

Vorbereitung für den Workshop „Einführung in Jupyter Notebooks“ an der Graduiertenakademie der TUHH

Marvin Kastner <marvin.kastner@tuhh.de>

6. Mai 2024

In diesem Dokument werden die Schritte aufgezählt, die von den Workshop-Teilnehmenden selbständig *vor* Beginn des Workshops auf dem eigenen Endgerät (PC, Laptop, Apple, ...) durchgeführt werden müssen. Dies ist notwendig, weil zur Einrichtung der Arbeitsumgebung große Mengen an Daten heruntergeladen und verarbeitet werden müssen.

Im Folgenden wird aufgezählt, welche Software und welche Dateien benötigt werden. Hier gibt es nur manchmal eine zeitliche Abhängigkeit, schauen Sie also gerne schon etwas voraus und parallelisieren Sie Aufgaben da, wo dies den Prozess beschleunigt. Der Text drumherum beschreibt, warum diese Schritte gemacht werden und zeigt ein paar Alternativen auf.

Installation von JupyterLab

Gehen Sie auf die Webseite <https://www.anaconda.com/distribution/> und laden Sie die Anaconda-Version für Ihr Betriebssystem herunter. Die Registrierung per Mail kann einfach übersprungen werden. Anaconda ist eine Python-Distribution, die die Entwicklungsumgebung *JupyterLab*, welche im Workshop verwendet wird, bereits mit ausliefert. Der Download-Bereich der Webseite sollte ungefähr wie in Abbildung 1 aussehen.

Falls Sie auf Ihrem Endgerät zwei Accounts, einen Administrator- bzw. root-Account und einen Account fürs alltägliche Arbeiten, verwenden, seien Sie bitte vorsichtig. Die Installation von Anaconda sollte mit dem Account fürs alltägliche Arbeiten durchgeführt werden. Eine Installation mit einem Nutzer mit erhöhten Rechten (z. B. *root* oder einem dedizierten Admin-Account) kann u. U. dazu führen, dass Anaconda nur für diesen einen Nutzer mit erhöhten Rechten installiert ist bzw. dass es zu Problemen mit der Rechteverwaltung kommt. In der Vergangenheit hat ein erneutes Installieren von Anaconda dieses Problem dann *nicht* beheben können. Lösungen hierfür lassen sich unter <https://docs.anaconda.com/anaconda/install/multi-user/> finden.

Rufen Sie nach dem Download den Installer auf und folgen Sie den Installationsschritten. Konsultieren Sie im Fehlerfall offizielle



Abbildung 1: Der Download-Bereich von Anaconda (Ausschnitt).

Quellen des Herstellers (wie z. B. <https://docs.anaconda.com/anaconda/install/>) oder Foren (wie z. B. <https://stackoverflow.com>).

Bezug der Seminar-Materialien

Klonen Sie das git-Repository <https://github.com/lkastner/einfuehrung-in-jupyter-notebooks.git>, damit Sie die Workshop-Materialien lokal haben – dies umfasst nur den Programmierteil. Am einfachsten ist es, wenn Sie die Dateien lokal unterhalb des benutzer-eigenen Ordners Dokumente ablegen (unter Windows typischerweise `C:\Users\BENUTZERNAME\Documents`). Denn dieser Ordner ist, wenn JupyterLab startet, direkt verfügbar. Falls Sie noch nie mit git gearbeitet haben, lesen Sie bitte die nächsten Absätze.

WARUM SOLLTE ICH GIT LERNEN? Für die Versionsverwaltung ist git quasi der Standard und wird immer häufiger auch außerhalb der Software-Entwicklung, aus der es ursprünglich stammt, eingesetzt. Deswegen lohnt es sich für (fast) alle, sich Fähigkeiten mit diesem Tool anzueignen. An der TUHH gibt es unter <https://collaborating.tuhh.de/> eine eigene GitLab-Instanz und die Graduiertenakademie bietet regelmäßig Workshops an, um sich mit der Bedienung vertraut zu machen. Gerade im Zusammenhang mit Jupyter Notebooks sind diese sehr zu empfehlen!

Das Versionierungssystem git ist ursprünglich als ein Kommandozeilentool entwickelt worden, welches über <https://git-scm.com/> heruntergeladen werden kann. Wer lieber grafische Oberflächen mag, kann sich eine von vielen GUI-Clients¹ aussuchen. Hier sollte neben dem Betriebssystem auch die ggf. kostenpflichtige Lizenz beachtet werden. Einige Lizenzen unterscheiden z. B. zwischen der privaten Verwendung und der Verwendung im Arbeitskontext. Bis zum Start des Workshops werden u. U. die Materialien noch überarbeitet oder erweitert. Aktualisieren Sie also bitte regelmäßig Ihre vorliegende Version über ein `git pull` bzw. durch das Klicken auf den Button „Pull“ im GUI-Client Ihrer Wahl.

