SCHA.T.S.I - Nutzerhandbuch

Hallo lieber Nutzer. Schön, dass Du dich entschieden hast SCHA.T.S.I eine Chance zu geben. In diesem Dokument findest du eine Anleitung, wie man SCHA.T.S.I einrichtet und wie man es nutzt.

Dabei wird hier unterschieden zwischen den Versionen <=1.4.1 und den nachfolgenden Versionen, da von dort an SCHA.T.S.I umstrukturiert wurde, um zukünftige Funktionen realisieren zu können.

Allgemein gilt es für alle Versionen, dass SCHA.T.S.I auf der Software "Docker" aufbaut. Diese Technik ermöglicht es, unseren Code zentral für alle Betriebssysteme schreiben und testen zu können und diesen dann in einen Container zu packen, welcher auf dem jeweiligen Betriebssystem gestartet werden kann (Siehe Installation beim jeweiligen Betriebssystem). Für das verständnis ist nur wichtig zu wissen, dass ein solcher Container alle Programme mitbringt, die er benötigt, sodass nur Docker selbst installiert werden muss, um das SCHA.T.S.I nutzen zu können.

Aufgaben

SCHA.T.S.I dient dazu wissenschaftliche Paper zu analysieren, den Text zu extrahieren und diesesn zu Untersuchen, um die Bedeutung für den Nutzer anhand von selbstgewählten Keywords festzulegen. Der Extrahierte Text wird untersucht, alle Wörter und Wortgruppen ohne Bedeutung entfernt und die Häufigkeit der auftretenden Wortgruppen von Bedeutung ermittelt. Diese werden mit den eigenen KeyWords abgeglichen um somit die Relevanz für den Nutzer zu ermitteln. Anschließend werden die Paper gerankt. Danach werden die Referenzen extrahiert und diese Informationen genutzt, um darzustellen, ob es Verbindungen zwischen den einzelnen wissenschaftlichen Veröffentlichungen gibt. Diese Zusammenhänge werden als Netz visualisiert.

Nachfolgend sind kurz die Strukturen von SCHATSI in den verschiedenen Versionen dargestellt, um besser zu verstehen, wie SCHA.T.S.I einzusetzen ist.

Den Quellcode, die Versionen und weiterführende Informationen sind auf der folgenden Github-Seite zu finden: SCHATSI-Github

Allgemeine Struktur bis Version 1.4.1

SCHA.T.S.I vor Version 1.4.2

SCHATSI

Laufzeit überwachen
Text extrahieren
Wörter zahlen
Bigrams und Trigrams
Ausdrücke filtern
Metadaten
Referenzen
Ranking erstellen

SCHATSI Machine Learning R

> Word-Cloud Topic-Modelling

Ganz Allgemein kann man sagen, dass bis Versionen 1.4.1 Alle Funktionen außer den Funktionen zum Elnsatz künstlicher Intelligenz in einem einzigen Container zusammengefasst sind (Aufteilung in SCHA.T.S.I und SCHA.T.S.I Machine Learner)

Diese können einzeln und voneinander unabhängig eingesetzt werden.

Allgemeine Struktur ab Version 1.4.2

SCHA.T.S.I ab Version 1.4.2

SCHATSI DataCleanser SCHATSI Ranker SCHATSI Machine Learning R

Laufzeit überwachen
Text extrahieren
Wörter zahlen
Bigrams und Trigrams
Ausdrücke filtern
Metadaten
Referenzen

Laufzeit überwachen Negatives Filtering Ranking erstellen

Word-Cloud Topic-Modelling

Bei Versionen >1.4.1 wurde SCHA.T.S.I in 3 einzelne Komponenten unterteilt.

- 1. Den SCHATSI DataCleanser,
- 2. Den SCHATSI Ranker und
- 3. Den SCHATSI_Machine_Learner_R

Diese 3 Komponenten erfüllen aufeinander aufbauende Funktionen, welche jedoch generell unabhängig voneinander sind, sodass diese auch einzeln ausgeführt oder wiederholt werden können.

2. Windows

2.1 Installation

Da Windows eine andere Umgebung als Docker selbst nutzt ist es für eine Installation wichtig, dass das "Windows Subsystem For Linux" (kurz WSL) auf dem Rechner vorhanden und auch einsatzbereit ist. Dies ist auf nahezu allen neueren Rechnern der Fall.

Sollten Probleme auftreten ist es ratsam zuerst zu prüfen, ob WSL korrekt läuft. Dazu gibt es auf der Website von Docker eine gute Anleitung: Docker für Windows

Ist WSL einsatzfähig, kann auf der folgenden Website eine Installationsdatei für Docker-Desktop heruntergeladen werden Docker-Desktop-Installation

Ist die Datei heruntergeladen kann diese (je nach Systemkonfiguration auch mit Administratorrechten) gestartet werden. Es öffnet sich ein Fenster welches durch die komplette Installation leitet. Ist die Installation beendet empfehlen wir einen Systemneustart und ein einmaliges Starten von Docker Desktop, um zu prüfen, ob die Installation erfolgreich war.

