EIN WEBBASIERTES MULTIPLE-KERNEL-SYSTEM FÜR DIE AUTOMATISIERTE ERSTELLUNG VON ANALYTISCHEN BERICHTEN.

Dr. Alexander Wagner

AGENDA

- Einführung
- Das Ziel des Projektes
- Das Konzept, die Struktur und die Funktionalitäten des Systems
- Analytischer Bericht
- Die System-Funktionen
- Prozessablauf der automatisierten Erstellung des analytischen Berichts
- Das System GUI mit Steuerelementen
- GitHub Repositorium des Systems
- Content des GitHub-Repositoriums des Systems
- Web-App des Systems
- Bereitstellen (Deploy) der Web-App in Streamlit Community Cloud
- Weiterentwicklung des Systems
- Zusammenfassung
- Folgerung

Einführung

Diese Arbeit gehört zum Forschungsgebiet "Automatisierte Erstellung von analytischen Berichten als webbasierte Applikationen". Das Ziel des Projektes besteht in der Automatisierung von Routineoperationen zur Erstellung von analytischen Berichten über die Standardisierung der Prozessabläufe des Datenmanagements bis hin zur Ablage des Berichtes, der Software und Daten in der Web-Cloud und der Verwaltung des Client-Cloud Prozesses. Zur Anwendung kommt ein breites Spektrum an Werkzeugen der Datenverarbeitung, der Softwareintegration und der Steuerung von IT-Prozessen.

Schlüsselwörter: Automatisierte Erstellung von analytischen Berichten, Python, Streamlit, Streamlit.io-Cloud, WebApp, SAS, Autoit, MS OFFICE.



30.03.2023

Das Ziel des Projektes

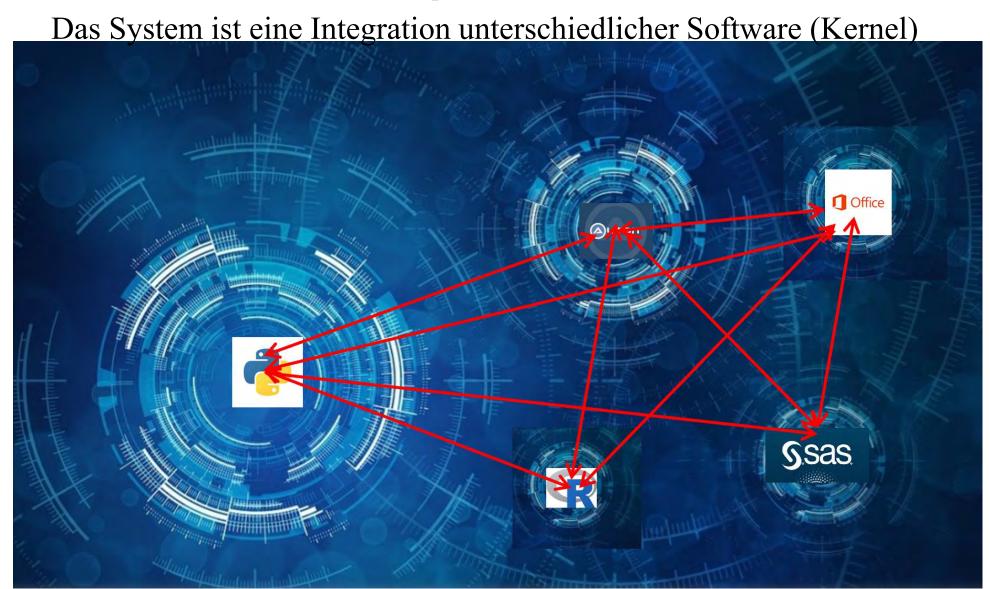
Das übergreifende Ziel dieses Projektes ist es, über die Implementierung einer automatisierten Erstellung von analytischen Berichten eine effiziente Steuerung der Kaskade aller Datenverarbeitungsschritte, die abschließende Ablage des Berichtes in der Web-Cloud und die Verwaltung des Client-Cloud Prozesses zu ermöglichen.

Die erste Aufgabe ist daher, alle Berichtsinformationen (Texte, Tabellen, Grafiken) in einem flexiblen Informations-System/GitHub Repositorium zu organisieren. Die zweite Aufgabe besteht darin, alle notwendigen analytischen und Daten-Management-Verfahren und Methoden in einem speziellen Programm-Tool zu integrieren in Form einer Projekt-GUI. In der dritten Phase wird der Bericht in der Web-Cloud abgelegt und die Verwaltung des Client-Cloud Prozesses durchgeführt.

