the right-hand side, you will find simple or single-term structures.



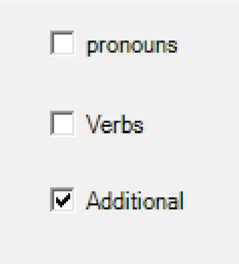
Structures composed by more than one term

de nada sirve
un abrir y cerrar de ojos

Simple or single-term structures

todos
ser

2) In the Remove tab, check the “Additional” box



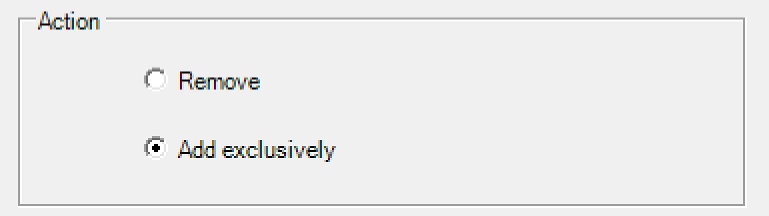
☐ pronouns

☐ Verbs

☑ Additional

- **Creating a space with only one “go list”**

  The procedure is similar to the previous one. In the Remove tab the “Additional” box is checked and in Structures > Select the structures in the “go list” are selected. The “Add exclusively” option must also be checked.



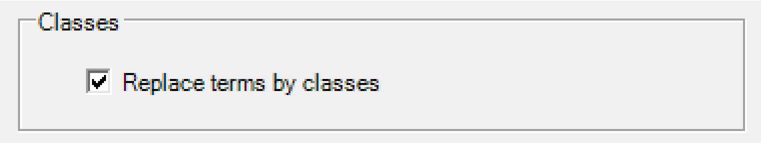
Action

○ Remove

● Add exclusively

Using this method you can select the remaining structures so that they are included in the analysis: for instance, an analysis with function words and adverbs and the selected additional structures. This information is entered in the Remove tab in the central panel.

- **Replace terms by classes:** To reduce the variability of some words, it can sometimes be interesting to include terms in categories and use those categories as terms. For instance, all mobile telephone brands could be included in the ClassMobile category. In this way, in the process both Nokia and Alcatel would be treated as the ClassMobile term, whose occurrences would increase. This information is entered in the Remove tab in the central panel.



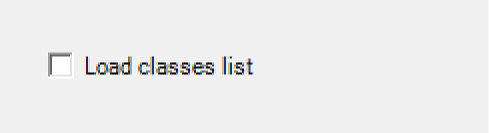
Classes

☑ Replace terms by classes

Class definition would be included in clases.txt in the directory.

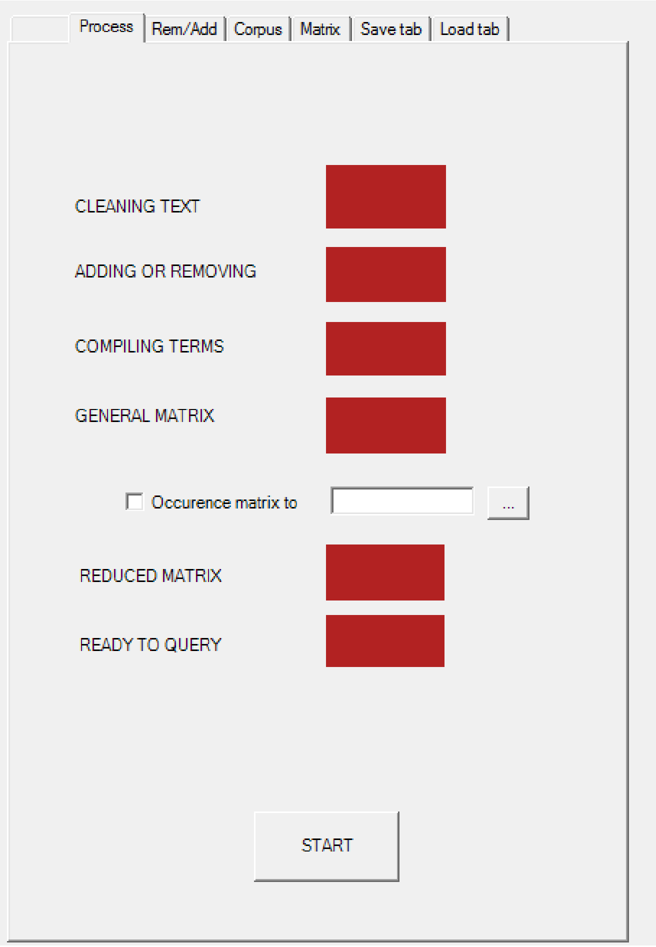
C:\Program Files\elsemantico.com\Gallito_2.0 64Bits\palabrasClases

If you want to load that has been previously processed using class definitions, it must be specified in the Load tab. The definitions used in the clases.txt must be provided.

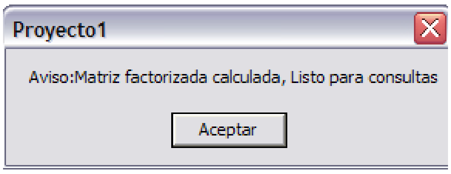


□ Load classes list

Once the parameters for corpus creation have been selected, open the Process tab and click on Start. Then the processes will be highlighted as they are executed. The total process can be slow and may even take days, depending on the options and the text corpus.

