

Welcome to COMP2041!

Software Construction: Techniques and Tools

Introductions

Course Resources

- Course website
 - <https://cgi.cse.unsw.edu.au/~cs2041/25T1>
- Course structure
 - Lectures, tutorials, labs, tests, assignments, exam
- Help
 - Forum: <https://discourse01.cse.unsw.edu.au/25T1/COMP2041/>
 - Emails
 - z5419252@unsw.edu.au
 - cs2041@cse.unsw.edu.au
 - Labs

Regular Expressions (regex)

- String matching!

- Unless a character has a special meaning it matches itself
 - e.g. **a** has no special meaning so it matches **a**.
- **p*** denotes zero or more repetitions of **p**.
 - e.g. **b*** matches the empty string and: **b**, **bb**, **bbb**, **bbbb** ...
 - note this is an infinite set of strings
- **pattern₁ | pattern₂** denotes the union of **pattern₁** and **pattern₂**.
 - e.g. **perl|python|ruby** matches any of: **perl**, **python** or **ruby**
 - | is sometimes called alternation
- Parentheses are used for grouping
 - e.g. **c(,c)*** matches: **c**, **c**, **c**, **c**, **c**, **c**, **c**, **c**, **c** ...
 - and **(d|e)*(f|g)** matches **f**, **g**, **df**, **dg**, **ef**, **eg**, **ddf**, **deg**, **edf**, **edg**, **eef**, ...
- backslash **** removes any special meaning of the following character
 - e.g. ***** matches an ***** instead of indicating repetition
- Any regular expression can be written using only **()*|**
 - but many syntax features are present for convenience & clarity.

- **.** (dot) matches any single character.
- Square brackets provide convenient matching of any **one** of a set of characters.
- **[listOfCharacters]** matches any **single** character from **listOfCharacters**.
 - e.g. **[aeiouAEIOU]** matches any (English) vowel.
- A shorthand is available for ranges of characters **[first – last]**
 - e.g. **[a-e]** **[a-z]** **[0-9]** **[a-zA-Z]** **[A-Za-z]** **[a-zA-Z0-9]**
- Square brackets matching can be inverted with an **^**
- **[^listOfCharacters]** matches any **single** character except those in **listOfCharacters**.
 - e.g. **[^a-e]** matches any character *except* one of the first five lowercase letters
- Other characters lose their special meaning inside bracket expressions.
- e.g. **[^X]*X** matches any characters up to and including the first X

grep

- Command that searches text for lines that match a regular expression