

Seng 265-Lab 08

szehtabi@uvic.ca

Regular expressions

- A language-independent approach for expressing string patterns
- Used in lexical analyzers for compilers
- Used for input validation
- Used for parsing xml files, html files, etc

Metasymbols

- dot (.)
- *, +, ?
- a{5}, a{3,7}
- [abc], [^abc]
- a|b
- ()
- \
- \d, \D
- \w, \W
- \s, \S
- \b, \B
- ^, \$

re module

- `re.match()`
 - returns a match object
- `re.search()`
 - returns a match object
- match object
 - `groups`
 - `start`
 - `end`
- `re.findall()`
 - returns a list of matched strings
- raw string literal

Exercise

1. Write a python script that finds the index of the first occurrence as well as the total number of occurrences of the word 'hobbit' in file 'the-hobbit.in'
2. Write a python script to extract id, name, email addresses and phone numbers from the file 'data.txt'

Do not use split unless justified!