Search String Parser

Release 0.3.0

Oct 04, 2018

Contents

Contents:

CHAPTER 1

Overview

1.1 Search String Parser

docs	
tests	
package	

Parse a more general search syntax to conform with a particular SQL dialect.

Currently, this is implemented using ply with a general lexer and a parser for generating PostgreSQL-specific search queries.

- Free software: BSD license

1.2 Installation

```
pip install searchstringparser
```

1.3 Documentation

https://searchstringparser.readthedocs.org/

1.4 Development

To run the all tests run:

```
tox
```

CHAPTER 2

Installation

At the command line:

```
pip install searchstringparser
```

CHAPTER 3

Usage

To use Search String Parser in a project:

```
import searchstringparser
```

or directly import one of the *lexers* or *parsers*, e.g.,

```
>>> from searchstringparser import GeneralSearchStringLexer
>>> from searchstringparser import PostgreSQLTextSearchParser
```

You can then use an instance of the lexer for your own parser or parse a search query string.

```
>>> parser = PostgreSQLTextSearchParser()
>>> parser.parse("find my term")
'find & my & term'
```

CHAPTER 4

Reference

Top-level package that exposes lexer and parser classes.

4.1 searchstringparser.lexer

class searchstringparser.lexer.general.**GeneralSearchStringLexer**(*illegal='ignore'*, ***kw_args*)

Bases: object

__init__(*illegal='ignore'*, ***kw_args*)

A composite class of a ply.lex.lex.

This is the setup step necessary before you can iterate over the tokens.

Parameters

- **illegal** (*{'record', 'ignore', 'error'} (optional)*) – Action to be taken when illegal characters are encountered. The default is to record them but continue without regarding them.
- **kw_args** – Keyword arguments are passed to the ply.lex.lex call.

get_illegal()

Return encountered illegal characters.

Returns

- *None* – If no illegal characters occurred.
- *Tuple* – A pair of lists that contain the illegal characters and the positions where they occurred.

input(*data*)

Add a new string to the lexer.

This is the setup step necessary before you can iterate over the tokens.

Parameters **data** (*str*) – Any string.

print_tokens(*data*)
Print all tokens in a string.

First iterates through all tokens found and prints them to sys.stdout. Then prints illegal characters if any occurred.

Parameters **data** (*str*) – Any string.

Exposes the following classes:

- *GeneralSearchStringLexer*

4.2 searchstringparser.parser

class searchstringparser.parser.postgresql.**PostgreSQLTextSearchParser**(*lexer=None*, ***kw_args*)

Bases: object

This parser implements the following rules using the tokens generated by an appropriate lexer. The goal is to generate a string for PostgreSQL full text search that conforms with the syntax understood by the function tsquery or to_tsquery.

The following rules are implemented which generate the correct query string.

```
expression : expression expression
           | expression AND expression
           | expression OR expression
           | NOT expression
           | LPAREN expression RPAREN
           | QUOTE term QUOTE
           | WORD WILDCARD
           | WORD

term : term SPACE term
       | term term
       | LITERAL_QUOTE
       | SYMBOL
```

__init__(*lexer=None*, ***kw_args*)
Parser instantiation.

Parameters

- **lexer** (*ply.lex (optional)*) – Any ply.lex lexer instance and generates the tokens listed in the rules. The default uses a GeneralSearchStringLexer instance.
- **kw_args** – Keyword arguments are passed to the ply.yacc.yacc call.

get_illegal()
Inspect encountered illegal characters.

Returns

- *None* – If no illegal characters occurred.
- *Tuple* – A pair of lists that contain the illegal characters and the positions where they occurred.

parse(*query*, ***kw_args*)
Parse any string input according to the rules.

Parameters

- **query** (*str*) – A string expected to conform with the search query syntax.
- **kw_args** – Keyword arguments are passed on to the ply.yacc.yacc.parse call.

Returns

A string that can directly be passed to the PostgreSQL functions tsquery or to_tsquery.

Return type str

Exposes the following classes:

- *PostgreSQLTextSearchParser*

CHAPTER 5

Contributing

Contributions are welcome, and they are greatly appreciated! Every little bit helps, and credit will always be given.

You can contribute in many ways:

5.1 Types of Contributions

5.1.1 Report Bugs

Report bugs at https://github.com/AGHerwig/searchstringparser/issues.

If you are reporting a bug, please follow the presented issue template since it is designed to ultimately make helping you easier and thus faster.

5.1.2 Fix Bugs

Look through the GitHub issues for bugs. Anything tagged with "bug" and "help wanted" is open to whoever wants to implement it.

5.1.3 Implement Features

Look through the GitHub issues for features. Anything tagged with "enhancement" and "help wanted" is open to whoever wants to implement it.

5.1.4 Write Documentation

As any open source project, searchstringparser could always use more and better documentation, whether as part of the official docs, in docstrings, or even on the web in blog posts, articles.

5.1.5 Submit Feedback

The best way to send feedback is to file an issue at https://github.com/AGHerwig/searchstringparser/issues.

If you are proposing a feature:

- Explain in detail how it would work.
- Keep the scope as narrow as possible, to make it easier to implement.
- Remember that this is a volunteer-driven project, and that contributions are welcome ;)

5.1.6 Get Started!

Ready to contribute? Here's how to set up `searchstringparser` for local development.

1. Fork the `searchstringparser` repo on GitHub.
2. Clone your fork locally:

```
git clone git@github.com:<your_name_here>/searchstringparser.git
```

3. Install your local copy into a virtualenv. Assuming you have virtualenvwrapper installed, this is how you set up your fork for local development:

```
mkvirtualenv searchstringparser
cd searchstringparser/
pip install -e .
```

4. Create a branch for local development:

```
git checkout -b name-of-your-bugfix-or-feature
```

   Now you can make your changes locally.

5. When you're done making changes, check that your changes pass the quality control:

```
tox
```

To get tox, just pip install it into your virtualenv.

6. Commit your changes using semantic commit messages and push your branch to GitHub:

```
git add .
git commit -m "feat: your detailed description of your changes"
git push origin name-of-your-bugfix-or-feature
```

7. Submit a pull request to this repository through the GitHub website.

CHAPTER 6

Authors

- Moritz Emanuel Beber - https://github.com/AGHerwig/searchstringparser

CHAPTER 7

Changelog

7.1 0.3.0 (2018-10-04)

- Fix a bug in the token parser.
- Update the infrastructure.

7.2 0.2.3 (2015-09-29)

- Publish on PyPi

7.3 0.2.2 (2015-09-29)

- Complete documentation

7.4 0.2.1 (2015-09-28)

- Add more helpful error messages

7.5 0.2.0 (2015-09-28)

- Add complete integration with Travis, Appveyor, Coveralls, and Read the Docs
- Increase test coverage to 100%
- Make source compatible with Python 2.6 - 3.4

7.6 0.1.1 (2015-09-21)

- Fix GeneralSearchStringLexer regex
- Fix PostgreSQLTextSearchParser rules
- Add first set of tests

7.7 0.1.0 (2015-09-16)

- Initial class layout

CHAPTER 8

Indices and tables

- genindex
- modindex
- search

Python Module Index

S

Index

Symbols

G

I

P

S