Dieses Projekt ist eine Fortsetzung verschiedener früherer Projekte.



Das Konzept, die Struktur und die Funktionalitäten des Systems



Analytischer Bericht

Der analytische Bericht besteht aus den folgenden Standardelementen, und zwar:

- Frontseite
- Inhaltsverzeichnis
- Überschriften und Fußnoten
- Texte
- Tabellen
- Grafiken
- Listen

Das System entwickelt alle diese Komponenten mit Hilfe der Python DOCx-Bibliothek, SAS ODS und speziellen Autoit-Funktionen. Der Bericht wird in Form eines MS-WORD- und PDF-Dokumentes erstellt und im Repositorium GitHub abgelegt

Tabellen und Graphiken. Beispiel: DOCx-Tabelle

Die Tabellen werden im Regelfall mit Hilfe von Python DOCx-Bibliothek erstellt; Tabellen, die mit SAS Proc Report erzeugt werden, können ebenfalls problemlos in den Bericht integriert werden. Die Erstellung von Graphiken wiederum erfolgt mit Hilfe von Python-Bibliotheken, SAS/Graph und SAS® 9.4 ODS Graphics Prozeduren, Funktionen R-System, etc.

	Gesamt Population	Altersgruppen				
	Median	0-14	15-24	25-54	55-64	65+
Aruba	39.3	17.64	12.78	41.72	14.28	13.59
Australia	38.7	17.8	12.79	41.45	11.83	16.14
Austria	44.0	14.01	11.07	42.42	13.23	19.26
Azerbaijan	31.3	22.95	14.84	45.39	10.17	6.64
Bahamas	32.0	22.55	16.4	44.14	9.16	7.75
The Bahrain	32.3	19.08	15.65	56.04	6.28	2.95
Bangladesh	26.7	27.76	19.36	39.73	6.93	6.23
Barbados	38.6	17.97	12.74	44.06	13.43	11.81
Belarus	40.0	15.78	10.29	44.76	14.21	14.95
Belgium	41.4	17.16	11.34	40.05	12.86	18.58
	© Dr. Alexander Wagner. All Rights Reserved					

DOCx-Tabelle, Python-script (Fragment)

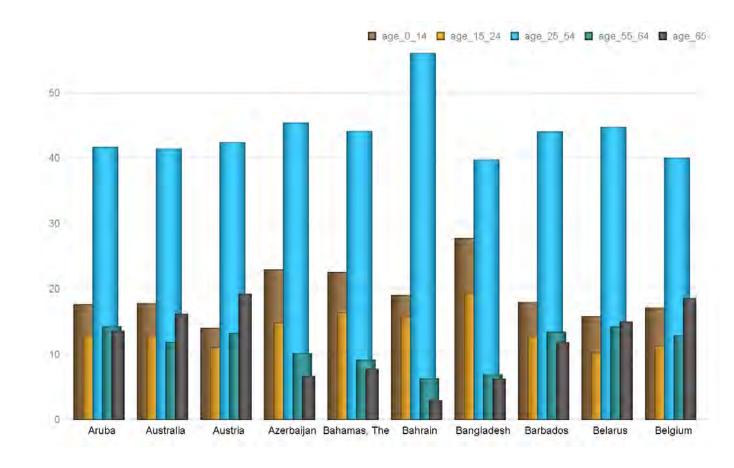
```
doc.add_heading("Tabelle No"+str(k+1), 2)
doc.add_heading("Altersverteilung für ausgewählte Länder nach
WHO: " + str(country.values[k, 1]), 3)
    t = doc.add table(n rows+2, n cols, style="Table Grid")
    set repeat table header(t.rows[0])
    for j in range(n_cols):
      if include index:
         ha=1
      else:
        t.cell(0,j).text = df3.columns[j]
    for i in range(2, n rows+2):
      for j in range(n_cols):
         if include index:
           t.cell(i, j).text = str(df3.values[i-2,j])
           t.cell(i, j).paragraphs[0].paragraph_format.alignment =
                     WD TABLE ALIGNMENT.RIGHT
         else:
          t.cell(i, j).text = str(df3.values[i,j])
      t.cell(i, 0).paragraphs[0].paragraph_format.alignment =
                     WD TABLE ALIGNMENT.LEFT
```

```
t.allow autofit = True
    t.style = 'Colorful Shading Accent 1'
    column_names0 = ["Country", "Median"]
    column_names = ["0-14", "15-24", "25-54", "55-64", "65+"]
    heading_cells = t.rows[0].cells
    for j in range(0, 1):
      heading_cells[j].text = column_names0[j]
       heading cells[i].width = widths[i]
    heading_cells = t.rows[1].cells
    for j in range(0, len(column_names)):
      heading_cells[j+2].text = column_names[j]
      heading cells[i].width = widths[i]
      cell = t.cell(1, j+2)
      cell.paragraphs[0].alignment =
          WD PARAGRAPH ALIGNMENT.CENTER
      cell.paragraphs[0].alignment =
          WD PARAGRAPH ALIGNMENT.CENTER
      paragraph = cell.paragraphs[0]
run.add_picture(".\\INDATEN\\who-logo-png.png", Cm(1), Gm(1))
```

