

Identification and Sanitization of Sensitive Information using AI/ML

Opportunity for a 6-month internship

Security Research @ SAP Labs France
Sophia-Antipolis – France

Maintaining security is a constantly shifting task, and we need to respond with continuous learning and research. The portfolio of SAP Security Research contains those topics that we believe are most important for SAP's security future.

SAP's vision to secure business is built on 3 ideals: **Zero-Vulnerability**, to harden the software by eliminating vulnerabilities, **Defensible Application**, to enable the software to identify and prevent attacks, and **Zero-Knowledge**, to make any theft of data useless through encryption.

Considering these aspects, SAP Security Research covers the following focal areas: Anonymization for Big Data, Secure Internet of Things, Software security analysis, Open-source analysis, Deceptive application, Applied cryptography, Quantum technology, and Machine Learning as enabler for the next generation of security.

[Security Research](#) proposes a 6-month internship in its Sophia-Antipolis offices (Mougins, France).

INTERNSHIP TOPIC

Information from social media, IoT devices, network logs is produced in vast amount. The processing of such information, especially when it comes in semi-structured or unstructured forms, requires attention and resources. In domains like cyber security and data protection&privacy, special requirements apply to different pieces of information, and the same requirements are often applied to the artefacts such pieces are part of. Automatic processing of such artefacts and especially when they are composed of semi-structured or unstructured text parts, is in general not easy and requires fine-grained approaches for text classification and extraction.

The internship will focus on the study, development and enhance a number of proofs-of-concept of supervised and unsupervised approaches to information detection/extraction tasks. In particular, the intern will explore the most recent advancements in the domain (see [1,2,3]) for their inclusion in a number of prototypes for processing sensitive information (be them personal data according to GDPR, cyber security threat information or attack logs). Natural Language Processing but also techniques for classification of structured information will be considered. The objective is to classify relevant pieces of information in order to trigger the appropriate follow-up processing. The work will benefit from a number of already developed prototypes for the classification of pieces of information in free text, using deterministic and AI/ML techniques.

In the above-described context, the specific goals of the internship are as follows:

- Application of the recent approaches based on Language Models [1,2,3] for NLP and multi-task learning
- Evaluation of results on different information sources (Github, social media, internal SAP data)
- Integration of the results into existing prototypes

Technologies/techniques involved are: Python, Java, Docker, and AI/Machine Learning (RNN, CNN in particular)

We expect that 40% of time will be dedicated to research activities, and 60% to development

CANDIDATE PROFILE

- University Level: Last year of MSc or less if the student has a good profile
- Good knowledge of the Python programming language, Java is a plus
- Good knowledge of data science and machine learning algorithms
- Interest in research work
- Fluency in English (working language)
- Good oral and written communication skills

INTERNSHIP CONTEXT

SAP

Founded in 1972, SAP has grown to become the world's leading provider of business software solutions. SAP is market leader in enterprise application software. The company is also the fastest-growing major database company. Globally, more than 77% of all business transactions worldwide touch an SAP software system. With more than 347.000 customers in more than 180 countries, SAP includes subsidiaries in all major countries. SAP is the world's largest inter-enterprise software company and the world's third-largest independent software supplier, overall. SAP solutions help enterprises of all sizes around the world to improve customer relationships, enhance partner collaboration and create efficiencies across their supply chains and business operations. SAP employs more than 98.600 people.

Security Research at SAP Labs France, Sophia Antipolis

Based at SAP Labs France Mougins, Security Research Sophia-Antipolis addresses the upcoming security needs, focusing on increased automation of the security life cycle and on providing innovative solutions for the security challenges in networked businesses, including cloud, services and mobile.

STANDARD INTERNSHIP PACKAGE

- *Salary*: depending on the length of the internship and your diploma.
- *Lunch*: SAP Labs France has a local cafeteria; interns contribute 2,63 €uro/lunch, like other SAP employees.
- *Holidays*: French Bank Holidays
 - January 1st; April 12th, April 13th, May 1st, May 8th, May 21st, June 1st, July 14th; August 15th, Nov 1st and 11th; December 25th
- *Travel*: no trip will be paid by SAP.
- *Accommodation*: SAP can propose an accommodation for the duration of your internship. The accommodation is subsidized by SAP: the intern pays half of the rental cost: 342€ for a 1-room apartment or 442€ for a 2-room apartment (Choice depending on the availability).

REFERENCES

- [1] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018). <https://arxiv.org/pdf/1810.04805.pdf>
- [2] Peters, Matthew E., et al. "Deep contextualized word representations." *arXiv preprint arXiv:1802.05365* (2018). <https://arxiv.org/abs/1802.05365>
- [3] Akbik, Alan, Duncan Blythe, and Roland Vollgraf. "Contextual string embeddings for sequence labeling." *Proceedings of the 27th International Conference on Computational Linguistics*. 2018. <http://alanakbik.github.io/papers/coling2018.pdf>

CONTACTS AND PROCEDURE

Please candidate by clicking on this link:

<https://career5.successfactors.eu/sfcareer/jobreqcareer?jobId=234528&company=SAP&username=>

UPLOAD (all documents must be in English):

- Your CV
- Cover letter
- Any relevant documents