

**Philosophische** Fakultät III

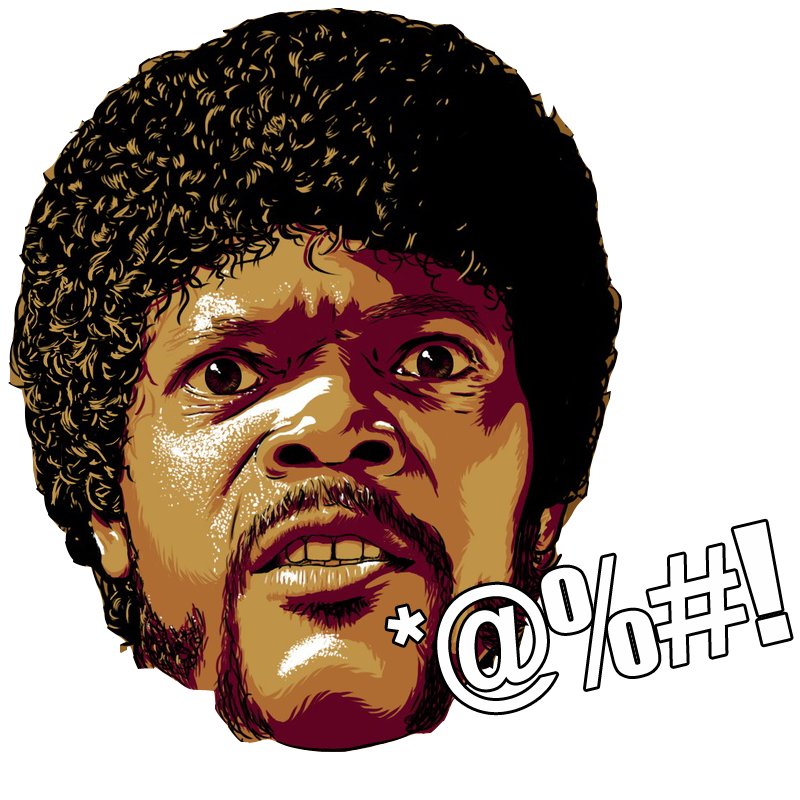
Sprach-, Literatur- und Kulturwissenschaften

Institut für Information und Medien, Sprache und Kultur (I:IMSK)  
Lehrstuhl für Medieninformatik

Projektseminar Mediengestaltung I: Informationsvisualisierung

MEI-M 05.3: Projektseminar Mediengestaltung

Sommersemester 2017



Cursewords and Deaths  
in Tarantino Movies

|  |  |
| --- | --- |
| Amanda Breuer Matr.Nr.:1769865  Medieninformatik, Informationswissenschaft, Philosophie  5. Semester B.A.  amanda.breuer@stud.uni-regensburg.de | Alicia Fritsche Matr.Nr.: 1781710  Medieninformatik, Vergleichende Kulturwissenschaft, Bildende Kunst  4. Semester B.A.  alicia.fritsche@stud.uni-regensburg.de |
| Louis Ritzkowski Matr.Nr.: 1816731  Medieninformatik, Informationswissenschaft 4. Semester B.A.  louis.ritzkowski@stud.uni-regensburg.de  Abgegeben am 31.07.2017 | Teresa Then Matr.Nr.: 1720261  Medieninformatik, Medienwissenschaft 5. Semester B.A.  teresa.then@stud.uni-regensburg.de |

**Abbildungsverzeichnis**

[Abb. 1: Bubble Chart – All Curses 6](#_Toc489282317)

[Abb. 2: Bubble Chart – Curses by Movie 6](#_Toc489282318)

[Abb. 3: Bubble Chart – Curses by Type 7](#_Toc489282319)

[Abb. 4: Line Graph 8](#_Toc489282320)

[Abb. 5: Sankey Flow 9](#_Toc489282321)

[Abb. 6: Bar Chart 10](file:///C:\Users\Amanda\Desktop\Projektdoku.docx#_Toc489282322)

