

Humanities Going Digital (HUGOD)

[2020-1-CZ01-KA226-HE-094363]

Organização de Dados
AntConc

Sessão 05

26/10/2022

Chiara Barbero
Sílvia Barbosa



AULAS

Organização dos dados



07

08

09

10

11

12

Análise de Dados



Trabalho final escrito



Extração de dados



Dúvidas



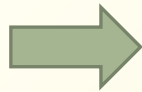
Apresentação de trabalhos





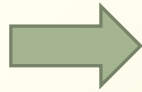
Pasta

- Nome do ficheiro deve ser claro (alfanumérico)
 - 20221231_JN_12345_original
 - 20221231_JN_12345_annotado
- Ficheiros independentes / ficheiro único
- A extensão dos ficheiros deve ser *plain text* – .TXT



- AntConc
- Hyperbase
- WordSmith Tools
- Sketch Engine





- Estes programas ajudam a analisar estatisticamente o *corpus* / o texto / a unidade lexical
- Identifica a quantidade de unidades lexicais, as frequências, os *hápax*
- Lematizar
-

6

ANTCONC



interface inicial

AntConc

File Edit Settings Help

1

Target Corpus

Name:

Files:

Tokens:

Progress 0%

KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud

Total Hits: 549 Page Size 100 hits 1 to 100 of 549 hits

| File | Left Context | Hit | Right Context |
|------|--------------|-----|---------------|
|------|--------------|-----|---------------|

Search Query Words Case Regex Results Set All hits Context Size 8 token(s)

Start Adv Search

Sort Options Sort center to left Sort 1 C Sort 2 1L Sort 3 2L Order by freq

Use Word cache Time taken (creating word list results): 0.1544 sec



AntConc - Global Settings

Category

- Colors
- Files
- Fonts
- Language Direction
- Searches
- Restore Settings
- Statistics
- Tags
- Tool Filters

Search Settings

Wildcards

| | |
|-----------|--|
| ? | Any one character |
| * | Zero or more characters |
| + | One or more characters |
| [abc,def] | Alternatives (comma separated in brackets) |
| | Search term 'OR' |

Os ficheiros TXT devem estar em Unicode (UTF-8)

