Unlock your potential
Ouriginal

Internship / Student Trainee – Deep Learning & Linguistic 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. And/or you are able to program in **Python** and have taken (or are taking) a course in **Machine Learning (ML), Deep Learning (DL),** or **Natural Language Processing (NLP).**
Exceptional candidates will have **programming** and **scripting** (Python, bash) experience and previous exposure to **Machine Learning(ML)/Natural Language Processing(NLP)**.

Job Description: Linguistic Data Collection/Annotation

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:

- **Thematic focus A: Corpus creation/linguistic annotation**

  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.

- **Thematic focus B: Deep Learning**

  We are prototyping textual similarity algorithms to help identify different types of plagiarism. This will include using toolkits such as spaCy or NLTK, as well as ML frameworks such as Tensorflow and Pytorch to write code and evaluating performance using standard metrics such precision, recall, f-score.

PlagScan GmbH
Subbelrather Stasse 15
50823 Cologne, Germany
Management Board:
Markus Goldbach, Johannes Knabe

Tel: +49 (0) 1522 990 8956
Fax: +49 (0) 221 75988996
Registered No.: AG Köln,
HRB 73381
VAT ID: DE267078551

Follow us
/ouriginal
www.ouriginal.com

Unlock your potential
Ouriginal

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 Computer Science / Computational Linguistics background or similar
- Independent, committed and reliable work attitude
- English (you will be part of an international team in Cologne and across the Baltic Sea to Sweden)
- First project experience with Java/Python development
- System knowledge (for example Debian/Linux, Bash, SQL)

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

PlagScan GmbH
Subbelrather Stasse 15
50823 Cologne, Germany
Management Board:
Markus Goldbach, Johannes Knabe

Tel: +49 (0) 1522 990 8956
Fax: +49 (0) 221 75988996
Registered No.: AG Köln,
HRB 73381
VAT ID: DE267078551