[Abb. 7: Movie Ratings 12](file:///C:\Users\Amanda\Desktop\Projektdoku.docx#_Toc489282323)

Inhalt

[1. Einleitung 3](#_Toc489281828)

[2. Umsetzung 4](#_Toc489281829)

[2.1 Bubble Chart 4](#_Toc489281830)

[2.2 Line Graph 7](#_Toc489281831)

[2.3 Sankey Flow 8](#_Toc489281832)

[2.4 BarChart 9](#_Toc489281833)

[2.5 Movie Ratings 11](#_Toc489281834)

[3. Gestaltungskonzept 12](#_Toc489281835)

[Projektmanagement 13](#_Toc489281836)

# Einleitung

Die vorliegende Projektarbeit stellt eine Informationsvisualisierung von Schimpfwörtern und Todesfällen in ausgewählten Tarantino-Filmen dar. Diese wurde mittels der JavaScript-Bibliothek D3 implementiert.

Bei dem hierfür verwendeten Datensatz handelt es sich um die Daten aus sieben Filmen ("Reservoir Dogs", "Pulp Fiction", "Jackie Brown", "Kill Bill: Vol. 1", "Kill Bill: Vol. 2", "Inglourious Basterds" und "Django Unchained") des Filmemachers Quentin Tarantino. Dieser ist bekannt für seine Werke, die viele Kraftausdrücke und Gewaltszenen beinhalten. Oliver Roeder machte es sich zur Aufgabe eben diese Szenen zu analysieren, indem er jedes gesagte Schimpfwort sowie Figuren-Tode unter Angabe der exakten Zeit quantifizierte. Seine Ergebnisse stellte er auf Github zur Verfügung, die als Grundlage dieser Arbeit dienen (https://github.com/fivethirtyeight/data/blob/master/tarantino/tarantino.csv). Zusätzlich wurde ein zweiter Datensatz angelegt, der Metadaten der genannten Filme enthält, sowie das Erscheinungsjahr und die Filmlängen. Außerdem wurden die Film-Bewertungen der IMDb (Internet Movie Database) und „Rotten Tomatoes“ aufgenommen. Die beiden genannten Datensätze liegen als CSV-Datei vor und können somit mittels der Methodik von D3 aufbereitet werden.

Die Idee dieses Projekts ist es, eine ansprechende und unterhaltsame Exploration der Daten bereitzustellen, die auch im wissenschaftlichen Kontext der Medienwissenschaft (z.B. der Filmanalyse) verwendet werden kann. Dabei wurden folgende Ziele verfolgt:

1. Darstellung der Anteile einer Wortkategorie an den gesamten Schimpfwörtern eines Films (Bubble Chart)
2. Vergleichender Überblick, welche Wörter in welchem Verhältnis in den verschiedenen Filmen vorkommen bzw. welche Wörter ausgeschlossen sind (Sankey Flow)
3. Vergleich der Gesamtanzahl an Schimpfwörtern zwischen Filmen (Bubble Chart: Curses by Movie, Sankey Flow, Bar Chart)
4. Vergleich der Anzahl an Todesfällen zwischen Filmen, Vergleich der Quote der Todesfälle im Verhältnis zu Schimpfwörtern (Bar Chart)
5. Vergleich der Film Ratings über die Filme hinweg, Vergleich des Ratings eines Films auf unterschiedlichen Plattformen (Movie Ratings)

# Umsetzung

## Bubble Chart

Die Schimpfwörter wurden bei der Aufbereitung der Daten nach den verschiedenen Kategorien „ass“, „shit“, „fuck“, „racial“, „genital“, „blasphemy“ und „other“ geordnet, in denen bedeutungsähnliche Worte oder solche desselben Wortstammes zusammengefasst wurden.

