Internship / Student Trainee in Linguistics: Data Collection/Annotation

Top Features: IMPACT, FREEDOM, DIVERSITY

We are:

… a prospering tech company developing cutting-edge algorithms and user interfaces for **full-text analyses**. Our cloud software PlagScan (SaaS) is on a mission to set the **universal standard for plagiarism checking**, in order to enable a fair and objective examination/valuation of scientific and commercial text content.
With a team of fifteen people, being easy going but downright ambitious, we are working on a product which already today helps more than **100.000 users every week** decide: Original or copy!
Very recently we joined forces with a sister company in Sweden and are combining our knowledge for the next level, re-releasing our service as 'Ouriginal' by the start of 2021.

You are:

… a technically savvy student looking for a rewarding and challenging **internship** which exposes you to modern methods of **computational linguistics** such as data collection, annotation, corpus creation, and (for those with previous exposure) implementation and evaluation of **Deep Learning** algorithms. You either have knowledge of **Spanish** (at least at the B2 level) and are comfortable working with textual editors/word processors and performing **web searches** with engines such as Google or Bing in order to find text in online sources.

Exceptional candidates will have some exposure to **computational linguistics, natural language processing** and **corpus linguistics.** This however is not mandatory.

Job Description:

PlagScan is processing huge amounts of text. On the one hand, users send us **thousands of documents** for plagiarism checking. On the other hand, we cover research papers and online content comprehensively to deliver the **best, most accurate results.** Currently, we are prototyping Deep Learning algorithms which help to identify plagiarism. Accordingly, we offer the following internship project topics:

We need to put Spanish textual data into a machine-readable format to use for training of ML algorithms. This task also involves uses search engines to obtain original sources and constructing ML datasets to be used for training and testing of algorithms.

Internship / Student Trainee – Deep Learning & Linguistic Data Collection/Annotation

Top Features: IMPACT, FREEDOM, DIVERSITY

As an intern / student trainee, you don't have to know everything but must be super motivated to learn everything:

- ∉ A student with linguistics background or similar
- ∉ Independent, committed and reliable work attitude
- ∉ Good English (you will be part of an international team) and Spanish (for the specific task)

Basic parameters

Whether you are joining us as an intern or as a student trainee, these are the respective parameters:

	Internship	Student Trainee
Start date	flexible	flexible
Payment	Competitive monthly salary	Competitive hourly salary
Benefits	train-ticket, food'n'drinks, events, etc.	train-ticket, food'n'drinks, team events, etc.
Duration	3-6 months	6-24 months
Work Time	Full-time in Cologne	10-20h per week

We are looking forward to reading your application!
Simply send an email to jobs@plagscan.com

Follow us
/ouriginal
www.ouriginal.com