

BBC News Labs – Coding Exercise

Background

Within the BBC News Labs team, we work with multiple teams across BBC News including BBC World Service to help improve their workflows.

The BBC re-versions content in multiple languages for broadcast across the world. This production process often involves the creation and handling of scripts containing a mix of languages. One specific example of this would be a script translated from English to allow a voice-over artist to create a correctly timed voice-over for a programme broadcast on one of our World Service TV channels.

Programme Script file

In this exercise, we provide you with a JSON file containing a real-world example of one of these scripts, in this case Click from early January. This programme was originally broadcast in English¹, but a version was also created in Persian². We need to use this as data for one of our prototypes.

The supplied excerpt of script is split into sections containing English text from the original broadcast, and the corresponding Persian that was recorded as a voice-over, along with some other notes that would help the voice-over artist.

We also provide an example conversion into JSON as input into the API, which you are free to use.

Files included:

- `farsi-english-script.rtf`
- `farsi-english-script.json`

Exercise

For this exercise, we'd like to ask you

- to design an API that
 - runs on **localhost** port **8080**
 - uses JSON data as input
 - extracts and separates out the metadata, Farsi, and the English lines of text from the data.
- Include some brief documentation that would help someone integrate with your service.
- Write the API in JavaScript, TypeScript, or Python.
- Please add additional fields in the response of the API (other than meta, english and additional languages) in the output that you think will be useful.

¹ https://www.youtube.com/watch?v=Gvh_Fkq9VJw

² <https://www.youtube.com/watch?v=EtmoQH4ITCo>

For an example (from another language) if a sample input data were formatted as:

```
["OOV:", "Some text", "何か", "OOV:", "more text", "まだまだ"]
```

we would like the output to be:

```
{
  "script": [
    {
      "meta": ["OOV"],
      "japanese": "何か",
      "english": "Some text"
    },
    {
      "meta": ["OOV"],
      "japanese": "まだまだ",
      "english": "more text"
    }
  ]
}
```

Things we'd like you to consider:

- Extensibility of the API
- Data ingestion and output
- How to deploy it, host it, and maintain it

We suggest that you should aim to spend no more than 2-3 hours on the exercise.

Please submit via email to coding-exercises@bbcnewslabs.co.uk by Friday, 11 June 2021, a link to a Github repository or a zip file containing

- Your code
- A README file containing instructions to run your code
- Your API documentation
- Any other notes to consider

During the interview, we'd like to spend a few minutes discussing your solution and the approach you've taken in a little more detail.

If you have any questions or need any clarifications relating to this exercise, please email coding-exercises@bbcnewslabs.co.uk.