**Carl von Ossietzky Universität Oldenburg** / 26111 Oldenburg

# Master's Thesis: Semantic Descriptions of Data Streams with Adaptive Windowing

Carl von Ossietzky
Universität
Oldenburg

Fakultät II - Informatik, Wirtschafts- und Rechtswissenschaften
**Department für Informatik**

**ABTEILUNG WIRTSCHAFTSINFORMATIK VLBA VERY LARGE BUSINESS APPLICATIONS**

Prof. Dr.-Ing. Jorge Marx Gómez
Tel.: 0441 798 4470
jorge.marx.gomez@uol.de

**Sekretariat**
Julia Franke
Tel. 0441 798 - 4478
Fax: 0441 798 - 4472
julia.franke@uol.de

Oldenburg, den 02.03.2021

**Standort**
Campus Haarentor, A4 3-315
Ammerländer Heerstraße 114-118
26129 Oldenburg

**Postanschrift**
26111 Oldenburg

**Paketanschrift**
Ammerländer Heerstraße 114–118
26129 Oldenburg

**Bankverbindung**
Landessparkasse zu Oldenburg
IBAN DE46 2805 0100 0001 9881 12
BIC SLZODE22

**Steuernummer**
6422008701

www.uol.de

## Background

Data streams can be thought of as data that is continuously flowing. This usually implies a one-pass constraint (i.e., you must do something with the incoming data point, or it is lost forever). Unfortunately, in some scenarios, it is impossible to either store all the streams in memory, send them to the cloud, or to process them with an arbitrary number of iterations [1]. Therefore, it becomes essential to gain quick insights from the data and explicitly represent its meaning (i.e., semantics ) to scale-up the well-known DIKW hierarchy [2].

One particular characteristic of streams is the sequential arrival of data points. Therefore, the time domain is an essential aspect of analytics. Depending on the data type, we can have time-series data (i.e., continuous data such as the speed) or discrete sequences (e.g., categorical data such as the gear number). A commonly used method to aggregate this type of data is by using time windows. Nevertheless, selecting the proper size of a time window is a use-case specific trade-off, and a fixed size limits the general applicability. To cope with this limitation, there are works based on the principle of adaptive windowing [3].

## Description

The purpose of this thesis is to generically extract human-understandable meaning out of data streams using an adaptive windowing approach. The technical research problem that this thesis is intended to address is to:

- Improve the semantic enrichment of data streams, by designing an artifact that describes the meaning of data over time using adaptive windowing; so that it enables data stream analytics without a use-case specific time window.

## Tasks

This position considers six months for the thesis. Some of the tasks to expect are: literature review on adaptive windowing and analytics applied to sensor data streams, dataset preparation, and use case definition.

VERY LARGE
BUSINESS
APPLICATIONS
Carl von Ossietzky
Universität Oldenburg

Carl von Ossietzky
Universität
Oldenburg

### Expected skills

- Expertise in at least one high-level programming language (e.g., Python)
- General data engineering skills: data structures, data analysis (e.g., pandas)
- General knowledge about data streams (i.e., time-series, discrete sequences)
- Confidence in statistics and probability
- English language proficiency in both speaking and writing

### Optional skills

- Experience with machine learning and deep learning
- Git distributed version control system

### Related Research Areas

- Data Stream Mining, Statistics, Machine Learning, Data Stream Analytics, Semantics

### Notes

- Supervision is offered in English
- Starting date: as soon as possible

### References

[1] Gama, J. (2010). ***Knowledge discovery from data streams***. Chapman & Hall/CRC.
[2] Rowley, J. (2007). ***The wisdom hierarchy: Representations of the DIKW hierarchy***. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706
[3] Bifet, A., & Gavaldà, R. (2007). ***Learning from Time-Changing Data with Adaptive Windowing***. *Proceedings of the 2007 SIAM International Conference on Data Mining*, 443–448. https://doi.org/10.1137/1.9781611972771.42
[4] Aggarwal, C. C. (2015). ***Data Mining***. Springer International Publishing. https://doi.org/10.1007/978-3-319-14142-8

### Contact

**Daniel Alvarez, M.Sc.**
External PhD Student
University of Oldenburg
Department of Computing Science
Chair of Business Informatics (Very Large Business Applications)
Ammerländer Heerstr. 114-118
26129 Oldenburg
Tel. +49 176 / 2111-7353
E-Mail: daniel.alvarez@uni-oldenburg.de