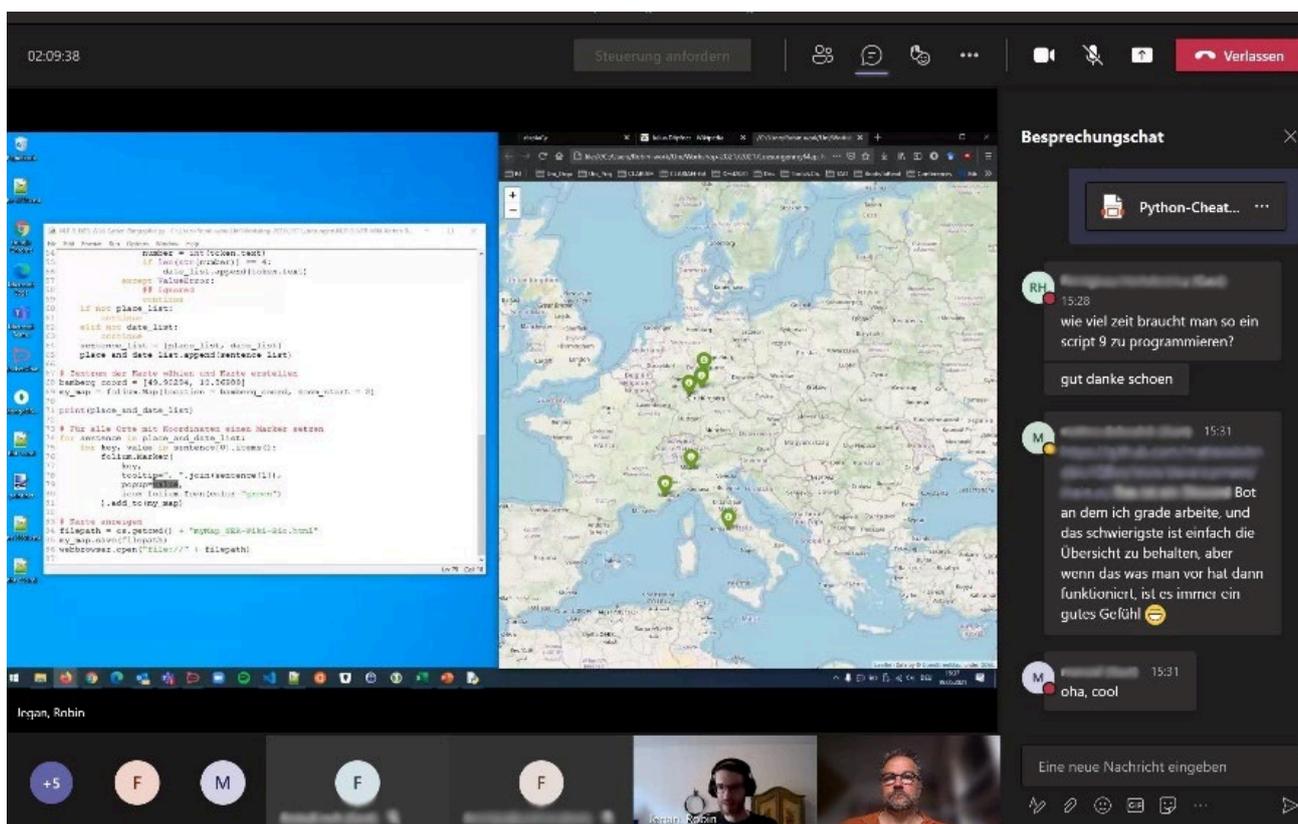


## Nachbericht: Wikipedia durchforsten mit Python

Ein Angebot im Rahmen des Schülerforschungszentrums der TechnologieAllianzOberfranken (TAO)

Der Workshop „Wikipedia durchforsten mit Python“ fand im Mai 2021 statt, aufgrund der Corona-Beschränkungen im digitalen Format, das heißt als Online-Workshop. Am Workshop nahmen 9 Schülerinnen und Schüler (Klasse 9-11) aus oberfränkischen Gymnasien teil, die mithilfe der Programmiersprache Python verschiedene Anwendungen im Zusammenhang mit Natürlicher Sprachverarbeitung bearbeiteten.