Beispiel: SAS-Graphik

Ein SAS-Macro für die Erstellung der Grafiken

```
%MACRO GRAPX(G=);
   Title:
   %COLOR(K=&G):
   ods graphics / reset width=8in height=5in noscale noborder
   imagename="GRAPH&G";
   proc sgplot data=DATEN.DEMOGRA(where=(GR=&G)) nowall
   noborder:
    styleattrs datacolors=(&STRING);
   vbar country / response=age_0_14 nostatlabel barwidth=0.6 nooutline dataskin=pressed
     discreteoffset=-0.15 baselineattrs=(thickness=0);
   vbar country / response=age_15_24 nostatlabel barwidth=0.5 nooutline dataskin=pressed
     discreteoffset=-0.05 baselineattrs=(thickness=0);
   vbar country / response=age_25_54 nostatlabel barwidth=0.4 nooutline dataskin=pressed
      discreteoffset=0.05 baselineattrs=(thickness=0);
   vbar country / response=age_55_64 nostatlabel barwidth=0.3 nooutline dataskin=pressed
      discreteoffset=0.15 baselineattrs=(thickness=0);
   vbar country / response=age_65 nostatlabel barwidth=0.2 nooutline dataskin=pressed
       discreteoffset=0.25 baselineattrs=(thickness=0);
       keylegend / location=outside position=topright noborder
   valueáttrš=(size=8 color=gray);
   xaxis fitpolicy=thin display=(nolabel noticks noline) valueattrs=(size=8 color=BLACK);
   yaxis grid display=(noline noticks nolabel noline) valueattrs=(size=8 color=gray);
   run;
```



Die System-Funktionen

Das System realisiert die folgenden Funktionen:

- Mit Hilfe von Python, SAS, R, etc. werden Tabellen und Grafiken für die Integration in den Bericht erstellt. Alle diese Berichtselemente werden lokal in einem speziellem Projekt-Directory abgelegt.
- Die vorgefertigten Text-Abschnitte für die Integration in den Bericht werden in einem speziellen Projekt-Directory als Word-Dokumente abgelegt.
- Mit Hilfe der Python DOCx-Bibliothek wird der Bericht erstellt.
- Der Bericht wird zunächst lokal auf einem speziellem Projekt-Directory abgelegt, danach in das ein GitHub Repositorium gepusht.
- Die Korrektur bzw. Nachbesserung der Programm-Codes und Berichts wird in der Regel lokal im GUI-Fenster vorgenommen und danach nach GitHub gepusht.
- Die spezielle Autoit UDF und Python Def realisieren mit Hilfe von speziellen Autoit Steuerelementen (Buttons, Pop- und Contents-Menu) die Download- bzw. Push/Upload Daten Transfer-Prozesse und Update des Systems.