IST ES FÜR DEN WORKSHOP ZWINGEND NOTWENDIG, GIT ZU LERNEN? Falls Ihnen git unbekannt ist und Sie keine Zeit dafür haben, sich mit git auseinanderzusetzen, gibt es auch die Möglichkeit, den Inhalt als ZIP-Ordner herunterzuladen. Klicken Sie dafür auf den Button, wie er in Abbildung 2 zu sehen ist. Falls Lernmaterialien später noch angepasst werden, müssen Sie diese dann allerdings erneut herunterladen und in einem neuen Ordner entpacken.

¹ Eine Liste ist auf <https://git-scm.com/download/gui/win> zu finden.

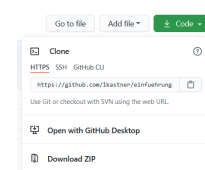


Abbildung 2: Ein GitHub-Repository bietet verschiedene Möglichkeiten zum Bezug der Inhalte an.

Installation der Bibliotheken

Mit dem Anaconda Navigator können die im Workshop später benötigten Bibliotheken automatisch installiert werden. In der Abbildung 3 sehen Sie unten den Button „Import“ (im Screenshot mit einer (1) markiert). Klicken Sie diesen an. Damit öffnet sich das Fenster „Import new environment“. Klicken Sie hier auf den Ordner in der Zeile, die mit „Specification File“ beginnt (mit einer (2) markiert). Danach öffnet sich ein Fenster (mit einer (3) markiert), in dem Sie dann zu den heruntergeladenen Dateien navigieren können. Wählen Sie die `environment.yml` aus den soeben bezogenen Dateien aus (sie liegt auf der obersten Ebene des Projektordners) und schließen den Vorgang ab. Das Erstellen der Umgebung nimmt für gewöhnlich einige Zeit in Anspruch.

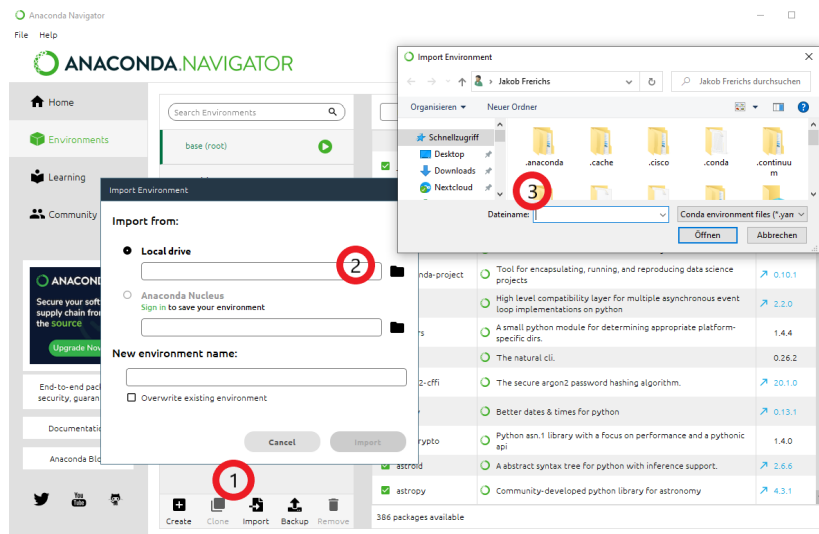


Abbildung 3: Der Anaconda Navigator erlaubt das Importieren von `environment.yml`-Dateien.

WARUM IST DIESER SCHRITT NOTWENDIG? Die Jupyter Notebooks, die Sie soeben heruntergeladen haben, benötigen spezielle Bibliotheken. Einige werden bereits bei der Installation mit ausgeliefert, andere müssen aber noch nachinstalliert werden. Auf der obersten Ebene des git-Repositorys befindet sich die Datei `environment.yml`, in der alle benötigten Bibliotheken aufgelistet werden. Diese Datei sollten Sie nun bereits auf Ihrem Endgerät in einem lokalen Ordner vorliegen haben. Die Struktur der Datei `environment.yml` ist von Anaconda vorgegeben und erlaubt es, die Abhängigkeiten von Bibliotheken automatisch aufzulösen. Damit man an einem PC in verschiedenen Projekten unterschiedliche Versionen einer gleichen Bibliothek haben kann, strukturiert Anaconda die zu einem Projekt gehörenden Bi-

bibliotheken standardmäßig in Umgebungen (eng. Environments). Für das Seminar erstellen wir die Umgebung `intro-jupyter-notebooks` basierend auf der gegebenen `environment.yml`.

ABER IST ES DENN NOTWENDIG, HIER MIT EINER GUI ZU ARBEITEN? Natürlich gibt es auch ein Kommandozeilentool, das bereits mit Anaconda ausgeliefert worden ist. Es heißt `conda`. Unter Windows wird dies je nach Auswahl während der Installation nicht in die Pfad-Variable mit aufgenommen. Dann steht Ihnen immer noch die *Anaconda Prompt* zur Verfügung. Weitere Informationen gibt es auf <https://docs.anaconda.com/anaconda/install/verify-install/#conda>. Auf <https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#creating-an-environment-from-an-environment-yml-file> wird erläutert, wie eine existierende `environment.yml` eingelesen werden kann. Folgen Sie der Anleitung und erstellen Sie die passende Umgebung.