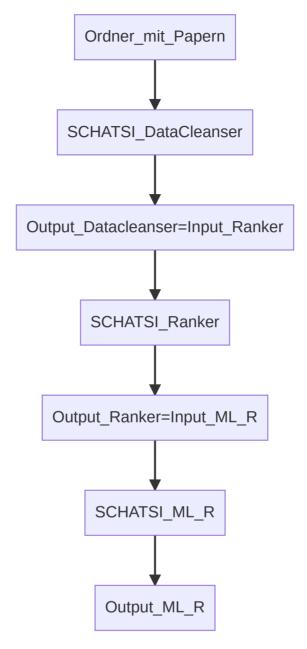
2.2 Einsatz

Ist Docker einsatzbereit, so kann nun die eigentliche Software bezogen werden.

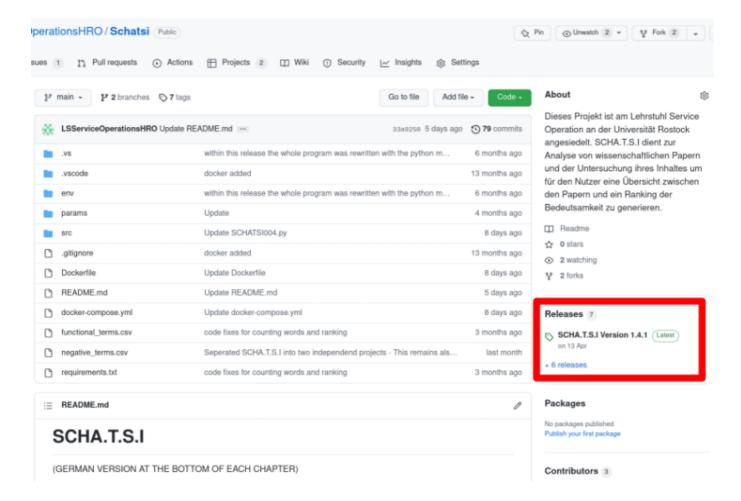
SCHATSI ist bis zur Version 1.4.1 in zwei, bei darauf folgenden Versionen in drei Komponenten aufgeteilt. Diese sind im jeweiligen Repository auf Github zu finden

- SCHATSI_DataCleanser
- SCHATSI Ranker
- SCHATSI_Machine_Learner_R

Die Nutzung jedes dieser Komponenten funktioniert exakt gleich, sodass wir hier die Nutzung der ersten Komponente erklären. Sollen die anderen ebenfalls zum Einsatz kommen müssen diese nur aus dem jeweiligen Repository heruntergeladen werden. Als Input-Ordner sind dann jeweils die Output-Ordner der Vorgänger, wie im Diagramm erklärt auszuwählen, doch dazu gleich in der vollständigen Erklärung mehr.



Auf der Github-Seite des Projektes finden Sie auf der rechten Seite die aktuellste Releaseversion, jeweils für ein bestimmtes Betriebssystem. SCHATSI-Repository (siehe Bild)



Sie können die Software in einem .zip-Ordner Herunterladen.

Nun müssen Sie diese nur noch Entpacken. Enthalten sind:

- Ein Ordner namens "params"
- Eine .csv-Datei "functional terms.csv"
- Ab Version 1.4.2 eine .csv-Datei "negative terms.csv"
- Eine Readme-Datei mit Hinweisen zum Einsatz von SCHATSI
- Eine Datei namens "docker-compose.yml" Diese kann einfach ignoriert werden, da sie nur intern benötigt wird
- Eine Starter-Datei namens "SCHATSI RUN" mit welcher Sie SCHATSI zum laufen bringen

2.2.1 Vor dem Start

Nun können Sie SCHATSI das erste mal starten. Sicherlich haben Sie einen Ordner, in dem alle Paper enthalten sind, die sie analysieren möchten - das ist gut! Denn das wird unser Input-Ordner.

Wichtig: In diesen Ordner müssen Sie noch zwei Dinge verschieben:

- Den Ordner "params"
- Die Datei "functional_terms.csv"

Weiterhin ist jetzt der Zeitpunkt gekommen, wo Sie sich überlegen müssen, welche Begriffe für ihre Analyse oder ihr Thema entscheidend sind. Gemeint sind dabei Begriffe, nach denen SCHA.T.S.I in den Papern suchen soll. Dabei können ganz unterschiedliche Ausdrücke gewäht werden und je konkreter diese zu ihrer Aufgabe passen desto hochwertiger wird die Arbeit von SCHA.T.S.I ausfallen.

