# Projektdefinition

## Randbedingungen:

* Beginn: 03.06.2019
* Geplantes Ende: 18.07.2019
* Standort: Variabel
* Projektleitung: Sonia Meehan
* Projektteam: Kai Haubrich, Marco Kovacic, Samuel Mezger, Marija,   
   Aleksandar Kostov, Théo Roncoletta
* Budget: keines

## Projektziel (was ist zu erreichen):

Dieses Projekt dient dazu eine Webapplikation zu entwickeln die es einem User möglich macht Wikipedia als einen „force-directed graph“, also einem Graph der sich durch simulierte Physik automatisch ausrichtet, einzusehen und mit ihm zu interagieren. Dies wird durch vier Hauptteile realisiert:

1. einen Webcrawler, welcher die Wikipedia-API und andere Quellen benutzt um diese Daten zu erheben,
2. eine Datenbank die die Daten des Webcrawlers als XML einspeichert und diese leicht zugänglich für andere Programmteile macht,
3. eine „Data-Refinery“ welche die Daten des Webcrawlers aufarbeitet (Überflüssige Verweise streichen, doppelte Daten löschen etc.)
4. und eine GUI welche die Daten aus der Datenbank abfragt und in einem „force-directed graph“ visualisiert, der anhand von Filter-Parametern vom User verändert werden kann

Der User kann hierbei den Startpunkt wählen von welchem aus Wikipedia-Link-Daten gecrawlt werden und einen Filter anhand welchem die gigantischen Datenmenge von Wikipedia gefiltert wird. Aktuell besteht der Filter allein aus Aufrufzahlen, dies kann aber noch in der Zukunft erweitert werden. Des weiteren soll der User selbst Noden und Kanten in den Graph einfügen um die Daten anhand seines eigenen Ermessens zu erweitern.

## Projektzielsetzung (wie wird dies erreicht):

* Die verfügbaren Ressourcen (Projektteilhaber) werden anhand ihrer Erfahrung und ihrer Willigkeit auf das Projekt gleichmäßig aufgeteilt.
* Eine große und ausgiebige Testphase wird festgeleg, damit garantiert werden kann, dass die verschiedenen Programmteile alleine und vor allem zusammen funktionieren.
* Klar definierte Schnittstellen festlegen, damit die Kommunikation zwischen den einzelnen Bestandteilen möglichst klar definiert ablaufen kann.

## Was gehört nicht zum direkten Projektziel („Nice to have“ aber nicht unbedingt nötig)

* Verschiedenste wechselbare Darstellungsmodi der Daten wie ein 3D-GUI, textuelle Darstellung
* Gruppierung der Daten anhand verschiedenster Tags (z.B. Säugetier, Reptil, Politiker etc.)
* Höchstoptimisierte Datenbank welche hunderte von GB an Daten ohne Probleme dynamisch speichern, ergänzen und zurückliefern kann

Visualisierung des   
Programmaufbaus:

Edit and add

User

Data funnel

Datenbank

Internet

Crawler

Screen

GUI