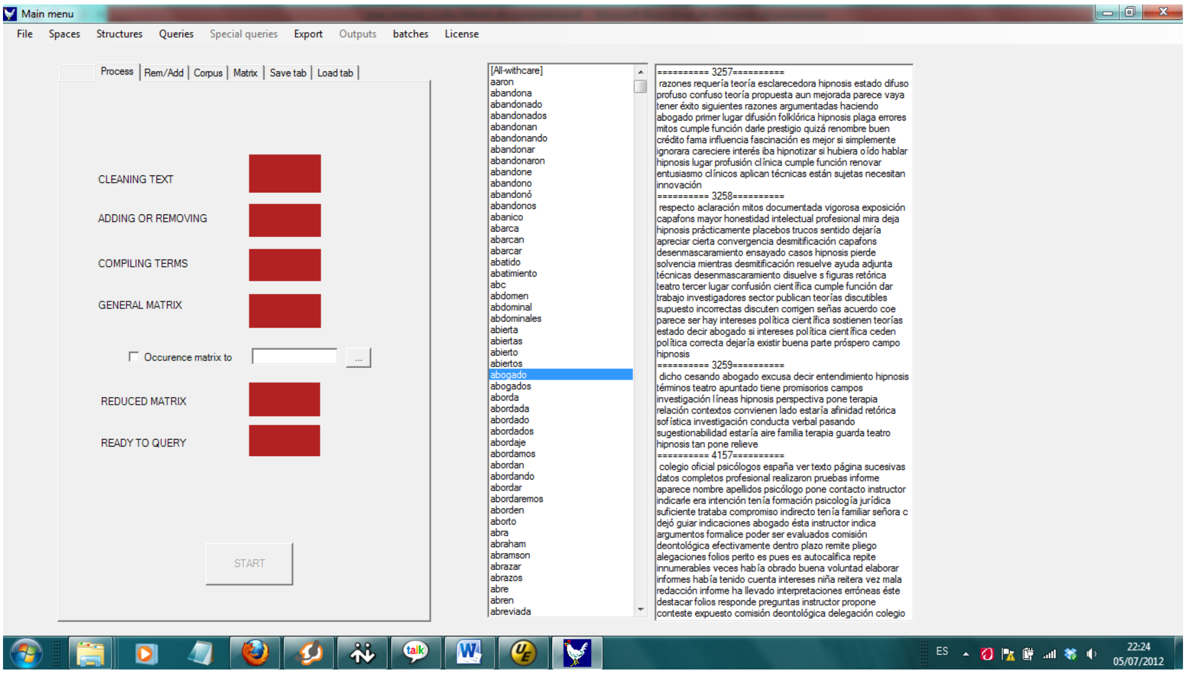


The final process ends with the following message:



Aviso:Matriz factorizada calculada, Listo para consultas

Once this warning is accepted, terms and documents will be loaded on the right-hand side and it will be possible to perform operations on the semantic space.



```
[All-withcare]
aaron
abandona
abandonado
abandonados
abandonan
abandonando
abandonar
abandonaron
abandone
abandono
abandonó
abandonos
abanico
abarca
abarcan
abarcar
abatido
abatimiento
abc
abdomen
abdominal
abdominales
abierta
abiertas
abierto
abiertos
abogado
abogados
aborda
abordada
abordado
abordados
abordaje
abordamos
abordan
abordando
abordar
abordaremos
aborden
aborto
abra
abraham
abramson
abrazar
abrazos
abre
abren
abreviada
```

```
========== 3257==========
 razones requería teoría esclarecedora hipnosis estado difuso
profuso confuso teoría propuesta aun mejorada parece vaya
tener éxito siguientes razones argumentadas haciendo
abogado primer lugar difusión folklórica hipnosis plaga errores
mitos cumple función darle prestigio quizá renombre buen
crédito fama influencia fascinación es mejor si simplemente
ignorara careciere interés iba hipnotizar si hubiera oído hablar
hipnosis lugar profusión clínica cumple función renovar
entusiasmo clínicos aplican técnicas están sujetas necesitan
innovación
========== 3258==========
 respecto aclaración mitos documentada vigorosa exposición
capafons mayor honestidad intelectual profesional mira deja
hipnosis prácticamente placebos trucos sentido dejaría
apreciar cierta convergencia desmitificación capafons
desenmascaramiento ensayado casos hipnosis pierde
solvencia mientras desmitificación resuelve ayuda adjunta
técnicas desenmascaramiento disuelve s figuras retórica
teatro tercer lugar confusión científica cumple función dar
trabajo investigadores sector publican teorías discutibles
supuesto incorrectas discuten corrigen señas acuerdo coe
parece ser hay intereses política científica sostienen teorías
estado decir abogado si intereses política científica ceden
política correcta dejaría existir buena parte próspero campo
hipnosis
========== 3259==========
 dicho cesando abogado excusa decir entendimiento hipnosis
téminos teatro apuntado tiene promisorios campos
investigación líneas hipnosis perspectiva pone terapia
relación contextos convienen lado estaría afinidad retórica
sofística investigación conducta verbal pasando
sugestionabilidad estaría aire familia terapia guarda teatro
hipnosis tan pone relieve
========== 4157==========
 colegio oficial psicólogos españa ver texto página sucesivas
datos completos profesional realizaron pruebas informe
aparece nombre apellidos psicólogo pone contacto instructor
indicarle era intención tenía formación psicología jurídica
suficiente trataba compromiso indirecto tenía familiar señora c
dejó guiar indicaciones abogado ésta instructor indica
argumentos formalice poder ser evaluados comisión
deontológica efectivamente dentro plazo remite pliego
alegaciones folios perito es pues es autocalifica repite
innumerables veces había obrado buena voluntad elaborar
informes había tenido cuenta intereses niña reitera vez mala
redacción informe ha llevado interpretaciones erróneas éste
destacar folios responde preguntas instructor propone
conteste expuesto comisión deontológica delegación colegio
```

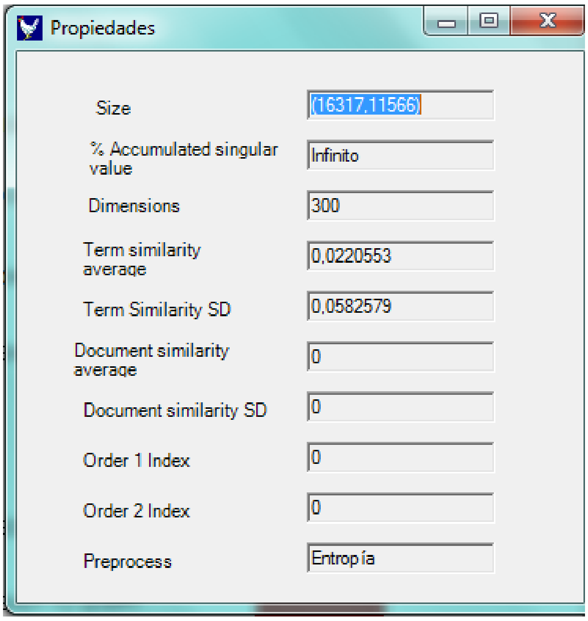
4.2 Operations on semantic spaces

Once a semantic space is created and loaded, operations can be performed on it.

