Unit 9 · Automate Your Work

WORKSHEET

Full marks 32

Name _____ Class _____ Roll no. ____ Date _____ Marks _____ / 32

1 Key words [4]

Fill in the blanks with words from the box. Two words in the box are not needed. Each blank carries 0.5 mark.

API	**automation**	**byte order mark**	**CSV**	**dry run**	**pixel**	**raw string**	**regular expression**	
robots.txt	**script**							

Making a computer do a job that a person used to do, step by step, is called _____, and a small program written to do such a job is a _____. Running a script so that it only shows what it would do, without doing it, is a _____. A spreadsheet saved as one row per line, with the values separated by commas, is a _____ file; Excel's CSV UTF-8 starts it with a _____. A description of what the text you want looks like, such as `9[678]\d{8}`, is a _____, best written in Python as a _____, with an r before the quotes. A website's _____ file asks programs not to visit some of its addresses.

2 The pieces of a pattern [3]

Match each piece of a regular expression, in Column A, with what it matches, in Column B. Two items in Column B are not needed. Each carries 0.5 mark.

Column A		Column B
1. `\d`	☐	**a.** any one character that is not a comma
2. `\s`	☐	**b.** the start of the text
3. `+`	☐	**c.** any digit
4. `?`	☐	**d.** the edge of a word
5. `[^,]`	☐	**e.** the thing before it, once or more
6. `\b`	☐	**f.** any one character but a new line
		g. the thing before it, once or not at all
		h. a space, tab or new line

3 Multiple choice [5]

Choose the correct option for each of the following questions.

i. Which pattern matches the whole of **Rs 2,500** and **Rs 75**, but not **Room 75**? [1]

a. `R\w+ [\d,]+`
b. `Rs [\d,]+`
c. `Rs \d+`
d. `[\d,]+`

ii. What does `re.findall(r"\d+", "Grade 9, roll ९")` give? **[1]**

a. `['9']` **b.** `['9', '९']`

c. `['Grade 9', 'roll ९']` **d.** `[]`

iii. Which text does `re.fullmatch(r"\d{3}-\d{4}", text)` match? **[1]**

a. 984-1000011 **b.** 661-2345

c. 6612345 **d.** Tel 661-2345

iv. A website's robots.txt says `Disallow: /results`. What does that mean for your program? **[1]**

a. The results pages are locked: no one can open them.

b. The site asks programs not to visit addresses starting with /results, and a polite program doesn't.

c. Your program may visit them once a day.

d. Only browsers may not open them.

v. A program reads a CSV saved from Excel, and `row["name"]` raises `KeyError: 'name'`, though printing the row shows a column called name. What is the most likely fix? **[1]**

a. Rename the column to Name.

b. Open the file with `encoding="utf-8-sig"`.

c. Save the file again with no names in the first row.

d. Use `row[0]` instead.

4 What does it print? [3]

Write exactly what the program prints, in the box under it.

```
1  import re
2
3  text = "Bus 12 at 7:45, fare Rs 35. Call 9841000011."
4  print(re.findall(r"\d+", text))
5  print(re.findall(r"Rs \d+", text))
6  print(re.sub(r"\d", "#", "Room 204"))
7  print(re.fullmatch(r"9[678]\d{8}", "9541000011"))
```

```
OUTPUT
```

5 Short answer [2]

Answer the following question.

Why is `string.Template` used for the certificates instead of `str.format()`? Why must each value be passed through `html.escape()` first?

6 Trace the program [3]

Fill in the trace table: one row for each phone number.

```
1  def clean_phone(text):
2      digits = "".join(c for c in text if c.isdigit())
3      if len(digits) == 13 and digits.startswith("977"):
4          digits = digits[3:]
5      if len(digits) == 10 and digits.startswith("9"):
6          return digits
7      return None
8
9  for text in ["+977-9801000044", "98-51000055", "984100006"]:
10     print(text, clean_phone(text))
```

`text`	`digits` after line 2	`digits` after line 4	Output
`+977-9801000044`			
`98-51000055`			
`984100006`			

7 Find and fix the bug [3]

Read the program, then answer.

The folder **Scans** holds three files: `fee receipt 2083-07-08.pdf`, `lab bill 2083-07-08.pdf` and `notice 2083-07-09.PDF`. This script names each by its date. It runs on Linux with no error.

```
1  from pathlib import Path
2
3  folder = Path("Scans")
4  for item in sorted(folder.iterdir()):
5      date = item.stem[-10:]                # the last 10 characters: the date
6      new = folder / (date + item.suffix.lower())
7      print(item.name, "->", new.name)
8      item.rename(new)
```

i. Which files are in the folder afterwards, and what has gone wrong?

ii. Write the lines that fix it, and say where they go.

iii. What should have been done before running it on the real folder?

8 Is it worth a script? [3]

Work it out in the space below.

A school clerk spends 30 minutes every week joining the sections' library lists. A script to do it would take 3 hours to write and test, and then 4 minutes a week to run and check. (i) In which week has the script paid for itself? (ii) How much time does it save in a school year of 40 weeks?

Answer: _____

9 Write a program [4]

Write the program in Python on the lines. Any program that gives the right result gets the marks.

The trip fee is Rs 1,500. A sheet saved from Excel as **fees.csv** has the columns `name` and `paid`. Write a program that reads it with the csv module and prints each student who has paid less, with how much they still owe, such as `Dawa Sherpa owes Rs 750`.

10 Think [2]

Think, then answer.

A program cleans phone numbers with `re.sub(r"\D", "", text)` and keeps the result if it has 10 digits. What happens to the text "Call 98410 00011 or 98510 00022"?

11 Activity: Test a pattern (not marked)

At a computer, write a pattern for a date such as 2083-07-08. First write texts that should and should not match; then test each with `re.fullmatch`.

Text	Should it match?	Did it?