Erlaubt sind dabei Ausdrücke aus einem, zwei oder drei Wörtern. Dies soll ermöglichen auch fachspezifische Ausdrücke wie "friction welding" oder "Internet of Things" zu erlauben.

1. Öffnen die dafür jetzt die Datei "functional terms.csv".

Sie sehen eine beinahe leere Tabelle abgesehen von den beiden Überschriften *term* und *cluster*.

- 2. Nun können Sie alle Ausdrücke, die Sie in ihre Untersuchung aufnehmen möchten untereinander in die Spalte *term* eintragen.
- 3. Die Spalte *cluster* dient dazu, Ausdrücke einzelnen Themengebieten zuzuordnen. Im späteren Verlauf von SCHA.T.S.I sollen diese genutzt werden, um die Paper auch in diese Bereiche einordnen zu können, sodass eine Übersicht entsteht, welches Paper Inhalte zu bestimmten Themenbereichen enthält.
 - Beispielhaft könnte eine Wirtschaftswissenschaftler, der eine wissenschaftliche Arbeit im Bereich Steuerwesen und Buchhaltung schreibt bestimmte Begriffe eher dem Bereich Steuern und andere in den Bereich Buchhaltung einordnen.
- 4. Speichern Sie das Dokument. Bei der Frage ob von Excel oder ähnlichen Programmen die Style-Information oder Formatierung geändert werden soll, sollte dies abgelehnt werden, da SCHA.T.S.I die folgende Form benötigt:

term; cluster

ausdruck1;cluster

Eine beispielhafte Tabelle ist hier abgebildet.

term	cluster
buchungssatz	Buchhaltung
umsatzsteuer	Steuerwesen
haben	Buchhaltung
soll	Buchhaltung
Gewinn	Steuerwesen

2.2.2 SCHA.T.S.I starten

Wenn alle Voraussetzungen aus den vorherigen Kapiteln erfüllt sind kann SCHA.T.S.I. jetzt gestartet werden.

Auf Windows ist das Starten nach den vorherigen Schritten sehr einfach.

Das einzige was Sie tun müssen ist die Datei "SCHATSI RUN.exe" durch Doppelklick zu starten.

- Es öffnet sich ein Fenster in welchem man sich durch die Ordner navigieren kann. Navigieren sie zum Ordner, welche ihre Paper enthält, wählen Sie Diesen aus und bestätigen sie Ihre Wahl.
- SCHA.T.S.I beginnt seine Arbeit und durchläuft einmal den gesamten Prozess, erstellt einen Ordner names Output und vermeldet auch, wenn seine Arbeit beendet ist.

In diesem Ordner liegen mehrere .csv-Dateien, welche die Daten der Analyse enthalten.

MICHTIG: Diese Anleitung erklärt die Nutzung einer der Komponenten von SCHA.T.S.I., damit ist der vollständige Funktionsumfang noch nichgt abgeschlossen. Die Prozesse wurden getrennt, um diese im einzelnen Wiederholen zu können. Wenn Sie nun die nächste Komponente nutzen wollen, dann müssen Sie diese nur aus Ihrem Repository herunterladen, entpacken und als Input-Ordner den gerade eben generierten Output-ordner auswählen, da die Ergebnisse der ersten Komponente der Input der zweiten sind (und nach dem gleichen Muster gilt dies auch für den Einsatz der 3. Komponente).

🔥 Der Inhalt des Outputs, wird in Kapitel 5 erläutert, damit Sie diesen nachvollziehen können.

3. Linux

Die Installation von Docker und der Einsatz von SCHA.T.S.I gestaltet sich zumeist reibungslos, da die Dockerumgebung ebenfalls auf Linux basiert. Es müssen lediglich die grundlegenden Programme installiert werden.

3.1 Installation

Um SCHA.T.S.I einsetzen zu können werden auf einem Linux-System die Programme "docker" und "docker-compose" benötigt. Diese sind auf den allermeisten Distributionen bereits in der Paketverwaltung vorhanden und können entweder über einen grafischen Paketstore oder das Terminal installiert werden. Hier werden einmal die Installationsmethoden für ein Debianbasiertes System und für ein Arch-basiertes System aufgezeigt.