AntConc - Tool Settings

Category

- KWIC
- Plot
- Cluster
- N-Gram
- Collocate
- Word
- Keyword

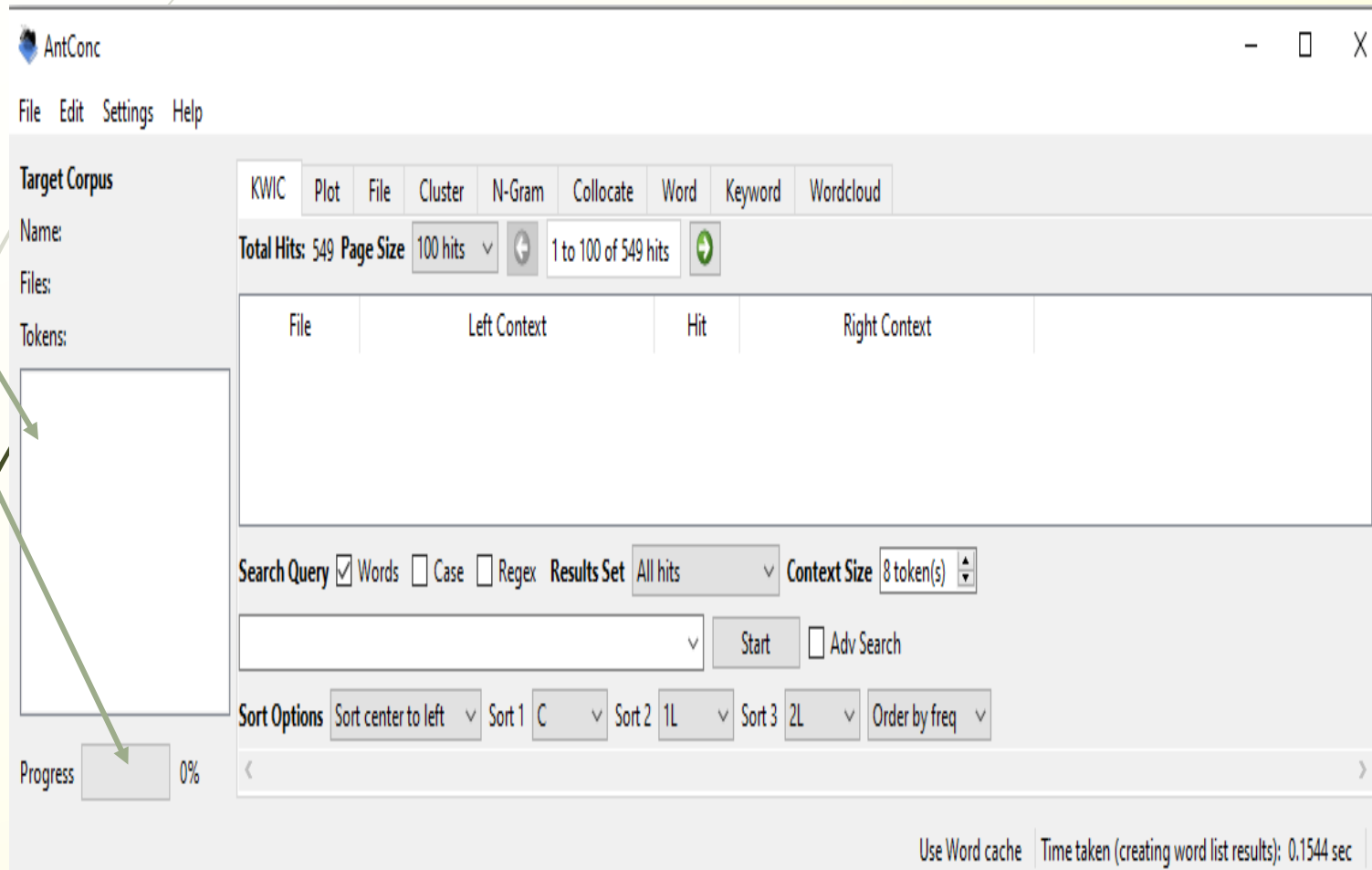
Word List Settings

List Type

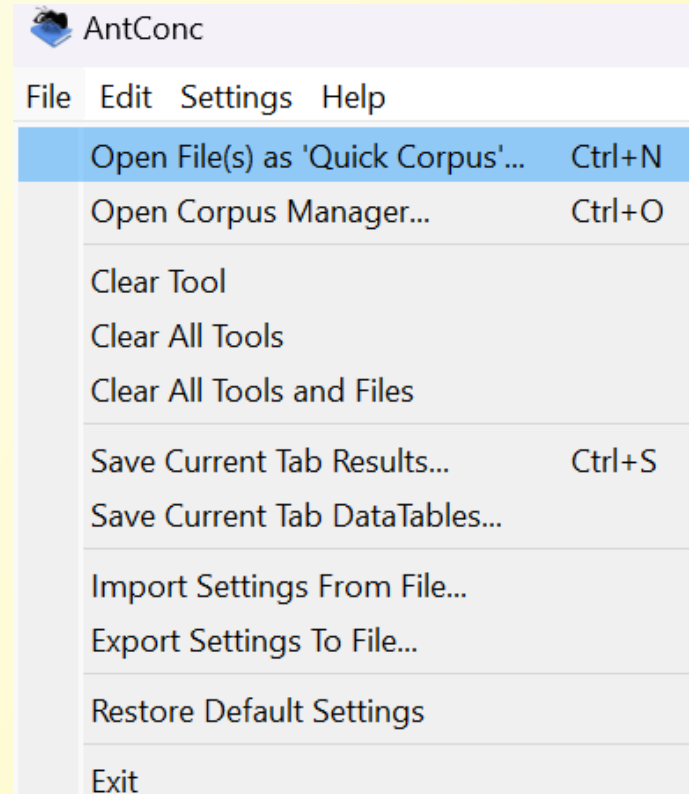
Word List

Display Options

- Show types
- Show headwords
- Show raw values
- Show normalized values



Para criar um *corpus*





AntConc

File Edit Settings Help

Target Corpus

Name: temp
Files: 8
Tokens: 48223

Art_01.txt
Art_02.txt
Art_03.txt
Art_04.txt
Pap_01.txt
Pap_02.txt
Pap_03.txt
Proceed_01.txt