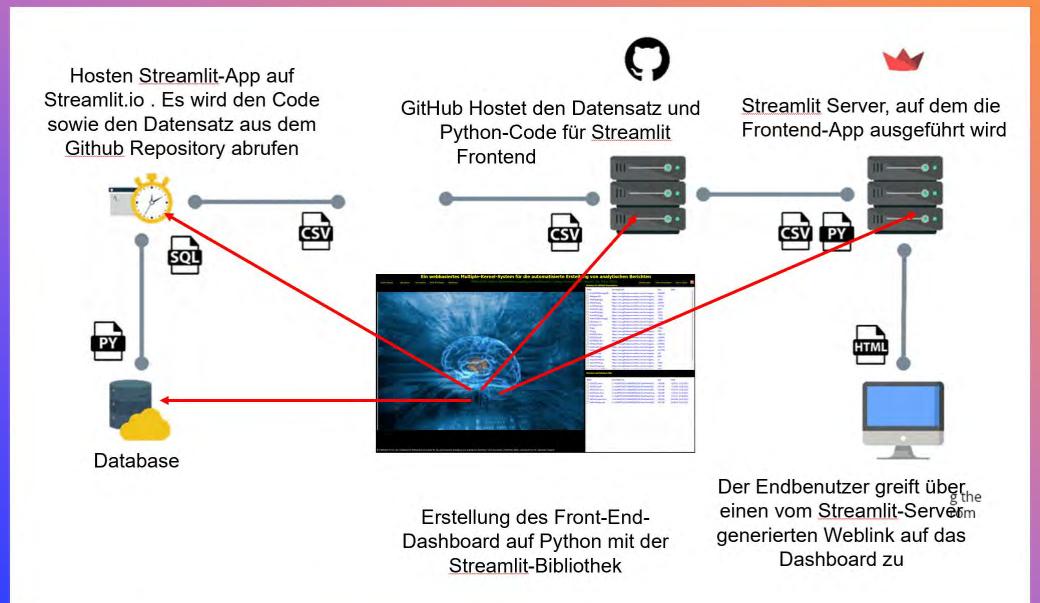
Prozessablauf der automatisierten Erstellung des analytischen Berichts

Der Prozessablauf für die Erstellung des analytischen Berichts erfolgt vollautomatisch. Durch Standard Point-and-Click Manipulationen wird der gesamte Prozess mit Hilfe von GUI-Elementen gesteuert. Die nachfolgende Folie stellt den Prozessablauf der automatisierten Erstellung des analytischen Berichts dar.

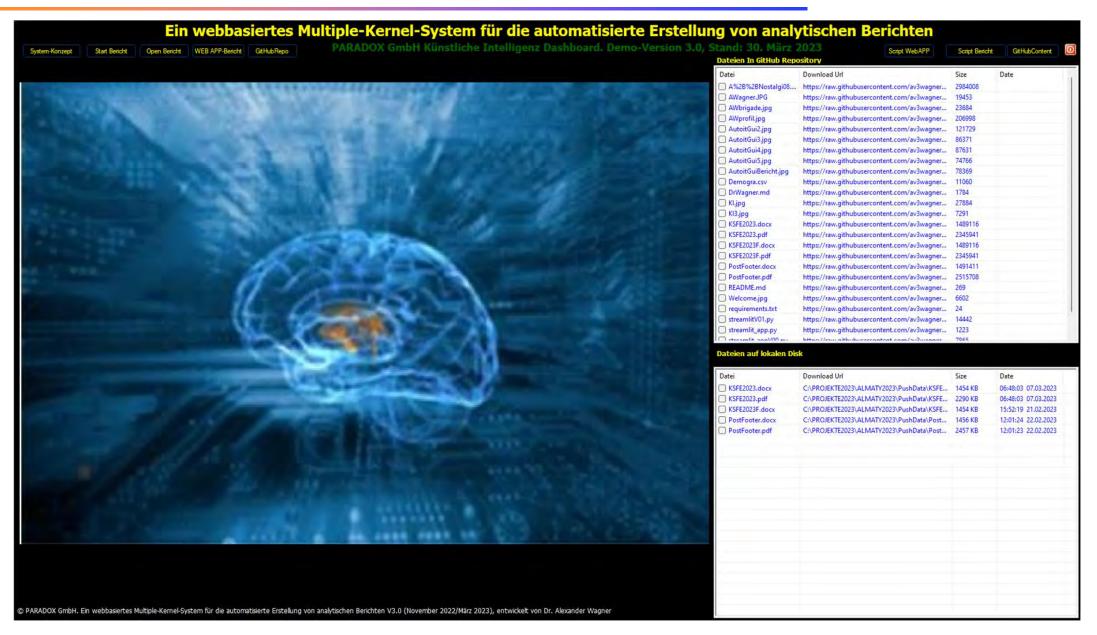
Der Prozessablauf besteht aus den folgenden Etappen bzw. Elementen:

- System-GUI als Kommando-Zentrale
- Data-Pushing
- Deployment der Web-App
- GitHub-Repositorium
- Streamlit.io-Cloud
- Produktion des Systems seitens der Endanwender