Für beide gilt: Öffne ein Terminal mit Strg+Alt+T

3.1.1 Ubuntu und andere Debian-basierte Systeme

Im Terminal müssen folgende Befehle ausgeführt werden. Dabei werden zuerst das Softwarerepository geupdated, anschließend benötigte Software installiert und die Installation von Docker selbst ermöglicht, welche Schlussendlich durchgeführt wird

```
sudo apt-get update
sudo apt-get install \
    ca-certificates \
   curl \
   gnupg \
    lsb-release
sudo mkdir -p /etc/apt/keyrings
curl -fsSL https://download.docker.com/linux/ubuntu/gpg | sudo gpg --dearmor -o
/etc/apt/keyrings/docker.gpg
echo \
  "deb [arch=$(dpkg --print-architecture) signed-
by=/etc/apt/keyrings/docker.gpg] https://download.docker.com/linux/ubuntu \
 $(lsb_release -cs) stable" | sudo tee /etc/apt/sources.list.d/docker.list >
/dev/null
sudo apt-get update
sudo apt-get install docker-ce docker-ce-cli containerd.io docker-compose-
plugin
```

3.1.2 Arch und Archbasierte Systeme

Für Arch und andere Distributionen wie Manjaro gestaltet sich die Installation schneller. Auch hier sollte zuerst das Softwarerepository aktualisiert werden. Anschließend können "docker" und "docker-compose" mittels der Softwareverwaltung pacman installiert werden

```
sudo pacman -Syyu
sudo pacman -S docker docker-compose
```

Anschließend kann es von Nöten sein den Rechner einmal neu zu starten, um die Änderungen zu übernehmen. Anschließend können wir mit dem Einsatz von SCHA.T.S.I beginnen.

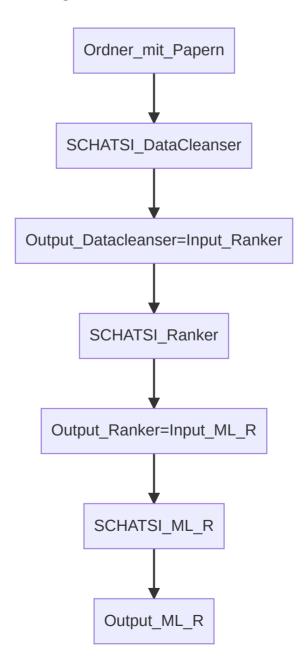
3.2 Einsatz

Ist Docker einsatzbereit, so kann nun die eigentliche Software bezogen werden.

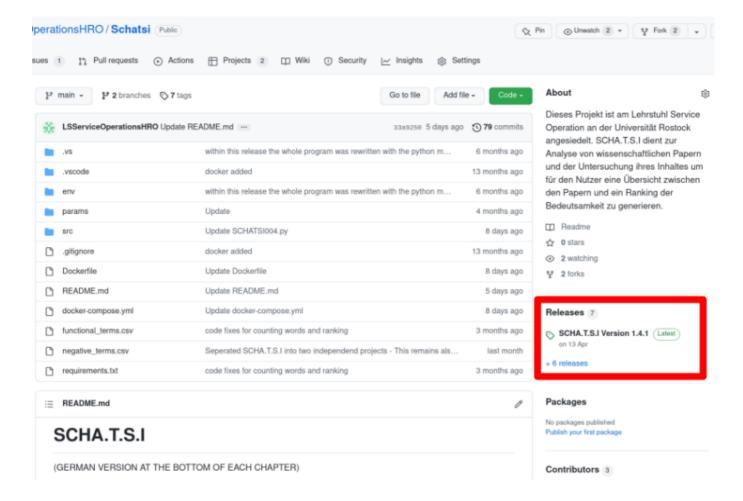
SCHATSI ist bis zur Version 1.4.1 in zwei, bei darauf folgenden Versionen in drei Komponenten aufgeteilt. Diese sind im jeweiligen Repository auf Github zu finden

- SCHATSI DataCleanser
- SCHATSI_Ranker
- SCHATSI_Machine_Learner_R

Die Nutzung jedes dieser Komponenten funktioniert exakt gleich, sodass wir hier die Nutzung der ersten Komponente erklären. Sollen die anderen ebenfalls zum Einsatz kommen müssen diese nur aus dem jeweiligen Repository heruntergeladen werden. Als Input-Ordner sind dann jeweils die Output-Ordner der Vorgänger, wie im Diagramm erklärt auszuwählen, doch dazu gleich in der vollständigen Erklärung mehr.



Auf der Github-Seite des Projektes finden Sie auf der rechten Seite die aktuellste Releaseversion, jeweils für ein bestimmtes Betriebssystem. SCHATSI-Repository (siehe Bild)



Sie können die Software in einem .zip-Ordner Herunterladen.

Nun müssen Sie diese nur noch Entpacken. Enthalten sind:

- Ein Ordner namens "params"
- Eine .csv-Datei "functional terms.csv"
- Ab Version 1.4.2 eine .csv-Datei "negative_terms.csv"
- Eine Readme-Datei mit Hinweisen zum Einsatz von SCHATSI
- Eine Datei namens "docker-compose.yml" Diese kann einfach ignoriert werden, da sie nur intern benötigt wird
- Eine Starter-Datei namens "SCHATSI RUN" mit welcher Sie SCHATSI zum laufen bringen

3.2.1 Vor dem Start

Nun können Sie SCHATSI das erste mal starten. Sicherlich haben Sie einen Ordner, in dem alle Paper enthalten sind, die sie analysieren möchten - das ist gut! Denn das wird unser Input-Ordner.