Start von JupyterLab

Nun ist JupyterLab einsatzbereit. Es kann nun über den Anaconda Navigator gestartet werden. In Abbildung 4 ist dies abgebildet. Zunächst wird links im Menü (im Screenshot mit einer (1) markiert) „Home“ ausgewählt. Im zweiten Schritt muss die Umgebung `intro-jupyter-notebooks` ausgewählt werden (mit einer (2) markiert). Danach startet ein Klicken auf `Launch JupyterLab` im Browser (mit einer (3) markiert). Standardmäßig öffnet sich nun ein neuer Tab im Browser und JupyterLab zeigt zunächst den Inhalt vom Benutzer-Ordner an (unter Windows typischerweise `C:\Users\BENUTZERNAME`). Dies sollte wie in Abbildung 5 aussehen.

Erste Schritte mit JupyterLab

Wenn Sie die Workshop-Materialien wie unter Abschnitt „Bezug der Seminar-Materialien“ angegeben abgelegt haben, können Sie in JupyterLab zum Ordner `00-Installationscheck` navigieren und dort das Jupyter Notebook (die Datei mit der Endung `.ipynb`) mit einem Doppelklick öffnen und dort überprüfen, ob Sie alle Zellen ausführen können. Falls es Fehlermeldungen gibt, wenden Sie sich gerne bereits im Vorwege an die Workshop-Leitung. Falls Sie noch keine Erfahrungen mit Python haben und Sie Interesse haben, sich in diese Sprache einzudenken, können Sie sich die Einführungsmaterialien unter `01-Einfuehrung-in-Python` anschauen.

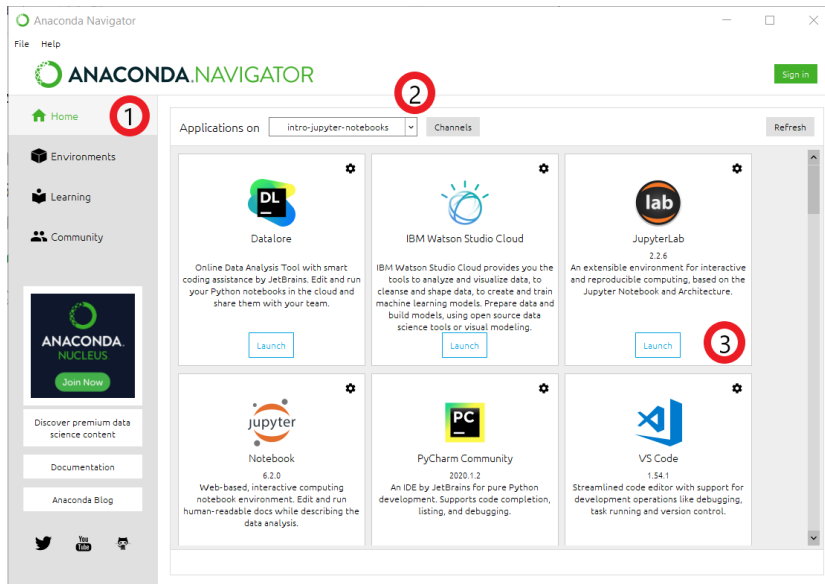


Abbildung 4: Aus dem Anaconda Navigator kann JupyterLab gleich in der richtigen Umgebung gestartet werden.

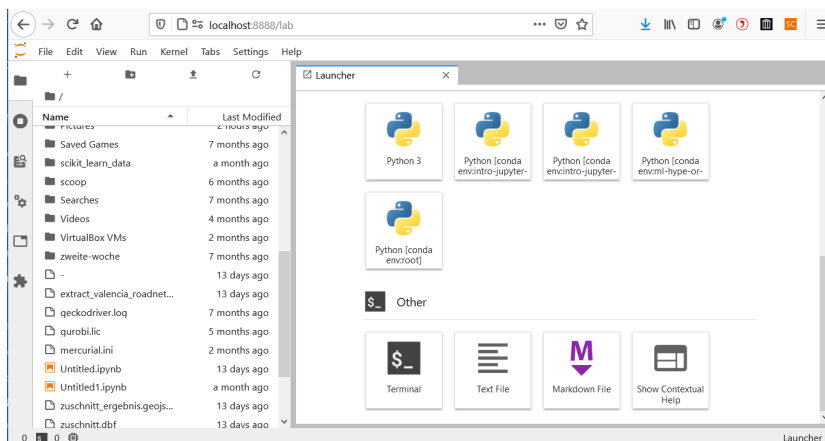


Abbildung 5: In JupyterLab können links alle Dateien und Ordner aus dem Benutzerordner betrachtet werden. Nun können die Dateien mit der Endung .ipynb mit einem Doppelklick geöffnet werden.