Progress 100%

KWIC Plot File Cluster N-Gram Collocate **Word** Keyword Wordcloud

Types 8915/8915 Tokens 48223/48223 Page Size 100 hits 1 to 100 of 8915 hits

| | Type | Rank | Freq | Range |
|----|---------|------|------|-------|
| 1 | de | 1 | 2429 | 8 |
| 2 | a | 2 | 1667 | 8 |
| 3 | e | 3 | 1380 | 8 |
| 4 | que | 4 | 1143 | 8 |
| 5 | o | 5 | 994 | 8 |
| 6 | da | 6 | 919 | 8 |
| 7 | do | 7 | 788 | 8 |
| 13 | pública | 13 | 387 | 8 |
| 14 | no | 14 | 372 | 8 |

Search Query Words Case Regex

Start Adv Search

Sort by Frequency Invert Order

Use Word cache | Time taken (creating word list results): 0.2978 sec

3



AntConc

File Edit Settings Help

Target Corpus
 Name: temp
 Files: 4
 Tokens: 391079

A Marquesa de Vale Neg
 Ambições - Ana de Cast
 Mistérios de Lisboa - Car
 Viagens na minha Terra _

4

KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud

Total Hits: 898 Page Size 100 hits 1 to 100 of 898 hits

| | File | Left Context | Hit | Right Context |
|---|---------------|--|-------|--|
| 1 | Mistérios ... | espondeu o padre, sorrindo-se --, chama-se | padre | Dinis Ramalho e Sousa... é como o mundo me |
| 2 | Mistérios ... | Preciso saber se está presente o reverendo | padre | Dinis Ramalho e Sousa. -- Um seu servo -- disse o |
| 3 | Mistérios ... | s que a nobreza da sua alma lhe custou. Ao | padre | Dinis Ramalho e Sousa legava o seu retrato, a |
| 4 | Mistérios ... | alguns meses depois Sebastião de Melo era | padre | Dinis Ramalho e Sousa. Nesse mesmo ano, entrou naEncar |
| 5 | Mistérios ... | arenta e quatro, mora um sujeito chamado | padre | Dinis Ramalho e Sousa. Tenha Vossa Senhoria a generosida |
| 6 | Mistérios ... | itamente venturosos, quando foi anunciado | padre | Dinis Ramalho e Sousa.Eugénia, como surpreendida, estren |
| 7 | Mistérios ... | ição competente fora tirado passaporte por | padre | Dinis Ramalho e Sousa. As suas investigações chegaram a |
| 8 | Mistérios ... | icórdia, com a condição de que ele doador, | padre | Dinis Ramalho e Sousa, seria recebido na enfermaria dos |
| 9 | Mistérios ... | arto particular do Hospital de S. José, estava | padre | Dinis Ramalho e Sousa. Ao lado do seu leito |

Search Query Words Case Regex Results Set All hits Context Size 10 token(s)

padre Start Adv Search

Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq

Progress 100%

Time taken (creating kwic results): 0.4655 sec



AntConc

File Edit Settings Help

Target Corpus

Name: temp
Files: 4
Tokens: 391079

KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud

Types 32673/32673 Tokens 391079/391079 Page Size 100 hits 1 to 100 of 32673 hits

| | Type | Rank | Freq | Range |
|----|--------|------|------|-------|
| 31 | seu | 31 | 1443 | 4 |
| 32 | minha | 32 | 1411 | 4 |
| 33 | dos | 33 | 1323 | 4 |
| 34 | meu | 34 | 1095 | 4 |
| 35 | quando | 35 | 1018 | 4 |
| 36 | das | 36 | 992 | 4 |
| 37 | padre | 37 | 898 | 4 |
| 38 | porque | 38 | 896 | 4 |
| 39 | sem | 39 | 884 | 4 |
| 40 | homem | 40 | 881 | 4 |
| 41 | muito | 41 | 869 | 4 |