- Michtig: In diesen Ordner müssen Sie noch zwei Dinge verschieben:
 - Den Ordner "params"
 - Die Datei "functional terms.csv"

Weiterhin ist jetzt der Zeitpunkt gekommen, wo Sie sich überlegen müssen, welche Begriffe für ihre Analyse oder ihr Thema entscheidend sind. Gemeint sind dabei Begriffe, nach denen SCHA.T.S.I in den Papern suchen soll. Dabei können ganz unterschiedliche Ausdrücke gewäht werden und je konkreter diese zu ihrer Aufgabe passen desto hochwertiger wird die Arbeit von SCHA.T.S.I ausfallen.

Erlaubt sind dabei Ausdrücke aus einem, zwei oder drei Wörtern. Dies soll ermöglichen auch fachspezifische Ausdrücke wie "friction welding" oder "Internet of Things" zu erlauben.

1. Öffnen die dafür jetzt die Datei "functional_terms.csv".

Sie sehen eine beinahe leere Tabelle abgesehen von den beiden Überschriften *term* und *cluster*.

- 2. Nun können Sie alle Ausdrücke, die Sie in ihre Untersuchung aufnehmen möchten untereinander in die Spalte *term* eintragen.
- 3. Die Spalte *cluster* dient dazu, Ausdrücke einzelnen Themengebieten zuzuordnen. Im späteren Verlauf von SCHA.T.S.I sollen diese genutzt werden, um die Paper auch in diese Bereiche einordnen zu können, sodass eine Übersicht entsteht, welches Paper Inhalte zu bestimmten Themenbereichen enthält.
 - Beispielhaft könnte eine Wirtschaftswissenschaftler, der eine wissenschaftliche Arbeit im Bereich Steuerwesen und Buchhaltung schreibt bestimmte Begriffe eher dem Bereich Steuern und andere in den Bereich Buchhaltung einordnen.
- 4. Speichern Sie das Dokument. Bei der Frage ob von Excel oder ähnlichen Programmen die Style-Information oder Formatierung geändert werden soll, sollte dies abgelehnt werden, da SCHA.T.S.I die folgende Form benötigt:

term; cluster

ausdruck1;cluster

Eine beispielhafte Tabelle ist hier abgebildet.

term	cluster
buchungssatz	Buchhaltung
umsatzsteuer	Steuerwesen
haben	Buchhaltung
soll	Buchhaltung
Gewinn	Steuerwesen

3.2.2 SCHA.T.S.I starten

Wenn alle Voraussetzungen aus den vorherigen Kapiteln erfüllt sind kann SCHA.T.S.I. jetzt gestartet werden.

Auf Linux ist das Starten relativ einfach.

- 1. Öffne ein Terminal und navigiere in den Ordner, in dem die Dateien *docker-compose.yml* und *SCHATSI_RUN* entpackt wurden. Bei den meisten Distributionen öffnet die Tastenkombination Strg+Alt+T ein Terminal. Anschließend kann mit dem Befehl cd /PFAD/ZUM/DATEIORDNER der Ort erreicht werden.
- 2. ist der Ordner im Terminal erreicht (Prüfung: Kommando 1s eingeben und mit Enter bestätigen Es werden die beiden Dateien aufgelistet) reicht es folgenden Befehl zum Start des Programms auszuführen:

```
sudo ./SCHATSI_RUN
```

Sie werden einmal nach Ihrem Passwort gefragt. Dieses geben Sie einmal ein und bestätigen die Eingabe mit Enter

- 3. Es öffnet sich ein Fenster in welchem man sich durch die Ordner navigieren kann. Navigieren sie zum Ordner, welche ihre Paper enthält, wählen Sie Diesen aus und bestätigen sie Ihre Wahl.
- 4. SCHA.T.S.I beginnt seine Arbeit und durchläuft einmal den gesamten Prozess, erstellt einen Ordner names *Output* und vermeldet auch, wenn seine Arbeit beendet ist.

In diesem Ordner liegen mehrere .csv-Dateien, welche die Daten der Analyse enthalten.

MICHTIG: Diese Anleitung erklärt die Nutzung einer der Komponenten von SCHA.T.S.I., damit ist der vollständige Funktionsumfang noch nichgt abgeschlossen. Die

Prozesse wurden getrennt, um diese im einzelnen Wiederholen zu können. Wenn Sie nun die nächste Komponente nutzen wollen, dann müssen Sie diese nur aus Ihrem Repository herunterladen, entpacken und als Input-Ordner den gerade eben generierten Output-ordner auswählen, da die Ergebnisse der ersten Komponente der Input der zweiten sind (und nach dem gleichen Muster gilt dies auch für den Einsatz der 3. Komponente).

open Inhalt des Outputs, wird in Kapitel 5 erläutert, damit Sie diesen nachvollziehen können.