Schematische Darstellung des Anwendungsbetriebs in Web-Plattformen

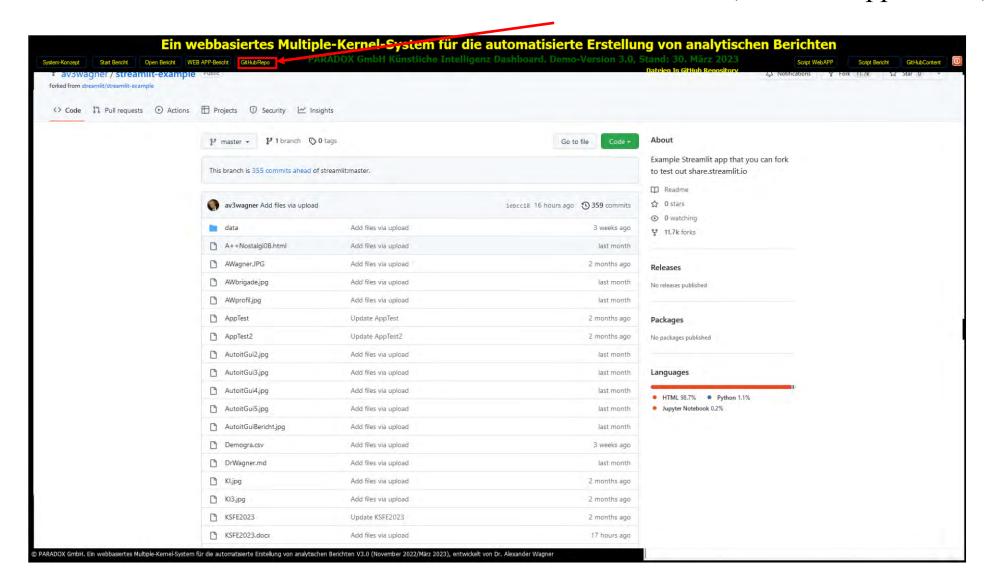


Das System GUI mit Steuerelementen



GitHub Repositorium des Systems

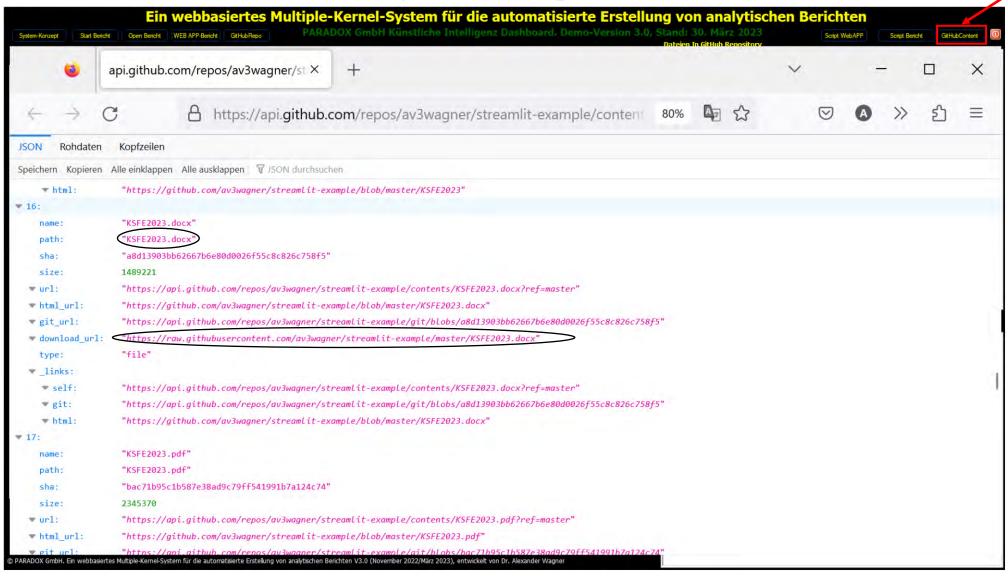
Alle Programme und Daten des Systems liegen im GitHub-Repositorium. Mit Hilfe von GUI-Buttons und **Context Menus** kann man die Dateien hoch- und herunterladen, die WebApp starten, etc.