These operations can be: comparing two terms, comparing two documents identified by a number, comparing two free texts entered by the user, extracting semantic neighbors (with simple or corrected cosines or with simple or corrected predication). The space can also be saved in a hard disc directory to be loaded at another time.

Space properties: In this option, the properties of the spaces on which the user is working can be viewed. Some of the indices will be disabled in the applications aimed at large linguistic corpora.

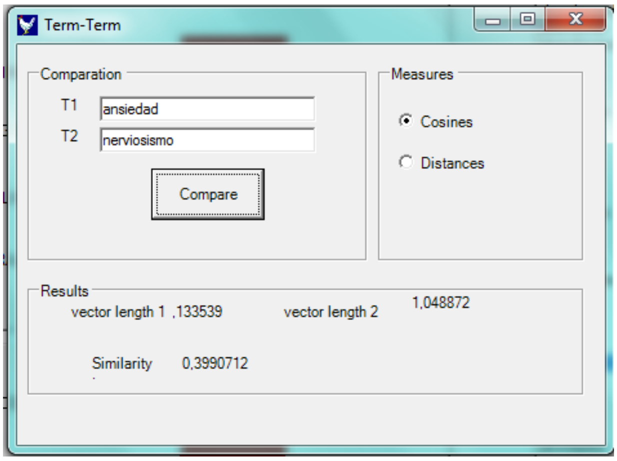
Space>Properties



Size	(16317,11566)
% Accumulated singular value	Infinito
Dimensions	300
Term similarity average	0,0220553
Term Similarity SD	0,0582579
Document similarity average	0
Document similarity SD	0
Order 1 Index	0
Order 2 Index	0
Preprocess	Entropía

Comparing two terms: Two specific terms can be compared by means of their cosine or the Euclidean distance between them.

Queries > Term-Term



T1 ansiedad

T2 nerviosismo

● Cosines

○ Distances

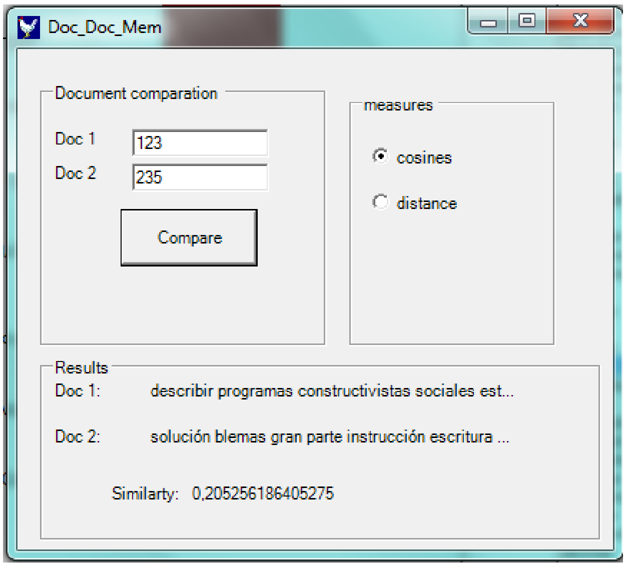
vector length 1 ,133539

vector length 2 1,048872

Similarity 0,3990712

Comparing two documents identified by a number: Two specific documents can be compared by means of the cosine or the Euclidean distance between them.

Queries > Document-Document



Doc 1 123

Doc 2 235

● cosines

○ distance

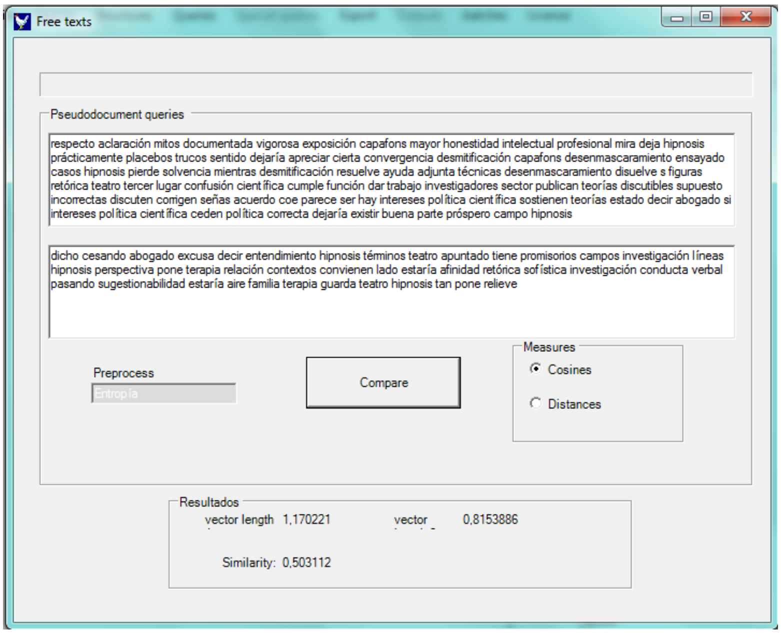
Doc 1: describir programas constructivistas sociales est...

Doc 2: solución blemas gran parte instrucción escritura ...

Similarty: 0,205256186405275

Comparing two freely produced texts: Two free texts can be compared by means of their cosine or the Euclidean distance between them. (In large-size spaces this process will take a few seconds. Process progress will be shown by means of a status bar).