Search Query Words Case Regex

padre Start Adv Search

Sort by Frequency Invert Order

Progress 100%

Clicando em “padre” aparece no menu KWIC as concordâncias

AntConc

File Edit Settings Help

Target Corpus

Name: temp
Files: 4
Tokens: 391079

KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud

Total Hits: 898 Page Size 100 hits 1 to 100 of 898 hits

| | File | Left Context | Hit | Right Context |
|---|---------------|--|-------|--|
| 1 | Mistérios ... | espondeu o padre, sorrindo-se --, chama-se | padre | Dinis Ramalho e Sousa... é como o mundo me |
| 2 | Mistérios ... | Preciso saber se está presente o reverendo | padre | Dinis Ramalho e Sousa. -- Um seu servo -- disse o |
| 3 | Mistérios ... | s que a nobreza da sua alma lhe custou. Ao | padre | Dinis Ramalho e Sousa legava o seu retrato, a |
| 4 | Mistérios ... | alguns meses depois Sebastião de Melo era | padre | Dinis Ramalho e Sousa. Nesse mesmo ano, entrou naEncar |
| 5 | Mistérios ... | arenta e quatro, mora um sujeito chamado | padre | Dinis Ramalho e Sousa. Tenha Vossa Senhoria a generosida |
| 6 | Mistérios ... | itamente venturosos, quando foi anunciado | padre | Dinis Ramalho e Sousa.Eugénia, como surpreendida, estren |
| 7 | Mistérios ... | ção competente fora tirado passaporte por | padre | Dinis Ramalho e Sousa. As suas investigações chegaram a |
| 8 | Mistérios ... | icórdia, com a condição de que ele doador, | padre | Dinis Ramalho e Sousa, seria recebido na enfermaria dos |
| 9 | Mistérios ... | arto particular do Hospital de S. José, estava | padre | Dinis Ramalho e Sousa. Ao lado do seu leito |

Search Query Words Case Regex Results Set All hits Context Size 10 token(s)

padre Start Adv Search

Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq

Progress 100%

Time taken (creating kwic results): 0.4655 sec



AntConc

File Edit Settings Help

Target Corpus
Name: temp
Files: 4
Tokens: 391079

KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud

Total Hits: 898 Page Size 100 hits 1 to 100 of 898 hits

| | File | Left Context | Hit | Right Context |
|---|---------------|--|-------|--|
| 1 | Mistérios ... | espondeu o padre, sorrindo-se --, chama-se | padre | Dinis Ramalho e Sousa... é como o mundo me |
| 2 | Mistérios ... | Preciso saber se está presente o reverendo | padre | Dinis Ramalho e Sousa. -- Um seu servo -- disse o |
| 3 | Mistérios ... | s que a nobreza da sua alma lhe custou. Ao | padre | Dinis Ramalho e Sousa legava o seu retrato, a |
| 4 | Mistérios ... | alguns meses depois Sebastião de Melo era | padre | Dinis Ramalho e Sousa. Nesse mesmo ano, entrou naEncar |
| 5 | Mistérios ... | arenta e quatro, mora um sujeito chamado | padre | Dinis Ramalho e Sousa. Tenha Vossa Senhoria a generosida |
| 6 | Mistérios ... | itamente venturosos, quando foi anunciado | padre | Dinis Ramalho e Sousa.Eugénia, como surpreendida, estren |
| 7 | Mistérios ... | ição competente fora tirado passaporte por | padre | Dinis Ramalho e Sousa. As suas investigações chegaram a |
| 8 | Mistérios ... | icórdia, com a condição de que ele doador, | padre | Dinis Ramalho e Sousa, seria recebido na enfermaria dos |
| 9 | Mistérios ... | arto particular do Hospital de S. José, estava | padre | Dinis Ramalho e Sousa. Ao lado do seu leito |

Search Query Words Case Regex Results Set All hits Context Size 10 token(s)

padre Start Adv Search

Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq

Progress 100%

Time taken (creating kwic results): 0.4655 sec

Para salvar as listagens de concordâncias

AntConc

File Edit Settings Help

- Open File(s) as 'Quick Corpus'... Ctrl+N
- Open Corpus Manager... Ctrl+O
- Clear Tool
- Clear All Tools
- Clear All Tools and Files
- Save Current Tab Results... Ctrl+S**
- Save Current Tab DataTables...
- Import Settings From File...
- Export Settings To File...
- Restore Default Settings
- Exit