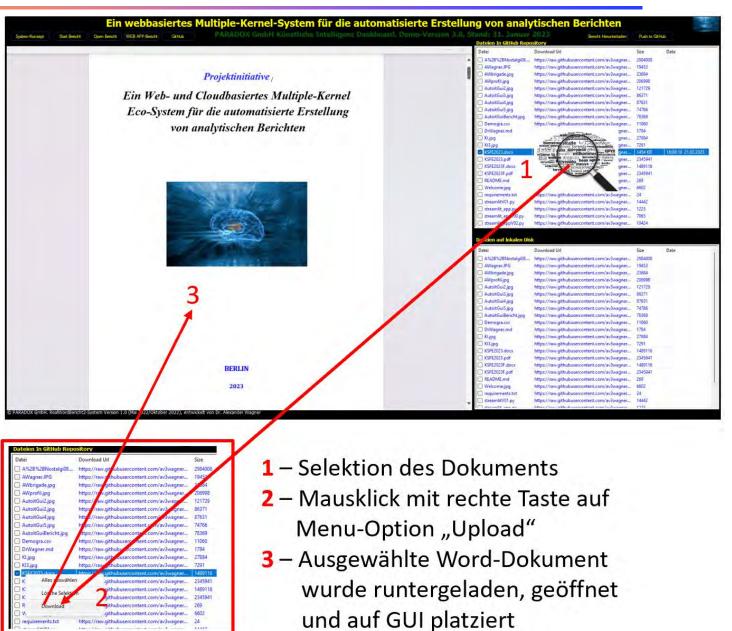
14

Content des GitHub-Repositoriums des Systems

Der Button "GitHubContent" öffnet das System des Repositorium-Content auf GUI.



Herunterladen des Berichts vom GitHub-Repo ins GUI-Fenster



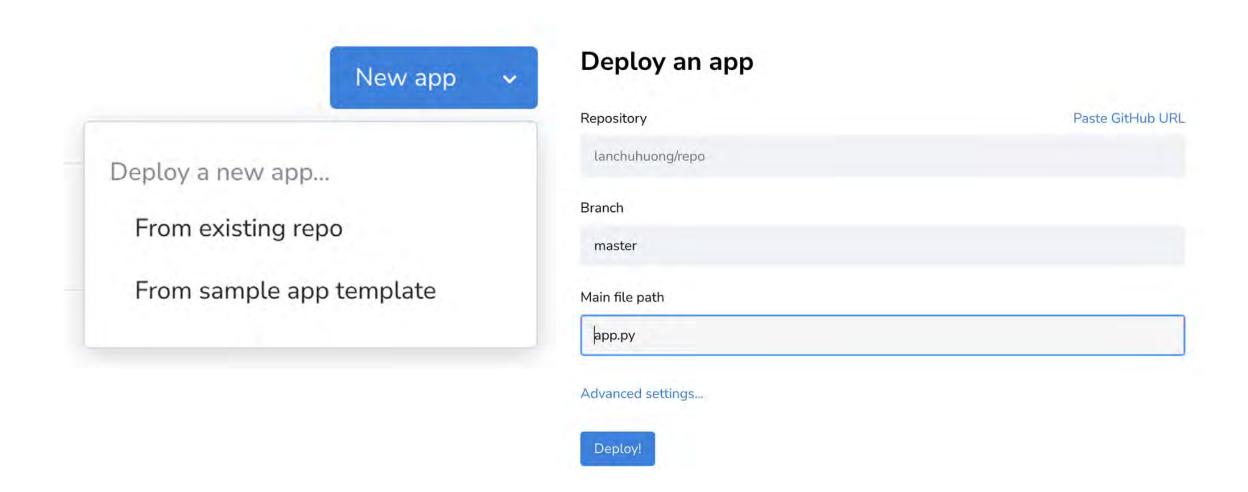
Web-App des Systems

Das System verwendet die Streamlit Community Cloud, eine Open-Source-Plattform für die Community, um Streamlit-Apps bereitzustellen und freizugeben. Der Prozess zum Bereitstellen der Streamlit-App in Streamlit Cloud funktioniert in drei einfachen Schritten:

- 1. Generierung der requirements.txt, die alle Abhängigkeiten für das Projekt auflistet. Dies wird benötigt, um Streamlit mitzuteilen, welche Bibliotheken installiert werden müssen, um den Code auszuführen. Hierzu kann man Pipregs verwenden, um die Datei basierend auf den Importen im Programm Code automatisch zu generieren.
- 2. Das Hosting des Codes erfolgt in einem öffentlichen GitHub-Repository. Hierbei ist sicherzustellen, dass das Repository die requirements.txt Datei enthält.
- 3. Registrieren an der share.streamlit.io.