Queries > Free texts



Pseudodocument queries

respecto aclaración mitos documentada vigorosa exposición capafons mayor honestidad intelectual profesional mira deja hipnosis prácticamente placebos trucos sentido dejaría apreciar cierta convergencia desmitificación capafons desenmascaramiento ensayado casos hipnosis pierde solvencia mientras desmitificación resuelve ayuda adjunta técnicas desenmascaramiento disuelve s figuras retórica teatro tercer lugar confusión científica cumple función dar trabajo investigadores sector publican teorías discutibles supuesto incorrectas discuten corrigen señas acuerdo coe parece ser hay intereses política científica sostienen teorías estado decir abogado si intereses política científica ceden política correcta dejaría existir buena parte próspero campo hipnosis

dicho cesando abogado excusa decir entendimiento hipnosis términos teatro apuntado tiene promisorios campos investigación líneas hipnosis perspectiva pone terapia relación contextos convienen lado estaría afinidad retórica sofística investigación conducta verbal pasando sugestionabilidad estaría aire familia terapia guarda teatro hipnosis tan pone relieve

Preprocess: Entropía

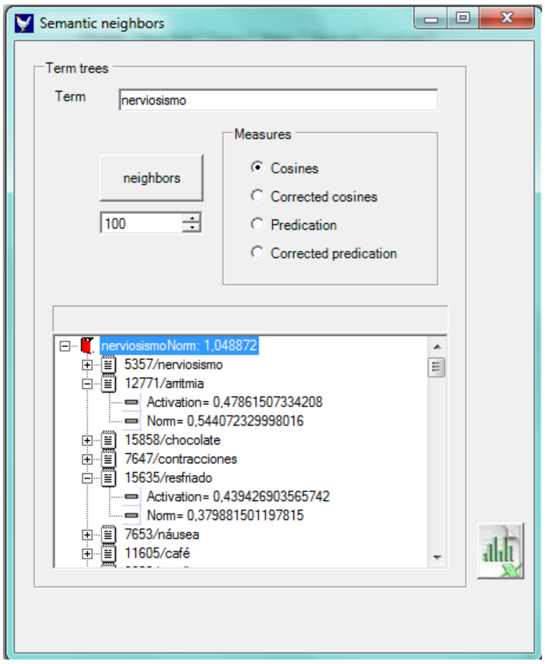
Measures
● Cosines
○ Distances

Resultados
vector length 1,170221 vector 0,8153886

Similarity: 0,503112

Extraction of semantic neighbors: The semantic neighbors of a specific term are extracted by various methods (cosines, corrected cosines, predication, corrected predication). Neighbor trees will be displayed as term neighbors are extracted. In addition, the number of neighbors to be extracted can be selected. The results can also be exported to Microsoft Excel.

Queries > Semantic neighbors

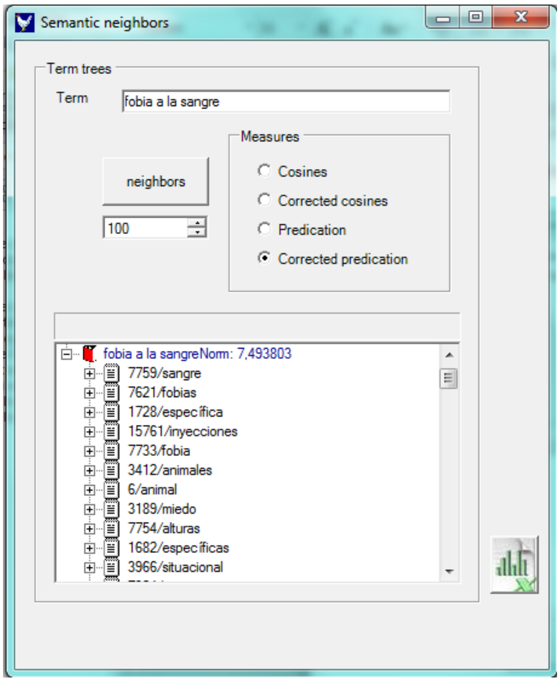


Term: nerviosismo

100

Measures
- ● Cosines
- ○ Corrected cosines
- ○ Predication
- ○ Corrected predication

```
nerviosismoNorm: 1,048872
   5357/nerviosismo
   12771/arritmia
      Activation= 0,47861507334208
      Norm= 0,544072329998016
   15858/chocolate
   7647/contracciones
   15635/resfriado
      Activation= 0,439426903565742
      Norm= 0,379881501197815
   7653/náusea
   11605/café
```



Term: fobia a la sangre

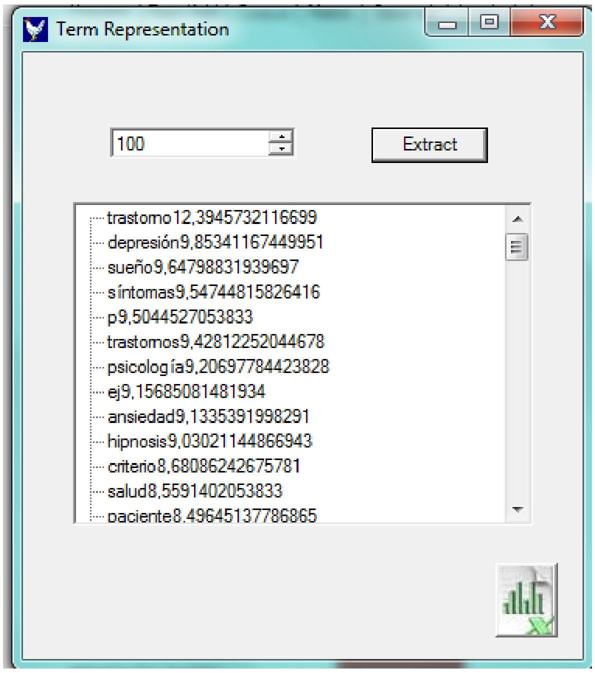
100

Measures
- ○ Cosines
- ○ Corrected cosines
- ○ Predication
- ● Corrected predication

```
fobia a la sangreNorm: 7,493803
   7759/sangre
   7621/fobias
   1728/específica
   15761/inyecciones
   7733/fobia
   3412/animales
   6/animal
   3189/miedo
   7754/alturas
   1682/específicas
   3966/situacional
```

Extraction of the most representative terms in the semantic space: The semantic neighbors of specific terms can be extracted by various methods.

