LECTURE 31

MORE ON REGULAR EXPRESSIONS

SOFTWARE LICENSING

MCS 260 Fall 2021
Emily Dumas

REMINDERS

- Project 3 due Friday at 6pm
- Suggested schedule: Make first submission to the autograder today so you have time to revise.

REGEX QUICK REFERENCE

- `.` — matches any character except newline
- `\s` — matches any whitespace character
- `\d` — matches a decimal digit
- `\w` — matches any "word character"
- `+` — previous item must repeat 1 or more times
- `*` — previous item must repeat 0 or more times
- `?` — previous item must repeat 0 or 1 times
- `{n}` — previous item must appear n times
- `(...)` — treat part of a pattern as a unit and capture as group
- `[...]` — match any one of a set of characters
- `A|B` — match either pattern `A` or pattern `B`.
- `^` — match the beginning of the string.
- `$` — match the end of the string or the end of the line.

RE MODULE QUICK REFERENCE

- **`re.search(pattern,text)`** — does `text` contain a match to the `pattern`? Return a match object or `None`.

- **`re.finditer(pattern,text)`** — return an iterable yielding all the non-overlapping matches as match objects.

- **`re.sub(pattern,replacement,text)`** — return `text` but with each match of `patttern` replaced by `replacement`.

EXAMPLE PROBLEM

Find all of the phone numbers in a string that are written in the format `319-555-1012`, and split each one into area code (e.g. `319`), exchange (e.g. `555`), and line number (e.g. `1012`).

SQUARE BRACKETS

Give a list of characters and to match any one of them.

`[abc]` matches any of the characters `a`,`b`,`c`.

`[^abc]` matches any character *except* `a`,`b`,`c`.

Supports dashed ranges, too.

`[A-Za-z]` matches any alphabet letter.

`[0-9a-fA-F]` matches any hex digit.

OR

`A|B` matches either pattern `A` or pattern `B`.

Use this inside parentheses to limit how much of the pattern is considered to be part of `A` or `B`, e.g.

```
[Hh](ello|i),? my name is (.*).
```

WARNING

The rest of this lecture talks about laws in the USA, but it is not legal advice. I am not a lawyer.

COPYRIGHT

Copyright is a set of protections granted to authors of "original works".

Software is protected by copyright. The creators of the software are considered the authors.

Key point: Software you write is automatically and immediately protected by copyright.

Copyright provides the authors the exclusive right to:

- Reproduce the work
- Make derivative works
- Sell copies of the work
- Authorize others to do things that would otherwise be prohibited by these protections
- Transfer ownership of copyright to another person or entity

Copyright eventually expires (currently 70 years after death of author).

If you find some code on the internet and there is no accompanying information that grants permission to

- Distribute the code
- Modify or use the code in your own program

then those things are typically prohibited.

SOFTWARE LICENSES

A software license is a document that grants a person or group permission to use a piece of software in certain ways.

Usually it permits certain actions that would otherwise be prohibited by copyright protections.

Whenever you find a program or bit of code on the internet, look for a license!

LICENSING EXAMPLE

Suppose I write an autograder program for use in Python teaching.

I might license the code for other instructors to use, with the condition it not be modified or used commercially.

For a fee, I might also license it to a company to modify and sell as a commercial product.

OPEN SOURCE

An important class of software license is an **open source license**, which grants anyone permission to:

- See the source code[*]
- Distribute the software and source code
- Make derivative works

Software that is not open source is **proprietary**.

[*] *Source code* means the text written in a computer programming language that was used to create the program. In Python, that's usually the same as the program itself.

There isn't universal agreement about the definition of "open source", but the definition from the Open Source Initiative is often used.

SOME POPULAR LICENSES

- **Public domain declaration** - Declares that the copyright owner waives all exclusive rights afforded by copyright. Most permissive license possible.

- **MIT License** - Very permissive. Only requires that a statement about the copyright ownership be included in all derivative works. Derivative works can have different licenses (e.g. may be proprietary).

- **GNU General Public License (GPL)** - More restrictive than MIT; distribution of a derived work is only allowed if no additional restrictions are applied. In particular, every derived work must be open source.

EXAMPLES

- The Python interpreter is open source. Its license is less restrictive than GPL.

- Linux is open source, licensed under the GPL.

- Microsoft Windows is proprietary.

REFERENCES

- CC0 from Creative Commons is an example of a public domain declaration.
- MIT license (the full text)
- The GNU GPL comes in several versions with different restrictions; details and full text: https://www.gnu.org/licenses/licenses.en.html
- Introduction to Copyright Law (short course from MIT Open Courseware)

REVISION HISTORY

- 2021-11-03 Initial publication