4. Installation unter MacOS (BETA)

4.1 Installation

Auf einem Mac gestaltet es sich relativ einfach, Docker als Basis von SCHA.T.S.I zum laufen zu bekommen.

Die folgende Website enthält eine detaillierte Anleitung für die Installation unter MacOS und eine Installationsdatei, sowohl für MacBooks mit Intel-Chip, als auch für MacBooks mit M1-Chip.

MacBooks konzipiert wurde und sich immernoch im Beta-Status befindet. Abstürze der Software und Fehlermeldungen können auftreten. Feedback ist erwünscht.

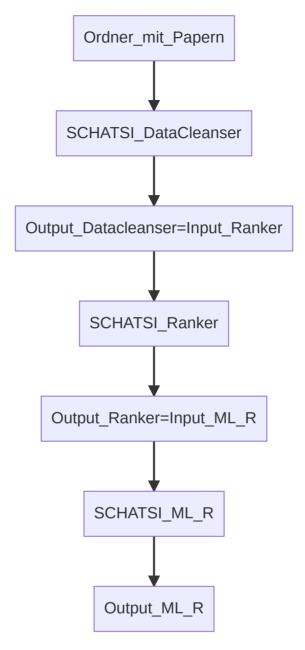
4.2 Einsatz

Ist Docker einsatzbereit, so kann nun die eigentliche Software bezogen werden.

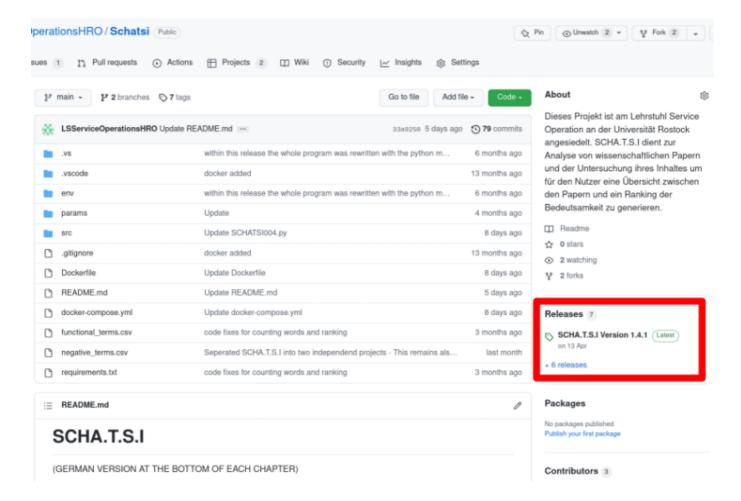
SCHATSI ist bis zur Version 1.4.1 in zwei, bei darauf folgenden Versionen in drei Komponenten aufgeteilt. Diese sind im jeweiligen Repository auf Github zu finden

- SCHATSI_DataCleanser
- SCHATSI Ranker
- SCHATSI Machine Learner R

Die Nutzung jedes dieser Komponenten funktioniert exakt gleich, sodass wir hier die Nutzung der ersten Komponente erklären. Sollen die anderen ebenfalls zum Einsatz kommen müssen diese nur aus dem jeweiligen Repository heruntergeladen werden. Als Input-Ordner sind dann jeweils die Output-Ordner der Vorgänger, wie im Diagramm erklärt auszuwählen, doch dazu gleich in der vollständigen Erklärung mehr.



Auf der Github-Seite des Projektes finden Sie auf der rechten Seite die aktuellste Releaseversion, jeweils für ein bestimmtes Betriebssystem. SCHATSI-Repository (siehe Bild)



Sie können die Software in einem .zip-Ordner Herunterladen.

Nun müssen Sie diese nur noch Entpacken. Enthalten sind:

- Ein Ordner namens "params"
- Eine .csv-Datei "functional terms.csv"
- Ab Version 1.4.2 eine .csv-Datei "negative terms.csv"
- Eine Readme-Datei mit Hinweisen zum Einsatz von SCHATSI
- Eine Datei namens "docker-compose.yml" Diese kann einfach ignoriert werden, da sie nur intern benötigt wird
- Eine Starter-Datei namens "SCHATSI RUN" mit welcher Sie SCHATSI zum laufen bringen

4.2.1 Vor dem Start

Nun können Sie SCHATSI das erste mal starten. Sicherlich haben Sie einen Ordner, in dem alle Paper enthalten sind, die sie analysieren möchten - das ist gut! Denn das wird unser Input-Ordner.

