HUGO

Highly Useful Graphical Output

Durchgeführt im Rahmen des Moduls „Big Data Engineering“ des Master Studiengangs Wirtschaftsinformatik an der Fachhochschule Münster.

Gruppenmitlieder:

Ben Lohrengel Matr. Nr.:

Sebastian Manns Matr. Nr.:

André Meier Matr. Nr.:

Prüfer

Erstprüfer: Lars George

Zweitprüfer:

Abgabedatum: 25.02.2016

Inhaltsverzeichnis

[Einleitung 3](#_Toc443831954)

[Idee 3](#_Toc443831955)

[Input 3](#_Toc443831956)

[Wikimedia 3](#_Toc443831957)

[Verarbeitung 3](#_Toc443831958)

[Output 3](#_Toc443831959)

[Verwendete Tools 3](#_Toc443831960)

[HBase 3](#_Toc443831961)

[Map Reduce 3](#_Toc443831962)

[Hadoop File System 3](#_Toc443831963)

[Oozie 3](#_Toc443831964)

[Ergebnisse 4](#_Toc443831965)

[Visionen 4](#_Toc443831966)

[Fazit 4](#_Toc443831967)

# Einleitung

## 

## Idee

Das Internet enthält heute eine unvorstellbare Menge an Videos. Im Dezember 2014 wurden alleine auf der Videoplattform YouTube, pro Minute, 12,5 Tage Material hochgeladen. All diese Videos sind einmalig und identifizieren sich aus einer bestimmten Kombination von Pixeln. Die darin enthaltenen Farbwerte machen den gesamten Videostream unverwechselbar.

Es ergibt sich nun die Fragestellung inwieweit das Bildmaterial der Videos auf seine Farbwerte hin analysiert und ausgewertet werden können. Als Ergebnis sollen zwei Farbwerte visualisiert werden. Zum einen ein durchschnittlicher Farbwert, basierend auf allen im Stream vorkommenden Pixel und zum anderen der am häufigsten vorkommende, also dominierende Farbwert. Das Ziel ist es eine Plattform für Nutzer zu schaffen, auf der sie die Ermittlung durchführen lassen können. Dazu ist es zunächst nötig den Link einer, prinzipiell beliebigen, Videoquelle einzugeben. Nach der Betätigung eines Buttons, soll sich das System eigenständig den Viedostream herunter laden und ihn lokal ablegen. Die darauffolgende interne Verarbeitung soll mit verschiedenen Tools aus dem Bereich des Big Data Engineering erfolgen. Zum Abschluss des Analyseprozesses soll dem Nutzer das Ergebnis in Form einer Durchschnittsfarbe sowie die ermittelte dominierende Farbe auf der Plattform präsentiert werden.

Ziel dieses Projektes namens “Highly Useful Graphical Output” kurz HUGO, ist es nun, die zuvor beschriebene Vorgehensweise zu realisieren.

## Input

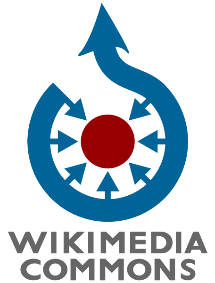
Zu Beginn des Prozesses wird ein Videostream benötigt, den ein Nutzer auf seine dominante bzw. Durchschnittsfarbe analysieren lassen möchte. Dabei müssen jedoch zwei Voraussetzungen erfüllt sein.

1. Der Stream muss online verfügbar und über einen direkten Downloadlink erreichbar sein.
2. Die Onlineplattform auf der sich der Stream befindet muss über eine frei zugängliche API erreichbar sein.

Ist beides erfüllt, so kann der Nutzer den http-Link zum Video kopieren und in der Eingabemaske des HUGO einfügen.

Jedoch wird bei den meisten Videoplattformen der zuvor erwähnte Punkt 2 leider nicht erfüllt. In der Regel sind es datenschutzrechtliche Gründe die große Anbieter wie „Youtube“ dazu bewegen keine freie API mehr anzubieten. Die Wikimedia Foundation hingegen bietet auf ihrer Website „wikimedia.org“ Videostreams an die basierend auf freien Lizenzen genutzt und vor allem mittels ihrer API „MediaWiki“ von Drittsystemen verwendet werden können.

**Wikimedia Commons**

Wikimedia definiert sich als weltweite Bewegung für freies Wissens. Unter dem Dach der gemeinnützigen Wikimedia Foundation werden verschiedenste Wikimedia-Projekte wie Wikipedias angeboten. Darunter befindet sich ebenfalls Wikimedia Commons, welche zum Ziel hat, als zentrales Medienarchiv in verschiedenen Sprachen für die Wikipedias zu dienen. Die Medienplattform stellt somit gemeinfreie und frei-lizensierte Medieninhalte wie Bilder, Audio- und Videodateien bereit. Derzeit werden über 30 Millionen Dateien zur Verfügung gestellt.

Die Gemeinfreiheit sowie das freie Lizensierungsmodell gestatten es Videoframes ohne die Verletzung von Persönlichkeits- und Datenschutzrechten sowie des Urheberrechts herunterzuladen und nicht-kommerziell zu nutzen. Die Videoinhalte können dabei über einen direkten Link erreicht werden. Das bedeutet, dass der Link nicht, wie bei anderen Plattformen, auf die Webseite des Videos führt, sondern direkt zum Download verwendet werden kann. Somit erfüllt Wikimedia Commons die beiden zuvor erwähnten Voraussetzungen für eine verwendbare Videoquelle. Dadurch wird diese Plattform im Folgenden als Videoherkunft für das HUGO-Projekt verwendet.

