

Home Assignment

You are given a collection of 30K html files containing tables, +30K metadata files. Those tables hold generated data. For each table (html file) there is a matching metadata (json file). The metadata holds a dictionary format, with 3 main sections:

1. Header

2. Body

3. Footer

For example, this is 0_table.html:

Table 26.34.97.17.47.27 Public affairs consultant		
John Garcia Lauren Duke Gary Krause		
George Group	1144 986	860
Ruiz Ltd	883% 1444	1124
Tucker, Coleman and Haney	56 989%	452%
Taylor, Johnson and Williams	265 1489%	575
Strong Ltd	1109 438%	912%
Simon-Robinson	28 425%	1065
Clarke Group	1271 68%	252
Creation: 22Mar2017 India		

The diagram shows three boxes on the right with arrows pointing to the table: 'Header' points to the first row, 'Body' points to the middle rows, and 'Footer' points to the last row.

We would like you to train a model that get an html file as an input and generate a corresponding json file which holds the correct metadata.

Please try to present your thought process through code with Readme.md on a github repository.

You can use any library you would like, such as: **Sklearn, PyTorch, Tesnorflow etc'**. The model will be tested on our newly generated data.

In the next interview, we will go over and discuss the assignment.
For any question you have, feel free to approach either:

Gilad at 052-8310326
(Gilad@beaconcure.com)

Or

Gil at 052-6718195
(Gil@beaconcure.com)

Best of luck !