AntConc

File Edit Settings Help

Target Corpus
 Name: temp
 Files: 8
 Tokens: 48223

Art_01.txt
 Art_02.txt
 Art_03.txt
 Art_04.txt
 Pap_01.txt
 Pap_02.txt
 Pap_03.txt
 Proceed_01.txt

6

KWIC Plot File Cluster N-Gram Collocate Word Keyword Wordcloud

Total Hits: 473 Total Files With Hits: 8

| Row | FileID | FilePath | FileTokens | Freq | NormFreq | Dispersion | Plot |
|-----|--------|----------------|------------|------|-----------|------------|------|
| 1 | 3 | Art_04.txt | 8860 | 40 | 4514.673 | 0.627 | |
| 2 | 2 | Art_03.txt | 8066 | 56 | 6942.723 | 0.678 | |
| 3 | 6 | Pap_03.txt | 2822 | 27 | 9567.682 | 0.678 | |
| 4 | 5 | Pap_02.txt | 5991 | 82 | 13687.197 | 0.762 | |
| 5 | 1 | Art_02.txt | 3658 | 58 | 15855.659 | 0.772 | |
| 6 | 7 | Proceed_01.txt | 5198 | 30 | 5771.451 | 0.783 | |
| 7 | 4 | Pap_01.txt | 5092 | 60 | 11783.189 | 0.795 | |
| 8 | 0 | Art_01.txt | 8536 | 120 | 14058.107 | 0.829 | |

Search Query Words Case Regex Results Set All hits Plot Zoom 1,00 x Overlay Color

arte Start Adv Search

Sort by Dispersion Invert Order

Progress 100%

Time taken (creating plot results): 0.77 sec



AntConc

File Edit Settings Help

Target Corpus
Name: temp
Files: 8
Tokens: 48223

Art_01.txt
Art_02.txt
Art_03.txt
Art_04.txt
Pap_01.txt
Pap_02.txt
Pap_03.txt
Proceed_01.txt

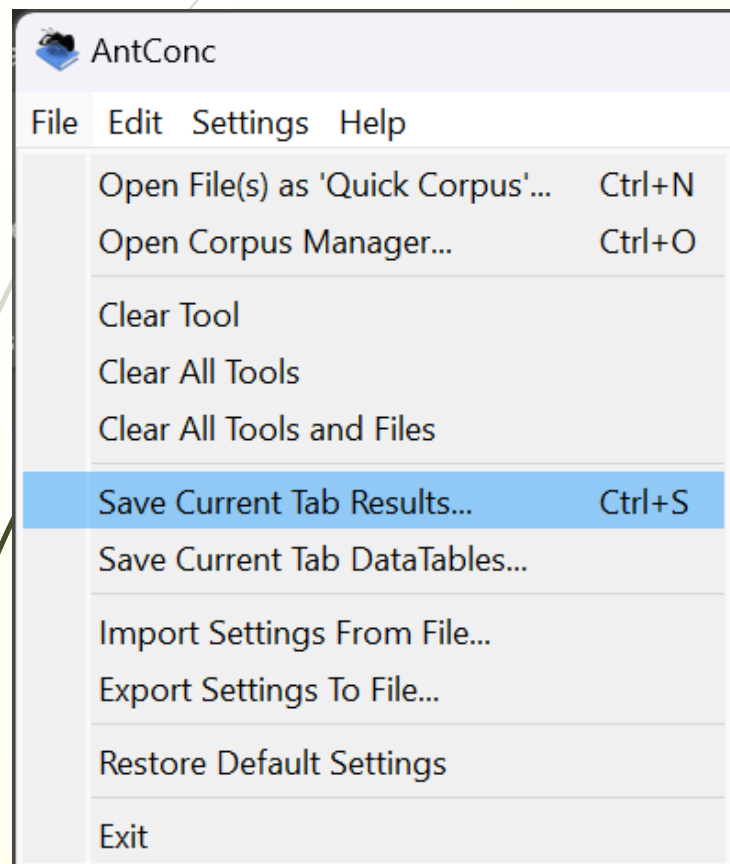
Collocate Types 13 Collocate Tokens 816 Page Size 100 hits 1 to 13 of 13 hits