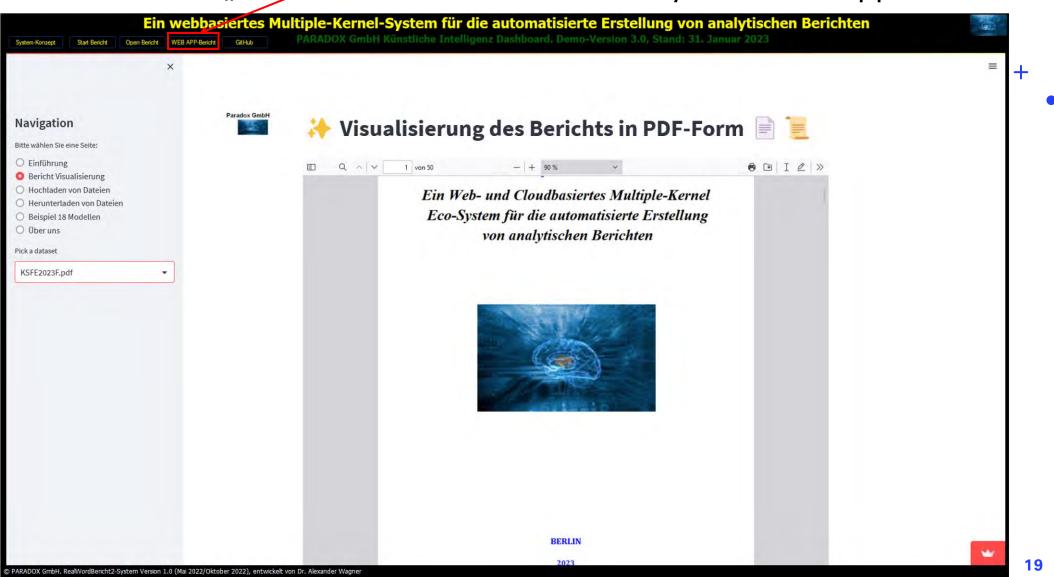
Die nachfolgende Folie stellt das Bereitstellen (Deployment) der WebApp in der Streamlit Community Cloud dar.

Bereitstellen (Deployment) der Web-App in Streamlit Community Cloud



Streamlit Web-App auf der GUI des Systems

Beim Klicken auf Button "WEB APP-Bericht" öffnet das System Web-Applikation auf GUI.



Weiterentwicklung des Systems

Das System funktioniert als Streamlit.io Web-Applikation und konnte mit seiner Effizienz vielfach überzeugen. Dennoch bestehen Ansatzpunkte für Nachbesserung und Weiterentwicklung. Die folgenden Erweiterungs- und Optimierungspotentiale werden priorisiert: •

- Verwendung des funktional und technologisch überlegenen Django-System (https://www.djangoproject.com/) statt Streamlit.
- Verwendung von funktional und technologisch überlegenen Cloud-Systemen (Amazon EC2, Microsoft Azure Cloud, etc.) statt Streamlit.io
- Korrektur bzw. Nachbesserung des Berichts in Cloud von mehreren Anwendern gleichzeitig
- Verwendung der JSON Steuer/Config-Datei für voll automatisierte Berichtserstellung
- Automatisierung des Uploads/Download Prozesses z.B. mit Hilfe von Windows-Task

ZUSAMMENFASSUNG



Es wurde ein webbasiertes Multiple-Kernel-System für die automatisierte Erstellung von analytischen Berichten entwickelt und in die Produktion gebracht.

Das System hat seine programm-technische und technologische Effizient gezeigt, funktioniert robust und kann für die Erstellung der analytischen Berichten verschiedener Art verwendet werden.

Weiterentwicklung und Nachbesserung des Systems folgt.

FOLGERUNG



In Kürze startet ein neues nichtkommerzielles Projekt "Automatisierung der Wissenschaftliche Forschung im Gesundheitswesen, Medizin und Pharmabranche mit Hilfe von KI-Methoden".

Das Projekt wird schätzungsweise 6-8 Monate dauern.

Wer jemand Interesse an einer spannenden hochintelligenten Arbeit mit abschließender wissenschaftlicher Publikation hat, dann herzlich Willkommen ins Team!