Abbildung Logo Wikimedia Commons

## GUI

Als zentrale Anlaufstelle zur Interaktion mit dem HUGO-Projekt, dient eine einzige Webseite. Hier soll zunächst eine Eingabemaske zu sehen sein. Der zuvor kopierte Link zum Videostream ist dort einzufügen. Im nächsten Schritt ist der Nutzer dazu angehalten anzugeben welches Ergebnis er haben möchte. Es kann entweder die dominierenden Farbe oder die Durchschnittsfarbe ermittelt werden. Nach der Betätigung eines Buttons zur Bestätigung, wird das entsprechende Video im Hintergrund vom System heruntergeladen und lokal gespeichert. Anschließend soll die interne Verarbeitung ebenfalls unsichtbar durchgeführt werden. Für den Nutzer ist auf der Webseite ein Lade- bzw. Fortschrittsbalken zu visualisieren, anhand dieser den Prozessfortschritt erkennet. Zudem kann die Beendigung einzelner Prozessabschnitte, wie bspw. der erfolgreiche Download durch Ausgabe eines Textes dem Nutzer erscheinen. Zum Abschluss der Verarbeitungskette sind die Ergebnisse in Form einer Grafik wieder auf der Webseite angezeigt.

## Output

In den vorherigen Abschnitten erfolgte die Definition des erwarteten Inputs, der zukünftig verwendeten Quelle sowie der graphischen Oberfläche. In diesem Abschnitt soll es nun kurz um eine Beschreibung des erwarteten/möglichen Ergebnisses bzw. Outputs gehen.

Nachdem eine Videoquelle erfolgreich eingelesen, verarbeitet und analysiert wurde soll auf der zentralen Weboberfläche die Ausgabe erfolgen. Abhängig von der zuvor getroffen Wahl des Nutzer, sind die Farbwerte hinsichtlich eines Ergebnisses zu ermitteln und in Form einer Grafik zu visualisieren. Dabei stehen folgende Analysemöglichkeiten zu verfügung.

1. **Durchschnittsfarbe** (average Color)

Ziel bei der Ergebnisform „Durchschnittsfarbe“ ist es den durchschnittlichen Farbwert pro Frame zu ermitteln.

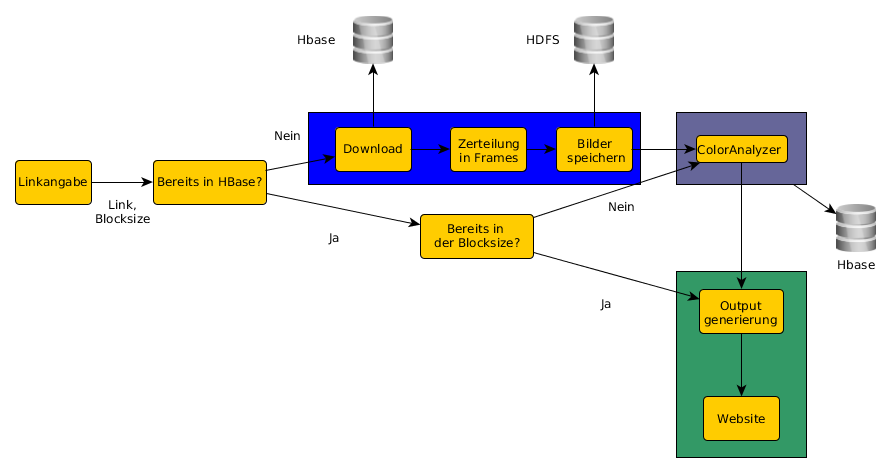
1. **Dominierende Farbe** (dominant Color)

Bei der Ergebnisform „Dominierende Farbe“ handelt sich um die Ermittlung jenes Farbwertes pro Frame, der am häufigsten vorkommt.

Unabhängig von dem gewünschten Ergebnis, sind die ermittelten RGB-Werte eines jeden Frames abzulegen. Zur abschließenden Darstellung müssen die einzelnen Durchschnitts- bzw. dominierenden Farbwerte nacheinander als Pixel ausgegeben und in einer Grafik zusammengefasst werden. Die so entstandene Grafik bildet somit eine chronologische Aneinanderreihung der Farbwerte eines Videostreams. Zuletzt ist dann die so entstandene Grafik auf der Weboberfläche auszugeben. Der Nutzer erhält damit das von ihm gewünschte Ergebnis und die Arbeit des HUGO-Projektes ist beendet.

* Webseite
* Video
* Blocksize
* Analyseverfahren
  + Durchschnittsfarbe
  + Dominante Farbe

## Verarbeitung



* Prüfen
* Ablaufmodell
* Download
* Zerlegen
* MapReduce Job
* Out Image generieren und abspeichern

## Output

* Webseite
* Bild wo für jeden Frame die Farben aneinandergereit sind

# Verwendete Tools

## HBase

## Map Reduce

* Pixel von Bildern verarbeiten
* Durchschnittswerte

## Hadoop File System

* Für Videodateien und Frames

## Oozie

* Workflow
* Tracking um den Status des Jobs an die Webseite zurückzugeben

Xuggler

* Was das Framework macht.

# Ergebnisse

# Visionen

* Was kann man noch machen
* Der Nutzer kann die Abtastrate selber wählen
* Evtl übertragbar auf den sound der videos

# Fazit

http://de.statista.com/statistik/daten/studie/207321/umfrage/upload-von-videomaterial-bei-youtube-pro-minute-zeitreihe/