Der Workshop wurde konzipiert und betreut von Robin Jegan (Wissenschaftlicher Mitarbeiter am Lehrstuhl für Medieninformatik an der Universität Bamberg) und Lutz Reuter (Kontaktlehrer des TAO-Schülerforschungszentrum). Zusätzliche Unterstützung wurde durch Jonas Troles (Mitarbeiter am TAO-Schülerforschungszentrum) angeboten, der insbesondere bei der Anmeldung und Organisation, aber auch während des Workshops tätig war.

Nachdem bereits einige Tage vor dem Workshop gemeinsam mit den Schülerinnen und Schülern in einem Web-Meeting erste Kontakte geknüpft und die technische Installation geprüft wurden, trafen sich die 9 Teilnehmerinnen und Teilnehmer mit den Betreuern am 21ten Mai via Microsoft Teams. Dort wurden zunächst in einer



Präsentation grundlegende Informationen zu Python vermittelt, bevor auf Bereiche der Natürlichen Sprachverarbeitung eingegangen wurde, die für den Workshop relevant waren.

Anschließend wurden den Schülerinnen und Schülern Programmierskripte übermittelt, die in zwei Schwierigkeitsgraden vorbereitet wurden, damit auch Schülerinnen und Schüler mit Vorkenntnissen in Python sowohl eine Herausforderung wie auch einen Mehrwert durch die schwierigeren Skripte erhalten konnten. In zwei Praxiseinheiten wurden die Skripte am Vormittag und am Nachmittag von den Teilnehmerinnen und Teilnehmern bearbeitet, die sie mithilfe des Python Editors IDLE (<https://www.python.org/downloads/>), welcher in der Standard-Installation von Python mitgeliefert wird, auf ihren eigenen PCs und Laptops anpassen und ausführen konnten.

In den Skripten wurden drei Anwendungsgebiete vermittelt. Zunächst wurden Daten und Inhalte aus der deutsch- sowie englischsprachigen Wikipedia extrahiert. Daraufhin wurden zwei grundlegende Techniken der Natürlichen Sprachverarbeitung eingeführt: Zum einen die Wortarterkennung, also die Einordnung von Wörtern in Nomen, Verben, Adjektiven, etc., und zum anderen die Eigennamenerkennung, das heißt die Identifikation von Worten, die Orte, Personen, Zeitangaben oder anderen Kategorien zugeordnet werden können. Anschließend wurden diese beiden Techniken kombiniert, das heißt die aus Wikipedia gewonnenen Daten wurden mithilfe der Methoden der Natürlichen Sprachverarbeitung analysiert. Das Ergebnis dieser Kombination waren beispielsweise Frequenzlisten, das heißt eine Auflistung und Summe der verwendeten Wortarten, oder Visualisierungen, die etwa alle Orte und Personen aus einem Wikipedia-Text farbig markiert im Browser darstellen. Als drittes Anwendungsgebiet wurden digitale Karten gewählt, die analog zu bekannten Kartenanbietern wie Google Maps oder OpenStreetMap, eine browserbasierte Ansicht auf eine Karte der Welt ermöglichen, auf der man bis auf Straßenebene hineinzoomen kann. Mithilfe der Python Skripte wurden diese Karten angelegt und Marker auf diesen platziert. Zuletzt wurden die drei genannten Anwendungsgebiete kombiniert, um Orte und Zeitangaben aus den Wikipedia Texten herauszufiltern, die verwendet wurden, um die Stationen in der Biographie einer Person mittels Marker auf einer Karte zu platzieren (Im Screenshot ist das Web-Meeting zu sehen, in dem das Ergebnis eines solchen Skriptes auf der Karte aufgerufen wurde).

Die Schülerinnen und Schüler konnten dabei eigene Texte auswählen, diese nach unterschiedlichen Kriterien bearbeiten und die Karten und dort platzierten Marker anpassen. Dabei wurden auftretende Probleme sowohl im Web-Meeting in der Gruppe mit allen Teilnehmerinnen und Teilnehmern wie auch in Einzelgesprächen via Microsoft Teams gelöst, um eine individuelle Betreuung zu ermöglichen.