TEAM



Dr. Alexander Wagner, Teilnehmer an KSFE-2015

30.03.2023



Literatur

- [1] Autolt website. https://www.autoitscript.com/site/
- [2] User Defined Function. https://de.wikipedia.org/wiki/User_Defined_Function
- [3] Base SAS. SAS Notes and Concepts for ODS, The RTF Destination.

https://support.sas.com/rnd/base/ods/templateFAQ/Template_rtf.html

[4] Lauren Haworth, Applying Microsoft Word Styles to ODS RTF Output.

http://www2.sas.com/proceedings/sugi30/043-30.pdf

[5] Kirk Paul Lafler Output Delivery System (ODS) – Simply the Basics. Software Intelligence Corporation, Spring Valley, California.

http://www.scsug.org/wp-content/uploads/2012/11/Output-Delivery-System-ODS-%E2%80%93-Simply-the-Basics-SCSUG-2012.pdf

[6] Carol Matthews, Elena Kalchenko, Pretty Please?! Making RTF Output "Pretty" with SAS. United Biosource Corporation, Blue Bell, Pennsylvania.

https://www.pharmasug.org/proceedings/2013/IB/PharmaSUG-2013-IB08.pdf

[7] SAS® 9.4 ODS Graphics: Procedures Guide, Sixth Edition.

http://support.sas.com/documentation/cdl//en/grstatproc/69716/HTML/default/viewer.htm#titlepage.htm

- [8] python-docx 0.8.11 documentation. https://python-docx.readthedocs.io/en/latest/
- [9] python-docx-template. https://docxtpl.readthedocs.io/en/latest/

Literatur (Fortsetzung 1)

- [9] python-docx-template. https://docxtpl.readthedocs.io/en/latest/
- [10] Automate Word Document (.docx) With Python-docx And pywin32
- [11] Jay, Automate Word Document (.docx) With Python-docx And pywin32. https://pythoninoffice.com/automate-docx-with-python/
- [12] Dr. Rik Voorhaar, How to edit Microsoft Word documents in Python. https://www.rikvoorhaar.com/python-docx/
- [13] PyGithub- documentation https://pygithub.readthedocs.io/en/latest/introduction.html
- [14] Abdou -Rockikz, Python- script- to -create GitHub- repository. https://iq.opengenus.org/create-github-repository-python/
- [15] Code- for- How- to- Use- Github- API- in- Python- Tutorial.
- https://www.thepythoncode.com/article/using-github-api-in-python
- [16] Streamlit-Dokumentation: https://docs.streamlit.io/
- [17] Marc, Introduction to Streamlit: Building Interactive Web Apps using Python. https://marccodess.medium.com/introduction-to-streamlit-building-interactive-web-apps-using-python-62c250ba858d
- [19] Alan Jones, Streamlit from Scratch: Embedding Images, Video and Audio. https://towardsdatascience.com/streamlit-from-scratch-embedding-images-video-and-audio-8b2e8b98fad4

Literatur (Fortsetzung 2)

- [20] Navid Mashinchi, The Current State of COVID-19 From 3 Different Perspectives. https://towardsdatascience.com/the-current-state-of-covid-19-from-3-different-perspectives-3fbaabcd0348
- [21] Ponshriharini, Deploy a Machine Learning model using streamlit. https://medium.com/featurepreneur/deploy-a-machine-learning-model-using-streamlit-a7ef163c19b0
- [22] Sharad Mittal, Creating a Web app of a ML model using Streamlit. https://sharadmittal.hashnode.dev/creating-a-web-app-of-a-ml-model-using-streamlit.
- [23] Deliver Your Customer Behavior Analysis Report with Streamlit. https://medium.com/octopus-id-data/deliver-your-customer-behavior-analysis-report-with-streamlit-36f052d0a43c
- [24] A. Wagner, K.Krobot. Aktuariell diskontiertes Überleben aus webbasierten Rohdatenbanken, KSFE-2009. https://saswiki.de/display/KONFERENZEN/KSFE+2009?preview=/21267185/21267221/13.KSFE-2009-Wagner-Webbasierte_Rohdatenbanken.pdf#KSFE2009-DataMining/WebMining/Datenmanagement
- [25] A. Wagner, J. Holstiege. Ein SAS basiertes System zur automatisierten Auswertung und Berichterstellung von klinischen Studien, KDFE-2019.
- https://saswiki.de/display/KONFERENZEN/KSFE+2019?preview=/19726371/19726410/23_KSFE_2019_Wagn er -
- <u>Ein SAS basiertes System zur automatisierten Auswertung und Berichterstellung von klinischen Studien.pdf</u>