

Dear candidate,

We appreciate your interest in joining the Data Science team @Conrad and your participation in our recruitment process. Following is the case study for the position.

Imagine that we would like to deliver a new feature for enhancing our sellers' journey when onboarding new products on our Conrad Marketplace.

For that, we would like to provide a service that automatically detects particular specifications of added products from the related unstructured data it has e.g. title, description, ...

A sample of the available data is shown in the screenshot below.

Note: A data file called ds_ner_test_case.csv is shared with the task details

description	productHighlights	headline	Brand	Speicherkapazität	Farbe
· Kompakte, optische 7-Tasten-Funkmaus mit ...	Kompakte Funkmaus mit optimaler Oberflächenabtastung und präz...	Hama Knallbunt 2.0 Rot	Hama		Rot
· Kompakte, optische 7-Tasten-Funkmaus mit ...	Kompakte Funkmaus mit optimaler Oberflächenabtastung und präz...	Hama Knallbunt 2.0 Gelb	Hama		Gelb
Unsere innovativen WD Black-Laufwerke brin...	2 TB Speicherkapazität 7200 U/min und 64MB Cache S...	Western Digital 2TB SATA 3,5" WD20...	Western Dig...	2 TB	
Unsere innovativen WD Black-Laufwerke brin...	1 TB Speicherkapazität 7200 U/min und 64 MB Cache S...	Western Digital 1TB SATA 3,5" WD10...	Western Dig...	1 TB	
Unsere innovativen WD Black-Laufwerke brin...	4 TB Speicherkapazität 7200 U/min und 64MB Cache S...	Western Digital 4TB SATA 3,5" WD40...	Western Dig...	4 TB	
Unsere innovativen WD Black-Laufwerke brin...	3 TB Speicherkapazität 7200 U/min und 64MB Cache S...	Western Digital 3TB SATA 3,5" WD30...	Western Dig...	3 TB	
 Nutzen Sie das volle Potenzial von Thun...	512 GB Speicherkapazität Zwei SSDs und zwei Thunderbolt...	Buffalo DriveStation Mini Thunderbolt ...	Buffalo	512 GB	
 Nutzen Sie das volle Potenzial von Thun...	1 TB Speicherkapazität Zwei SSDs und zwei Thunderbolt™-A...	Buffalo DriveStation Mini Thunderbolt ...	Buffalo	1 TB	
Die Desktop-Festplatten WD Elements™ mit ...	2 TB Speicherkapazität Blitzschneller Datentransfer mit USB ...	WD Elements 2TB USB 3.0	WD	2 TB	
Die Desktop-Festplatten WD Elements™ mit ...	3 TB Speicherkapazität Blitzschneller Datentransfer mit USB ...	WD Elements 3TB USB 3.0	WD	3 TB	
Das HUAWEI Ascend Y300 bietet neben dem...	Brillanter 10,2 cm (4") IPS Touchscreen 1,0 GHz Dual-Core Pr...	Huawei Ascend Y300 im Tarif Congsta...	HUAWEI	4 GB	
 Klein ganz großartig. Es liegt k...	10,16 cm (4") Super-AMOLED Touchscreen 5 Megapixel Ka...	Samsung Galaxy S 3 Mini i8190 im Ta...	Samsung	8 GB	Blau

Tasks:

1. How do you think that the provided solution could enhance the sellers' experience in onboarding new products into our Marketplace?
2. If needed, transform the provided dataset into the needed format for training a Named Entity recognition model
3. Train and evaluate a model of your choice for tagging provided products' details with **recognized brands, storage capacity (Speicherkapazität), and colors (Farbe)**.
4. The shared data was prepared using a naive string matching algorithm for extracting the entities using partially available products' technical attributes. How do you think the data extraction and preparation could be improved?
5. Provide a Dockerized version of the implemented solution
6. Implement unit test(s) whenever that is needed

Optional:

7. Implement a simple Rest API with an endpoint getting as a parameter a product title and returning the tagged text
8. Implement an integration test for the implemented API route

Note: the provided data should not be made publicly available

Make sure to push the code to a **PRIVATE** Github repository with the needed instructions to run the solution code. Then, add the user **dstest22** as a collaborator to the project and send us an email with the link to the repo. If you have any questions feel free to send an email to saloua.litayem@conrad.de

We are looking forward to getting your solution (Github link) and discussing it.
Good luck.

Best regards,
Conrad Data Science Team