

# News Labs – Senior Software Engineer – Coding Exercise

## Background

The BBC produces 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. Part of our language technology research work aims to help improve such workflows.

## Programme Script File



In this exercise, we provide you with a file containing a real-world example of one of these scripts, in this case Click from early January. This programme was originally broadcast in English<sup>1</sup>, but a version was also created in Persian<sup>2</sup>. We need to use the text 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.

File:

- farsi-english-script.rtf

## Exercise

For this exercise, we'd like to ask you to:

- **Write some code to extract and separate out the Persian and the English lines of text from this script.**
- **Turn your code into a simple API.** While the example text is supplied here as an RTF file for convenience, the input to your API does not need to take a document of this format. The input and output formats are up to you.
- **Write brief documentation that would help someone integrate with your service.**

---

<sup>1</sup> [https://www.youtube.com/watch?v=Gvh\\_Fkg9VJw](https://www.youtube.com/watch?v=Gvh_Fkg9VJw)

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

Things we'd like you to consider:

- Alternative approaches for extracting the text for different languages from the script.
- The format and overall shape of your API.
- Appropriate input and output formats for your API.

Use a programming language and an approach with which you feel most comfortable.

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

Please submit, via an email to `coding-exercises@bbcnewslabs.co.uk`, *no later than Wednesday 6<sup>th</sup> March*, a link to a github repository containing:

- Your code.
- A README file containing instructions to run your code.
- Your API documentation.

Your code submission will be *reviewed anonymously*, so please don't include any identifying information within the code.

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`.