Queries > Representation

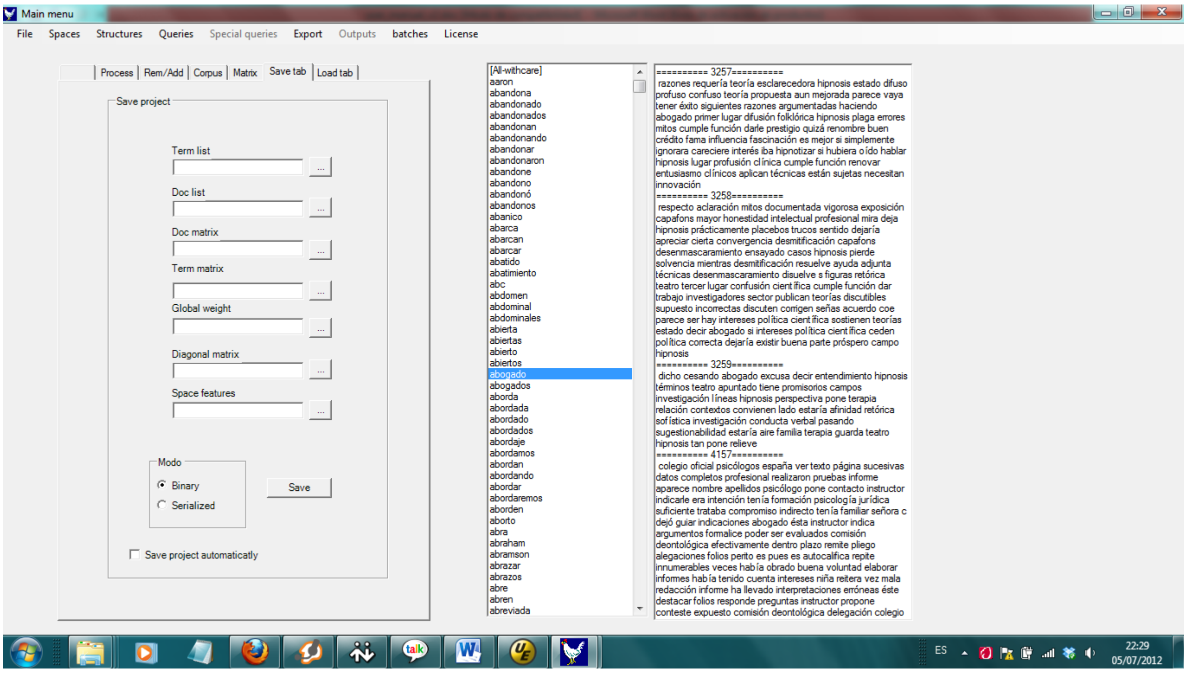


```
trastorno12,3945732116699
depresión9,85341167449951
sueño9,64798831939697
síntomas9,54744815826416
p9,5044527053833
trastornos9,42812252044678
psicología9,20697784423828
ej9,15685081481934
ansiedad9,1335391998291
hipnosis9,03021144866943
criterio8,68086242675781
salud8,5591402053833
paciente8,49645137786865
```

Saving a semantic space: A semantic space can be saved in the hard disc and reloaded when necessary with no need to creating it against.

Save tab

Then a name and path to save 7 variables must be selected. We recommend that you save all spaces in the same directory, which can be created in the same directory browsers as each of the variables., At the end, select the format (we recommend “binary”) and click on Save.



