

MLSS Chunker GUI

1 BASIC INFORMATION

1.1 *Overview and purpose of the tool*

The chunker is a tool which looks for sequences of part-of-speech tags in a tagged text, using one or more patterns that capture the typical members of, say, a noun phrase, and returning a list of chunks from the text that match the searched patterns.

1.2 *Using the chunker*

The GUI link is <http://metanet4u.research.um.edu.mt/Chunker.jsp>.

The GUI features:

- a panel of buttons representing the POS tags, showing the corresponding description on mouse over.
- a text box which allows the user to enter the pattern of the chunks, and
- a text area where the user inputs the text

The user types or pastes text into the large text box on the bottom of the page and then enters a search pattern into the smaller text box above:

The user can specify patterns of:

- POS; these being preceded by '_'. Example, _DDC _NN
- words; string containing no '_'. Example, il- ? (where the '?' represents anything which follows that word)
- tagged words; being the most specific. Example, il_ DDC ?

Instead of typing in the search patterns, the user can also click on the buttons in the panel on the left-hand side of the page.

Afterwards, the user clicks on “Process” to start the chunking process. The user is then directed to a page “Chunker Result”, which shows the tagged and extracted chunks in a text box.

2 TECHNICAL INFORMATION

2.1 *GUI description*

The GUI is available at <http://metanet4u.research.um.edu.mt/Chunker.jsp>

This is the first version of the chunker. Therefore, there may be bugs and room for improvement.

2.2 *Software dependencies and system requirements*

Not applicable.

2.3 *Execution instructions*

See under 1.2.

2.4 *Input/Output data formats*

Input data format: text string in an input text box, search patterns in the search box

Output data format: text sub-strings (i.e. tagged text chunks) of the input string, with one chunk per line

2.5 *Integration with external tools*

The GUI does not need any external tool.

3 CONTENT INFORMATION

A search pattern _NN (_MJ | _VP)

on an input string such as

Wara l-istorja vvintata tal-kandidat laburista Chris Cilia, issa feġġet storja oħra vvintata minn kandidata oħra tal-Malta Labour Party - storja li minn eżerċizzju sempliċi ta' verifika rriżulta li kienet gidba sfaċċata - purament għal skopijiet partiġġjani.

Nhar il-Hadd li għadda, 2 ta' Diċembru, fuq il-programm TVAM fuq l-istazzjon nazzjonali, il-kandidata Laburista Joanne Vella-Cuschieri, li

hi wkoll avukata tal-GWU, allegat li fl-iskola primarja tal-Mellieha qed issir maratona ta' ġbir ta' fondi biex jinxtraw il-bibien. Vella Cuschieri allegat li din l-iskola lanqas għandha flus biex tagħmel il-bibien u qalet li taċ-ċajt li hawn skejjel state-of-the-art meta imbagħad hawn oħrajn li lanqas biss għandhom bibien.

will result in the following output.

```
istorja_NN vvintata_VP
storja_NN oħra_MJ
kandidata_NN oħra_MJ
eżerċizzju_NN sempliċi_MJ
skopijiet_NN partiġġjani_MJ
istazzjon_NN nazzjonali_MJ
kandidata_NN Laburista_MJ
iskola_NN primarja_MJ
```

4 ADMINISTRATIVE INFORMATION

4.1 Contact

For further information and technical support running this tool, please send an email to Andrew Attard:
andrew.attard@um.edu.mt

4.2 Delivery medium (if relevant; description of the content of each piece of medium)

Online GUI at <http://metanet4u.research.um.edu.mt/Chunker.jsp>

4.3 Copyright statement and information on IPR

META-SHARE Commons BY-NC-SA.

5 RELEVANT REFERENCES AND OTHER INFORMATION

<http://metanet4u.research.um.edu.mt/>