olatool

Table of contents

1 olatool

This is a command-line tool that provides an interface to the Open Linguistic Architecture library.

2 Usage

```
olatool [options] [INPUTFILE]
```

If the INPUTFILE is not specified, it will read from stdin .

3 Options

Command-line options to the olatool :

- --lang=STRING - one or more to specify the language(s) of the input document. Possible values for STRING are:
  - an ISO 639 code (both two or three letter codes accepted), possibly extended with country and variant codes to specify an exact language locale
  - the string mono - it indicates that the document is monolingual; if used alone, the built-in language identifier will be utilised to guess the language; if used together with *one* specific language code (in a separate --lang option), it is superfluous; if used together with multiple language specifications it is meaningless and will be ignored
  - the string multi - it indicates that the document is multilingual; if used alone, the built-in language identifier will be utilised to guess the languages; if used together with *one* specific language code (in a separate --lang option), it triggers the language identifier to try to identify other languages than the given; if used together with multiple language specifications the specified languages take precedence, and it will be ignored.
- --statistics

4 Output

The output depends on the linguistic processing requested, but is always given to stout . The following lists the output for each processing type.

4.1 Speller

The speller gives as output the following:

- optionally a header with some statistical information, see below
- input token
- speller response (correct, incorrect, ...)
- possibly a(n) (empty) list of suggestions

The header has the following format (the given headers are just indicative of what could be put in the header):

```
//Date:
//Time:
//Time elapsed:
//Speller lexicon version:
//...
//Unrecognised words:
//Candidates for lexicalisation:
//...
```

The format of the rest of the speller output data is as follows:

4.2 Hyphenator

dfghjk