

# FileWordCharInfo Project

## Purpose and scope

---

This simple application try to analyze content of a pure text file locally or remotely (http) provided by counting words, spaces, letters and word repetition with a certain level of attention about large files, in fact, it tries to analyze the file while reading it.

## Changelog

---

- 0.1 :
  - Very first version

## Dependencies

---

- Node js version >= 20.0.0
- yarn 1.22

## Installation

---

Just execute yarn to fetch dependency packages,

## Usage

---

Just run **yarn start --file <File path/URL> --config**

### Command line arguments

- -h --help -> display help and exit
- -f --file <File path/URL> -> define source file (required)
- -c --config -> use selected configuration file.

## Test

---

A litte set of unit tests based on **jest** is located in the **\_\_tests\_\_** folder.

To execute tests just run **yarn test**

During tests all logs are displayed even if configuration are differently set.

## Configuration

---

Default configuration file is located at **config/config.json** path.

Configuration example:

```

{
  "global": {
    "minLogLevel": "INFO"
  },
  "wordCharAnalyzer": {
    "minWordRepetition": 10,
    "hyphenChar": "-",
    "removePunctuation": false,
    "wordPattern": ".*[a-zA-Z0-9]+.*"
  }
}

```

- **global** section is exported at global scope.
  - **minLogLevel** -> defines minimum log severity to display: one of 'EMERGENCY', 'ALERT', 'CRITICAL', 'ERROR', 'WARNING', 'NOTICE', 'INFO', 'DEBUG'
- **wordCharAnalyzer** section contains program functionalities related options
  - **minWordRepetitions** -> (required) when result is displayed, only words repeated at least this times value are shown
  - **hyphenChar** -> (optional) if present this value is used to try to reconstruct hyphenated words
  - **removePunctuation** -> (optional) if true all punctuation chars are removed from words
  - **wordPattern** -> (optional) just "define your own word". If present it defines what a word is to avoid counting groups of punctuation chars (for example)