Michtig: In diesen Ordner müssen Sie noch zwei Dinge verschieben:

- Den Ordner "params"
- Die Datei "functional_terms.csv"

Weiterhin ist jetzt der Zeitpunkt gekommen, wo Sie sich überlegen müssen, welche Begriffe für ihre Analyse oder ihr Thema entscheidend sind. Gemeint sind dabei Begriffe, nach denen SCHA.T.S.I in den Papern suchen soll. Dabei können ganz unterschiedliche Ausdrücke gewäht werden und je konkreter diese zu ihrer Aufgabe passen desto hochwertiger wird die Arbeit von SCHA.T.S.I ausfallen.

Erlaubt sind dabei Ausdrücke aus einem, zwei oder drei Wörtern. Dies soll ermöglichen auch fachspezifische Ausdrücke wie "friction welding" oder "Internet of Things" zu erlauben.

1. Öffnen die dafür jetzt die Datei "functional terms.csv".

Sie sehen eine beinahe leere Tabelle abgesehen von den beiden Überschriften *term* und *cluster*.

- 2. Nun können Sie alle Ausdrücke, die Sie in ihre Untersuchung aufnehmen möchten untereinander in die Spalte *term* eintragen.
- 3. Die Spalte *cluster* dient dazu, Ausdrücke einzelnen Themengebieten zuzuordnen. Im späteren Verlauf von SCHA.T.S.I sollen diese genutzt werden, um die Paper auch in diese Bereiche einordnen zu können, sodass eine Übersicht entsteht, welches Paper Inhalte zu bestimmten Themenbereichen enthält.
 - Beispielhaft könnte eine Wirtschaftswissenschaftler, der eine wissenschaftliche Arbeit im Bereich Steuerwesen und Buchhaltung schreibt bestimmte Begriffe eher dem Bereich Steuern und andere in den Bereich Buchhaltung einordnen.
- 4. Speichern Sie das Dokument. Bei der Frage ob von Excel oder ähnlichen Programmen die Style-Information oder Formatierung geändert werden soll, sollte dies abgelehnt werden, da SCHA.T.S.I die folgende Form benötigt:

term; cluster

ausdruck1;cluster

Eine beispielhafte Tabelle ist hier abgebildet.

term	cluster
buchungssatz	Buchhaltung
umsatzsteuer	Steuerwesen
haben	Buchhaltung
soll	Buchhaltung
Gewinn	Steuerwesen

4.2.2 SCHA.T.S.I starten

→ Wichtig: Standardmäßig wird auf MacOS das Ausführen von Software aus unnbekannter Quelle blockiert. Da SCHA.T.S.I. als Software nicht aus dem offiziellen Softwarestore heruntergeladen wird die Ausführung blockiert. Um dies zu ändern tun Sie folgendes:

```
Menü --> Systemeinstellungen --> Sicherheit & Datenschutz --> Allgemein --> 🔐
```

Dieses "Schloss" muss geöffnet sein, damit Software aus unbekannten Quellen ausgeführt werden kann.

- 1. Anschließend können sie das Programm *Terminal* öffnen und mittels des Befehls cd /PFAD/ZUM/DATEIORDNER/ zum Ordner navigieren, welcher die Dateien "dockercompose.yml" und "SCHATSI RUN MacOS" enthält.
- 2. Wenn der Ordner im Terminal erreicht ist (Prüfung: Eingabe des Befehls ls --> Wiedergabe der enthaltenen Dateien) geben Sie folgenden Befehl ein, um SCHA.T.S.I zu starten:

```
sudo ./SCHATSI_RUN_MacOS
```

- 3. Sie werden aufgefordert ihr Passwort einzugeben. Tun Sie dies und bestätigen Sie ihre Eingabe mit Enter .
- 4. Es öffnet sich ein Fenster in welchem Sie ihre Ordnerstruktur durchsuchen können. Navigieren Sie zum Ordner, welcher die zu untersuchenden Paper enthält und bestätigen Sie ihre Auswahl.
- 5. SCHA.T.S.I beginnt seine Arbeit und durchläuft einmal den gesamten Prozess, erstellt einen Ordner names *Output* und vermeldet auch, wenn seine Arbeit beendet ist.

In diesem Ordner liegen mehrere .csv-Dateien, welche die Daten der Analyse enthalten.

MICHTIG: Diese Anleitung erklärt die Nutzung einer der Komponenten von SCHA.T.S.I., damit ist der vollständige Funktionsumfang noch nichgt abgeschlossen. Die Prozesse wurden getrennt, um diese im einzelnen Wiederholen zu können. Wenn Sie nun die nächste Komponente nutzen wollen, dann müssen Sie diese nur aus Ihrem Repository herunterladen, entpacken und als Input-Ordner den gerade eben generierten Output-ordner auswählen, da die Ergebnisse der ersten Komponente der Input der zweiten sind (und nach dem gleichen Muster gilt dies auch für den Einsatz der 3. Komponente).

des Outputs, wird in Kapitel 5 erläutert, damit Sie diesen nachvollziehen können.