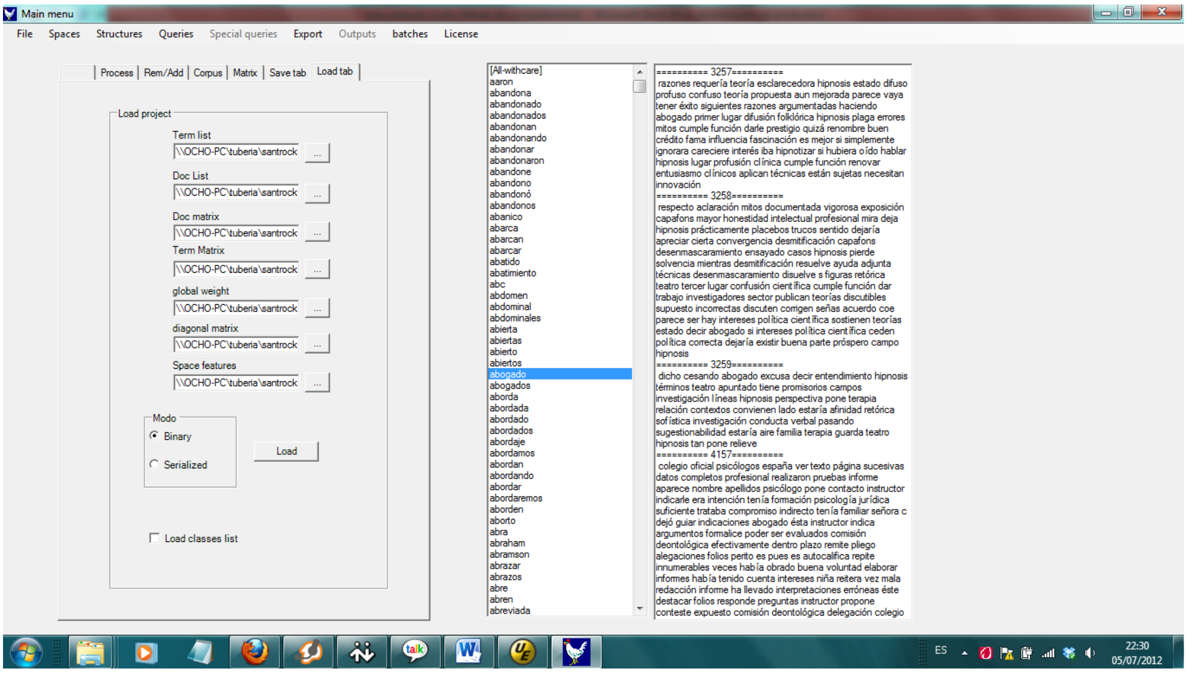
```
[All-withcare]
aaron
abandona
abandonado
abandonados
abandonan
abandonando
abandonar
abandonaron
abandone
abandono
abandonó
abandonos
abanico
abarca
abarcan
abarcar
abatido
abatimiento
abc
abdomen
abdominal
abdominales
abierta
abiertas
abierto
abiertos
abogado
abogados
aborda
abordada
abordado
abordados
abordaje
abordamos
abordan
abordando
abordar
abordaremos
aborden
aborto
abra
abraham
abramson
abrazar
abrazos
abre
abren
abreviada
```

```
========== 3257==========
 razones requería teoría esclarecedora hipnosis estado difuso
profuso confuso teoría propuesta aun mejorada parece vaya
tener éxito siguientes razones argumentadas haciendo
abogado primer lugar difusión folklórica hipnosis plaga errores
mitos cumple función darle prestigio quizá renombre buen
crédito fama influencia fascinación es mejor si simplemente
ignorara careciere interés iba hipnotizar si hubiera oído hablar
hipnosis lugar profusión clínica cumple función renovar
entusiasmo clínicos aplican técnicas están sujetas necesitan
innovación
========== 3258==========
 respecto aclaración mitos documentada vigorosa exposición
capafons mayor honestidad intelectual profesional mira deja
hipnosis prácticamente placebos trucos sentido dejaría
apreciar cierta convergencia desmitificación capafons
desenmascaramiento ensayado casos hipnosis pierde
solvencia mientras desmitificación resuelve ayuda adjunta
técnicas desenmascaramiento disuelve s figuras retórica
teatro tercer lugar confusión científica cumple función dar
trabajo investigadores sector publican teorías discutibles
supuesto incorrectas discuten corrigen señas acuerdo coe
parece ser hay intereses política científica sostienen teorías
estado decir abogado si intereses política científica ceden
política correcta dejaría existir buena parte próspero campo
hipnosis
========== 3259==========
 dicho cesando abogado excusa decir entendimiento hipnosis
términos teatro apuntado tiene promisorios campos
investigación líneas hipnosis perspectiva pone terapia
relación contextos convienen lado estaría afinidad retórica
sofística investigación conducta verbal pasando
sugestionabilidad estaría aire familia terapia guarda teatro
hipnosis tan pone relieve
========== 4157==========
 colegio oficial psicólogos españa ver texto página sucesivas
datos completos profesional realizaron pruebas informe
aparece nombre apellidos psicólogo pone contacto instructor
indicarle era intención tenía formación psicología jurídica
suficiente trataba compromiso indirecto tenía familiar señora c
dejó guiar indicaciones abogado ésta instructor indica
argumentos formalice poder ser evaluados comisión
deontológica efectivamente dentro plazo remite pliego
alegaciones folios perito es pues es autocalifica repite
innumerables veces había obrado buena voluntad elaborar
informes había tenido cuenta intereses niña reitera vez mala
redacción informe ha llevado interpretaciones erróneas éste
destacar folios responde preguntas instructor propone
conteste expuesto comisión deontológica delegación colegio
```

Loading a semantic space. In the same way, a semantic space can be loaded from the hard disc with no need to create it again.

Load Tab

A name and route to locate the 7 variables of the space to be loaded must be selected. In the same directory browser as each of the variables, you can browse and select the path for the relevant variable. At the end, select the format in which the space was saved (we recommend “binary”).



Term list: \\OCHO-PC\tuberia\santrock

Doc List: \\OCHO-PC\tuberia\santrock

Doc matrix: \\OCHO-PC\tuberia\santrock

Term Matrix: \\OCHO-PC\tuberia\santrock

global weight: \\OCHO-PC\tuberia\santrock

diagonal matrix: \\OCHO-PC\tuberia\santrock

Space features: \\OCHO-PC\tuberia\santrock

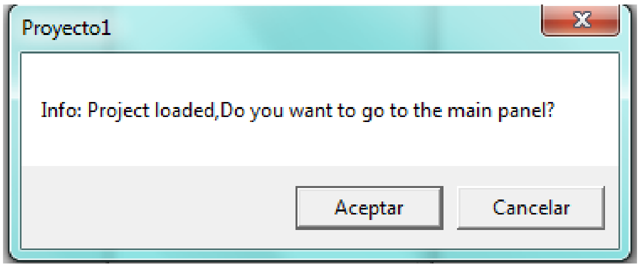
Modo: ● Binary ○ Serialized

□ Load classes list

[All-withcare]
aaron
abandona
abandonado
abandonados
abandonan
abandonando
abandonar
abandonaron
abandone
abandono
abandonó
abandonos
abanico
abarca
abarcan
abarcar
abatido
abatimiento
abc
abdomen
abdominal
abdominales
abierta
abiertas
abierto
abiertos
abogado
abogados
aborda
abordada
abordado
abordados
abordaje
abordamos
abordan
abordando
abordar
abordaremos
aborden
aborto
abra
abraham
abramson
abrazar
abrazos
abre
abren
abreviada

```
========== 3257==========
 razones requería teoría esclarecedora hipnosis estado difuso
profuso confuso teoría propuesta aun mejorada parece vaya
tener éxito siguientes razones argumentadas haciendo
abogado primer lugar difusión folklórica hipnosis plaga errores
mitos cumple función darle prestigio quizá renombre buen
crédito fama influencia fascinación es mejor si simplemente
ignorara careciere interés iba hipnotizar si hubiera oído hablar
hipnosis lugar profusión clínica cumple función renovar
entusiasmo clínicos aplican técnicas están sujetas necesitan
innovación
========== 3258==========
 respecto aclaración mitos documentada vigorosa exposición
capafons mayor honestidad intelectual profesional mira deja
hipnosis prácticamente placebos trucos sentido dejaría
apreciar cierta convergencia desmitificación capafons
desenmascaramiento ensayado casos hipnosis pierde
solvencia mientras desmitificación resuelve ayuda adjunta
técnicas desenmascaramiento disuelve s figuras retórica
teatro tercer lugar confusión científica cumple función dar
trabajo investigadores sector publican teorías discutibles
supuesto incorrectas discuten corrigen señas acuerdo coe
parece ser hay intereses política científica sostienen teorías
estado decir abogado si intereses política científica ceden
política correcta dejaría existir buena parte próspero campo
hipnosis
========== 3259==========
 dicho cesando abogado excusa decir entendimiento hipnosis
términos teatro apuntado tiene promisorios campos
investigación líneas hipnosis perspectiva pone terapia
relación contextos convienen lado estaría afinidad retórica
sofística investigación conducta verbal pasando
sugestionabilidad estaría aire familia terapia guarda teatro
hipnosis tan pone relieve
========== 4157==========
 colegio oficial psicólogos españa ver texto página sucesivas
datos completos profesional realizaron pruebas informe
aparece nombre apellidos psicólogo pone contacto instructor
indicarle era intención tenía formación psicología jurídica
suficiente trataba compromiso indirecto tenía familiar señora c
dejó guiar indicaciones abogado ésta instructor indica
argumentos formalice poder ser evaluados comisión
deontológica efectivamente dentro plazo remite pliego
alegaciones folios perito es pues es autocalifica repite
innumerables veces había obrado buena voluntad elaborar
informes había tenido cuenta intereses niña reitera vez mala
redacción informe ha llevado interpretaciones erróneas éste
destacar folios responde preguntas instructor propone
conteste expuesto comisión deontológica delegación colegio
```