| | Collocate | Rank | FreqLR | FreqL | FreqR | Range | Likelihood | Effect |
|----|---------------|------|--------|-------|-------|-------|------------|--------|
| 1 | pública | 1 | 291 | 8 | 283 | 8 | 1265.956 | 4.267 |
| 2 | da | 2 | 114 | 108 | 6 | 8 | 112.383 | 1.668 |
| 3 | contemporânea | 3 | 18 | 0 | 18 | 5 | 69.901 | 4.041 |
| 4 | urbana | 4 | 22 | 2 | 20 | 5 | 64.748 | 3.357 |
| 5 | de | 5 | 179 | 163 | 16 | 8 | 63.868 | 0.916 |
| 6 | a | 6 | 119 | 99 | 20 | 8 | 37.845 | 0.870 |
| 7 | obra | 7 | 17 | 17 | 0 | 5 | 32.829 | 2.557 |
| 8 | do | 8 | 6 | 4 | 2 | 4 | 30.451 | -2.358 |
| 9 | história | 9 | 9 | 9 | 0 | 4 | 18.244 | 2.639 |
| 10 | o | 10 | 16 | 2 | 14 | 5 | 17.685 | -1.278 |
| 11 | se | 11 | 8 | 4 | 4 | 4 | 17.118 | -1.688 |
| 12 | obras | 12 | 11 | 11 | 0 | 4 | 15.658 | 2.116 |
| 13 | ideário | 13 | 6 | 6 | 0 | 3 | 14.396 | 2.942 |

Search Query Words Case Reg Window Span From 2L To 2R Min. Freq 1 Min. Range 3

arte

Sort by Likelihood Invert Order



Para salvar os resultados das diferentes listagens (frequências, concordâncias, plot, ... para consultarem mais tarde / anexo):

File > Save Current Tab Results > KWIC_results (ex.)

IMPORTANTE: renomear o ficheiro para ser mais fácil saber a que resultados se referem

Antconc_results_wordlist_corpus_1

Antconc_results_concordância_arte

...



- Fazer *download* do programa AntConc
- Criar um *corpus* (usar uma amostra) > Adicionar o *corpus* ao programa
- Indique o tamanho do *corpus* (tokens/types)

Indique as 10 unidades mais frequentes (excluindo as palavras funcionais)



- Utilizando o separador KWIC extraia os contextos para as primeira 5 unidades lexicais mais frequentes (não funcionais), restrinja a lista de outputs para “10 hits” (ou seja os primeiros 10 resultados).
- Utilizando o separador Collocates extraia as colocações das primeiras 5 unidades lexicais mais frequentes (não funcionais), definindo a janela de coocorrência para 2 elementos à esquerda e 2 elementos à direita; restrinja a lista de outputs para “10 hits” (ou seja os primeiros 10 resultados) e liste os elementos coocorrentes.

ANTCONC / SKETCH ENGINE

... Outros programas

- É importante explorar as definições dos programas
- Ver os vídeos e manuais
- AntConc Guião para ajudar:

<https://www.laurenceanthony.net/software/antconc/releases/AntConc403/help.pdf>



Trabalho de casa

- Submeter até dia 31 de outubro o RESUMO

Avaliação



Para a próxima aula
Dúvidas sobre as pesquisas no
AntConc relativamente aos
trabalhos em curso

Não Esquecer: Preencher e submeter as tarefas indicadas em ATIVIDADES no MOODLE



| | |
|-----------|---|
| EN | The European Commission's support for the production of this publication does not constitute an endorsement of the contents, which reflect the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein. |
| PT | O apoio da Comissão Europeia à produção desta publicação não constitui um aval do seu conteúdo, que reflete unicamente o ponto de vista dos autores, e a Comissão não pode ser considerada responsável por eventuais utilizações que possam ser feitas com as informações nela contidas. |