5. Die Ergebnisse von SCHA.T.S.I

Egal welches Betriebssystem sie verwenden, nach einem erfolgreichen Ablauf von SCHA.T.S.I hat jede Komponente einen output-Ordner generiert, welcher die Ergebnisse in mehreren Dateien gesichert hat. Dabei sind auch Daten enthalten, die Sie für ihre eigenen Ergebnisse nicht direkt benötigen. Jedoch stellen diese eine Möglichkeit dar den Prozess nachzuvollziehen und stellen somit Transparenz her und können potenzielle Fehler aufdecken.

Insbesondere der SCHATSI_DataCleanser (die Erste Komponente, bis Version 1.4.1 einfach nur *SCHA.T.S.I* genannt) geneiert eine Vielzahl von Dateien, welche hier nacheinander erläutert werden.

Neben den neu erstellten Ergebnis-Dateien enhält der Ordner ebenfalls die Dateien *functional_terms.csv* und *negative_terms.csv*, welche aus dem Input-Ordner stammen und für die zweite Komponente ebenfalls von Bedeutung sind.

Neu generiert wurden die folgenden Dateien:

- SCHATSI_include.csv
- SCHATSI_runtime.csv
- SCHATSI_data_cleansing.csv
- SCHATSI terms.csv
- SCHATSI references.csv

Diese werden nun nacheinander kurz erläutert

5.1 SCHATSI include

Diese Datei gibt eine Übersicht, welche Paper von SCHATSI verarbeitet werden konnten und welche nicht. Die beiden Hauptgründe, warum eine Datei nicht verarbeitet werden konnte, sind zum Einen, dass es sich um eine Datei handelt, welche nicht im .pdf-Format vorliegt und zum anderen, dass beim Extrahieren des Textes ein Fehler aufgetreten ist.

Sollte ein Paper in den späteren Rankings und Visualisierungen nicht auftauchen, so kann hier nachvollzogen werden, ob ein Fehler auftrat.

5.2 SCHATSI runtime

In dieser Datei werden die Dauer der jeweils ablaufenden Teilprozesse erfasst. Es werden die einzelnen Teilaufgaben, eine Start- und Endzeit sowie eine Dauert der einzelnen Prozesse und des gesamten Programmes erfasst. Dies ist eher für die Entwickler des Programmes von Interesse, um die internen Prozesse von SCHA.T.S.I zu optimieren. Jedoch kann es auch für Sie als Nutzer interessant sein zu sehen, für welchen Aspekt der Analyse wie viel Zeit aufgebracht werden muss.

5.3 SCHATSI_data_cleansing

Diese Datei enthält laut aktuellem Stand (Version 1.4.1) eine tabellarische Übersicht über die einzelnen, erfolgreich untersuchten, Paper, deren Dateiformat und die Gesamtanzahl der Wörter im Textteil der Veröffentlichung. Diese ist später für das Ranking wichtig, um die Häufigkeit wichtiger Terme in Relation zur Gesamtlänge des Textes setzen zu können.

ZUKÜNFTIG sollen in dieser Datei auch die Metadaten der einzelnen Arbeiten automatisiert extrahier und gespeichert werden. Dazu zählen dann Autor, Jahr des Erscheinens sowie den Titel der Publikation, da häufig der Dateiname keinerlei Aufschluss über den publizierten Text liefert. Diese Metadaten bieten somit einen tabellarischen Überblick über die grundlegenden Fakten zu den einzelnen Papern. Diese Informationen werden später für die Verknüpfung und Visualisierung der Veröffentlichungen genutzt.

5.4 SCHATSI_terms

Diese Datei enhält die gesammelten Daten der Textextraktion und -Analyse. Dabei werden jeweils alle Terme (bestehend aus 1, 2 oder 3 elementaren Termen), sowie die Anzahl ihres Auftretens in einer Veröffentlichung erfasst.

Dabei handelt es sich, bereits um die gefilterten Terme, sodass Füllwörter (engl. "stopwords") nicht mehr enthalten sind. Diese Datei stellt die Grundlage für das Ranking der Dateien

untereinander dar, da sie die absoluten Zahlen über das Auftreten aller Terme in den Papern enthält.

5.5 SCHATSI_references

In der aktuellen Version befindet sich die Funktion für die Referenzextraktion in Wartung, sodass für jedes Paper jeweils der gesamte rohe Referenz-Text gespeichert wird.

Grund hierfür liegt in der Vielzahl von Ausgestaltungen und Zitationsstylen der einzelnen Paper, welche die Korrektheitsquote im Moment noch zu gering halten.

ZUKÜNFTIG sollen Hier jedoch für jedes Paper der Titel der Referenz, Erscheinungsjahr und der Referenzautor, für jedes untersuchte Paper aufgelistet werden. Dies wird nach einer Optimierung der Funktion in einer neuen Release-Version wiederveröffentlicht.