After this, click Load and wait for the following message:



Info: Project loaded,Do you want to go to the main panel?

Select Accept and the space will be loaded in the memory to perform the operations described above.

4.3 Batch processing

Actions can be performed by batches, specifying in a file the terms and characteristics of the operations to be performed.

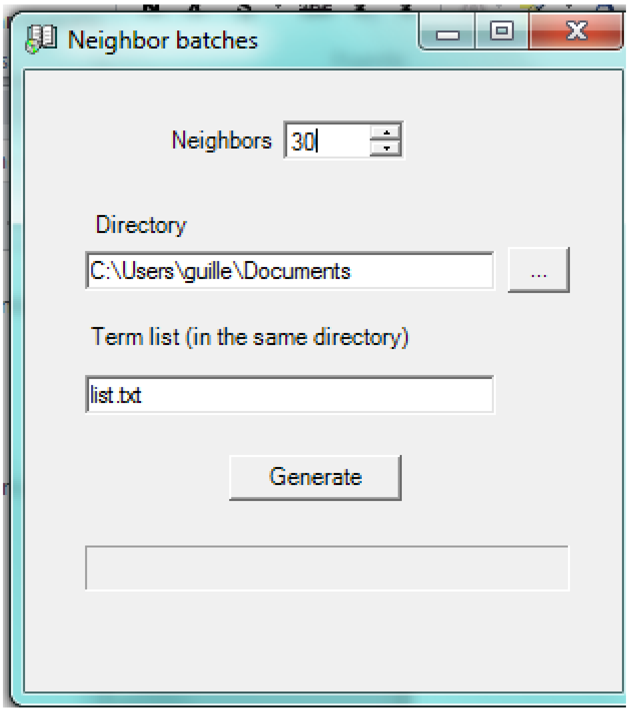
4.3.1 Neighbor batches

The first n semantic neighbors will be extracted from a number of terms. In the expandable menu, the following screen will appear, where the number of neighbors, the directory where the file specifying the terms, and the file itself are specified.

The file will have the following format

match
football
note

The same process will generate a file per term, specifying the neighbor, cosine, and vector length



Neighbor batches

Neighbors 30

Directory

C:\Users\guille\Documents

Term list (in the same directory)

list.txt

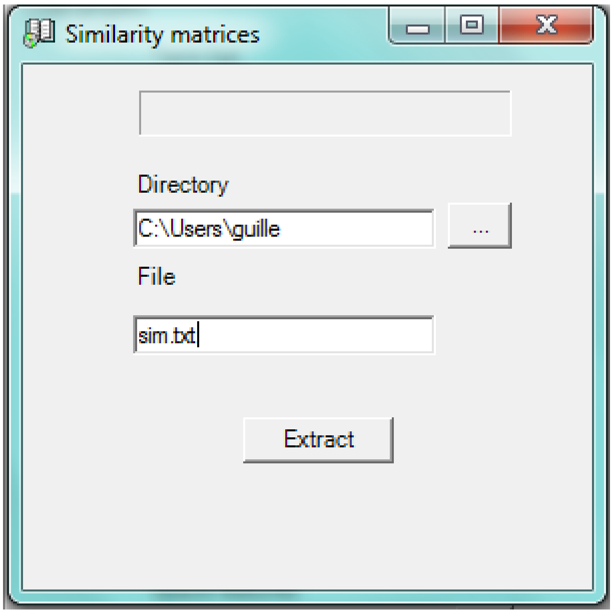
4.3.2 Similarity matrices

A matrix will be extracted to compare the neighbors with each other. This square matrix has ones in its diagonal and each cell represents the cosine of two neighbors. The directory for the reference file and the file name itself will be specified. Similarity matrices will be generated in that same directory.

The file format is the following:

match|200
football|300
note|300

Where a 200x200 will be extracted in the first case, 300x300 in the second one, etc.



Directory
C:\Users\guille

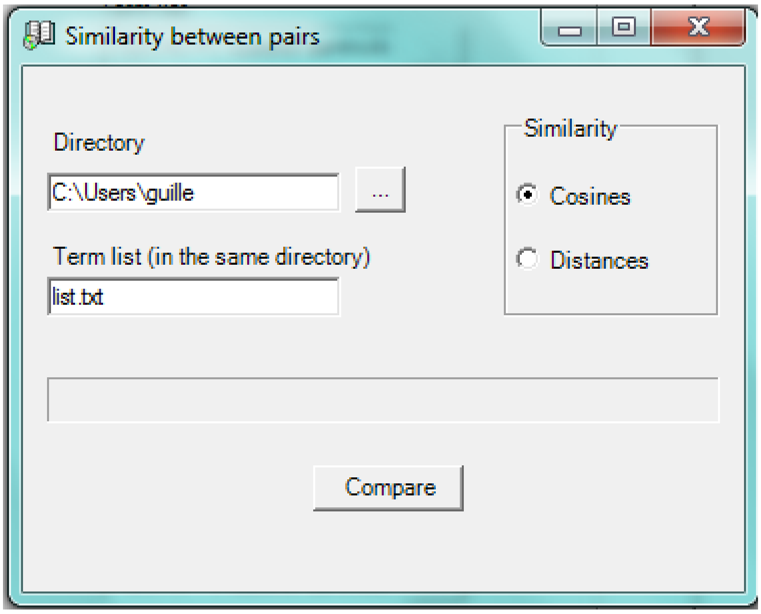
File
sim.txt

4.3.3 Similarity between pairs

This will generate the similarities between a number of term pairs. The directory for the reference file and name of the file itself will be specified.

