WORKSHEET

F8.6 Characters: ASCII and Unicode

Data representation · GCSE · OCR J277 1.2.4, AQA 8525 3.3.5, Edexcel 1CP2 2.2.1 · about 12 min

Name _____ Class _____ Date _____

What this lesson is about

Character sets, bits per character, and why Unicode was needed.

- ASCII
- The trouble with ASCII
- Unicode
- Storage: more characters cost more bits
- Working out text size

Questions

5 marks in all

1. How many characters can 7-bit ASCII represent? [1 mark]

2. Why was Unicode needed? [1 mark]

- ☐ **A** ASCII has no room for the world's other writing systems and symbols
- ☐ **B** ASCII was too slow
- ☐ **C** Unicode uses fewer bits
- ☐ **D** ASCII cannot store digits

3. ord('A') is 65. What is ord('D')? [1 mark]

4. A 12-character message is stored in 7-bit ASCII. How many bits is it? [1 mark]

5. What does this program print? [1 mark]

```
print(len("robot".encode("utf-8")), len("机器人".encode("utf-8")))
```

The task: message sizes

For the message `BugBot says hi`, print three lines: `characters: <n>`, `ascii bits: <n>` for 7-bit ASCII, and `utf-8 bytes: <n>`. Then print the 7-bit binary code of each of its first three characters, one per line, such as `B 1000010`. Work everything out from the message.

```
# the two lines every program starts with: the commands, then the robot
from bugbot import *
connect()

message = "BugBot says hi"
```

Plan your program here, then type it in and press Run.

Do it on the robot

www.bugbotlab.com/learn/f8-6-characters-ascii-and-unicode/

The simulator checks it and tells you when it passes. Nothing to install, no account.

Challenges

1. Find the smallest code for a lower-case letter, and explain how to turn any capital into lower case with arithmetic.
2. Send the robot's name as a string of 7-bit binary codes, and decode it again.
3. Which uses fewer bytes in UTF-8: a tweet in English or the same length of emoji? Measure it.