

Laws of Malta MT

1 BASIC INFORMATION

1.1 *Corpus composition*

The corpus contains the Laws of Malta in Maltese from the official government website. The unannotated raw text files were extracted from the pdf files that can be found on the website.

1.2 *Representation of the corpora (flat files, database, markup)*

Folder structure is: laws_mt/txt/FILE1.txt, ... FILEn.txt

There are no subfolders in the txt folder.

1.3 *Character encoding*

UTF-8

2 ADMINISTRATIVE INFORMATION

2.1 *Contact person (name, address, affiliation, position, telephone, fax, e-mail)*

Name: Ministry for Justice and Home Affairs

Address: 32/33 House of Catalunya, Marsamxett Road, Valletta VLT1955

Affiliation: <http://www.justiceservices.gov.mt/lom.aspx?pageid=24>

Telephone: (+356) 2295 7000

Fax: (+356) 2295 7348

e-mail: lawdrafting.unit@gov.mt

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

The resource will be uploaded on the MetaShare platform as a folder with text files.

2.3 *Copyright statement and information on IPR*

META-SHARE Commons BY-NC

3 TECHNICAL INFORMATION

3.1 *Directories and files*

A flat directory containing all the text files.

3.2 *Data structure of an entry*

Just text and line breaks taken over from the PDF files.

3.3 *Corpora size (nmb. of tokens, MB occupied on disk)*

69,9 MB on disk

4 CONTENT INFORMATION

4.1 *Type of the corpus (monolingual/multilingual, parallel/comparable, raw/annotated)*

The corpus is monolingual, but it is planned to align it with the respective English language corpus of Maltese legislation to achieve a parallel corpus at a later stage. The text consists of non-annotated raw data.

4.2 *The natural language(s) of the corpus*

Maltese

4.3 *Domain(s)/register(s) of the corpus*

Legalese

4.4 Annotations in the corpus (if an annotated corpus)

The corpus is not annotated.

4.4.1 Types of annotations (paragraph mark-up, sentence mark-up, lexical mark-up, syntactic mark-up, semantic mark-up, discourse mark-up)

N.A.

4.4.2 Tags (if POS/WSD/TIME/discourse/etc –tagged or parsed)

N.A.

4.4.3 Alignment information (if the corpus contain aligned documents: level of alignment, how it was achieved)

N.A.

4.4.4 Attributes and their values (if annotated)

N.A.

4.5 Intended application of the corpus

Machine translation

4.6 Reliability of the annotations (automatically/manually assigned) – if any

N.A.

5 RELEVANT REFERENCES AND OTHER INFORMATION

The law texts were downloaded as pdf files by a web crawler and the text extracted from them.

The website of the Laws of Malta is here: <http://www.justiceservices.gov.mt/lom.aspx?pageid=24>

Laws of Malta EN

1 BASIC INFORMATION

1.1 Corpus composition

The corpus contains the Laws of Malta in English from the official government website.

The unannotated raw text files were extracted from the pdf files that can be found on the website.

1.2 Representation of the corpora (flat files, database, markup)

Folder structure is: laws_en/txt/FILE1.txt, ... FILEn.txt

There are no subfolders in the txt folder.

1.3 Character encoding

UTF-8

2 ADMINISTRATIVE INFORMATION

2.1 Contact person (name, address, affiliation, position, telephone, fax, e-mail)

Name: Ministry for Justice and Home Affairs

Address: 32/33 House of Catalunya, Marsamxett Road, Valletta VLT1955

Affiliation: <http://www.justiceservices.gov.mt/lom.aspx?pageid=24>

Telephone: (+356) 2295 7000
Fax: (+356) 2295 7348
e-mail: lawdrafting.unit@gov.mt

2.2 Delivery medium (if relevant; description of the content of each piece of medium)
The resource will be uploaded on the MetaShare platform as a folder with text files.

2.3 Copyright statement and information on IPR
The resource is freely available for personal and non-commercial use. Government of Malta must be referred to as the source of the original documents.

3 TECHNICAL INFORMATION

3.1 Directories and files
A flat directory containing all the text files

3.2 Data structure of an entry
Just text and line breaks taken over from the PDF files.

3.3 Corpora size (nmb. of tokens, MB occupied on disk)
77 MB on disk

4 CONTENT INFORMATION

4.1 Type of the corpus (monolingual/multilingual, parallel/comparable, raw/annotated)
The corpus is monolingual, but it is planned to align it with the respective Maltese language corpus of Maltese legislation to achieve a parallel corpus at a later stage. The text consists of non-annotated raw data.

4.2 The natural language(s) of the corpus
English

4.3 Domain(s)/register(s) of the corpus
Legalese

4.4 Annotations in the corpus (if an annotated corpus)
There are no annotations in the corpus.

4.4.1 Types of annotations (paragraph mark-up, sentence mark-up, lexical mark-up, syntactic mark-up, semantic mark-up, discourse mark-up)
N.A.

4.4.2 Tags (if POS/WSD/TIME/discourse/etc –tagged or parsed)
N.A.

4.4.3 Alignment information (if the corpus contain aligned documents: level of alignment, how it was achieved)
N.A.

4.4.4 Attributes and their values (if annotated)
N.A.

4.5 Intended application of the corpus
Machine translation

4.6 Reliability of the annotations (automatically/manually assigned) – if any
N.A.

5 RELEVANT REFERENCES AND OTHER INFORMATION

The law texts were downloaded as pdf files by a web crawler and the text extracted from them.

The website of the Laws of Malta is here: <http://www.justiceservices.gov.mt/lom.aspx?pageid=24>