The file format will be the following:

tremendous|action
dog|cat
share|stock



Directory
C:\Users\guille

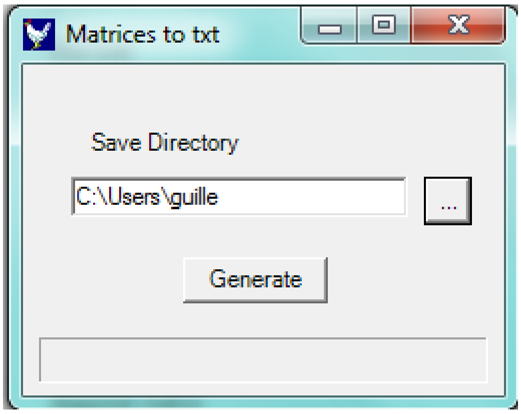
Term list (in the same directory)
list.txt

Similarity
● Cosines
○ Distances

4.4 Standard output

4.4.1 Matrices to plain text files

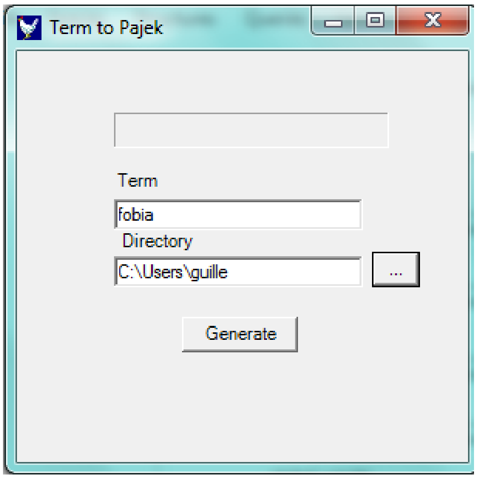
The Export drop menu includes the option “Matrices to .txt”. Just enter the directory in which you wish to generate the files, which will be the US, SV, and S matrices, as well as the weights assigned by the calculations to each word (log-entropy or log-idf) and the vector length for each term.



C:\Users\guille

4.4.1 Term to Pajek file

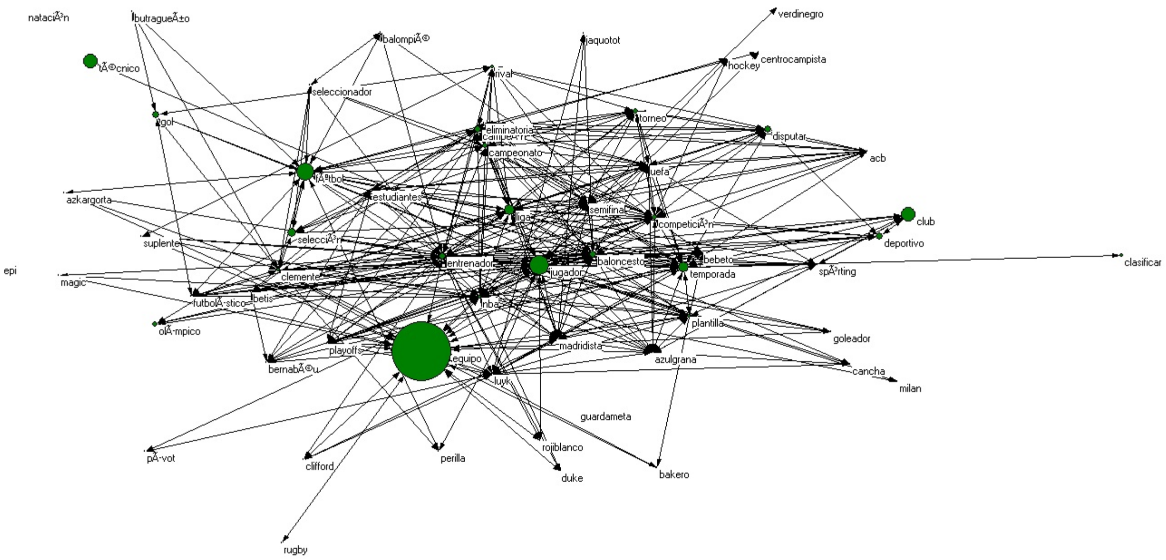
For visualization purposes, there is a Term to Pajek option in the Export drop menu. From this process a file is extracted that will serve as input for the Pajek program to generate visualization networks.



fobia

C:\Users\guille

Pajek is available at http://vlado.fmf.uni-lj.si/pub/networks/pajek/ and can be downloaded for free. This output will use cosines as similarity and vector length as node size. By entering the input, graphs such as the one below can be obtained.


nataciÃ³n
butragueÃ±o
tÃ©cnico
balompiÃ©
jaquotot
verdinegro
hockey
centrocampista
rival
seleccionador
gol
eliminatoria
campeÃ³n
campeonato
torneo
disputar
acb
fÃºtbol
uefa
estudiantes
azkargorta
semifinal
liga
competiciÃ³n
club
selecciÃ³n
deportivo
suplente
bebeto
epi
entrenador
jugador
baloncesto
temporada
spÃ³rting
clasificar
magic
clemente
futbolÃ-stico
betis
nba
olÃ-mpico
plantilla
goleador
madridista
playoffs
equipo
azulgrana
bernabÃ©u
cancha
luyk
milan
guardameta
pÃ-vot
rojiblanco
clifford
perilla
bakero
duke
rugby