Der Bubble Chart erlaubt in erster Linie eine quantitative Visualisierung der Schimpfwörter in Summe, aber auch interaktive Möglichkeiten für den User. Über drei Buttons kann zwischen einer Ansicht der Gesamtheit der Wörter („All Curses“), einer Sortierung je Film („Curses by Movie“) und einer weiteren Aufschlüsselung nach Schimpfwortkategorien innerhalb eines Films („Curses by Type“) gewechselt werden. So kann sich der User ein Bild darüber machen, in welchem Film welches Wort genau wie oft gesagt wird, wieviel insgesamt im Filmvergleich geflucht wird und welche Typen von Schimpfwörtern welchen Anteil innerhalb eines Films ausmachen. Durch Hovern über den einzelnen Bubbles erhält der Nutzer in Form eines Tooltips Details zu Anzahl der Vorkommnisse im jeweiligen Film mit dessen Erscheinungsjahr zum gewählten Wort.

Für die Darstellung von Häufigkeiten und Gruppierungen erschien der Bubble Chart mit force layout am besten geeignet. Es ist naheliegend, dass sich die Proportionalität von Häufigkeiten gut als Größenverhältnisse von geometrischen Formen visualisierend lässt und intuitiv interpretiert wird. Da die Motivation für die Visualisierung einem linguistischen Interesse entspringt, erschien es sinnvoll, die oben genannte Kategorisierung der Wörter vorzunehmen. Hiervon lassen sich mögliche Metadaten ableiten, beispielsweise welcher Typ von Kraftausdrücken in Relation zu den anderen Typen überwiegt. Somit wird ein breites Potential an weiterführenden Fragestellungen erschlossen.

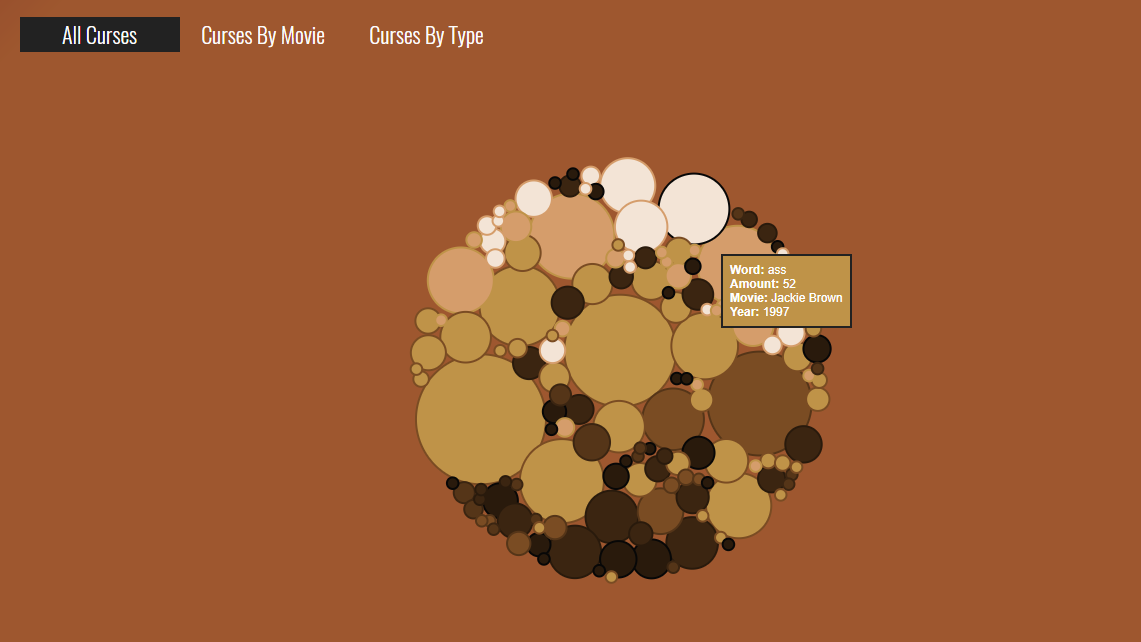


Abb. 1: Bubble Chart – All Curses



Abb. 2: Bubble Chart – Curses by Movie

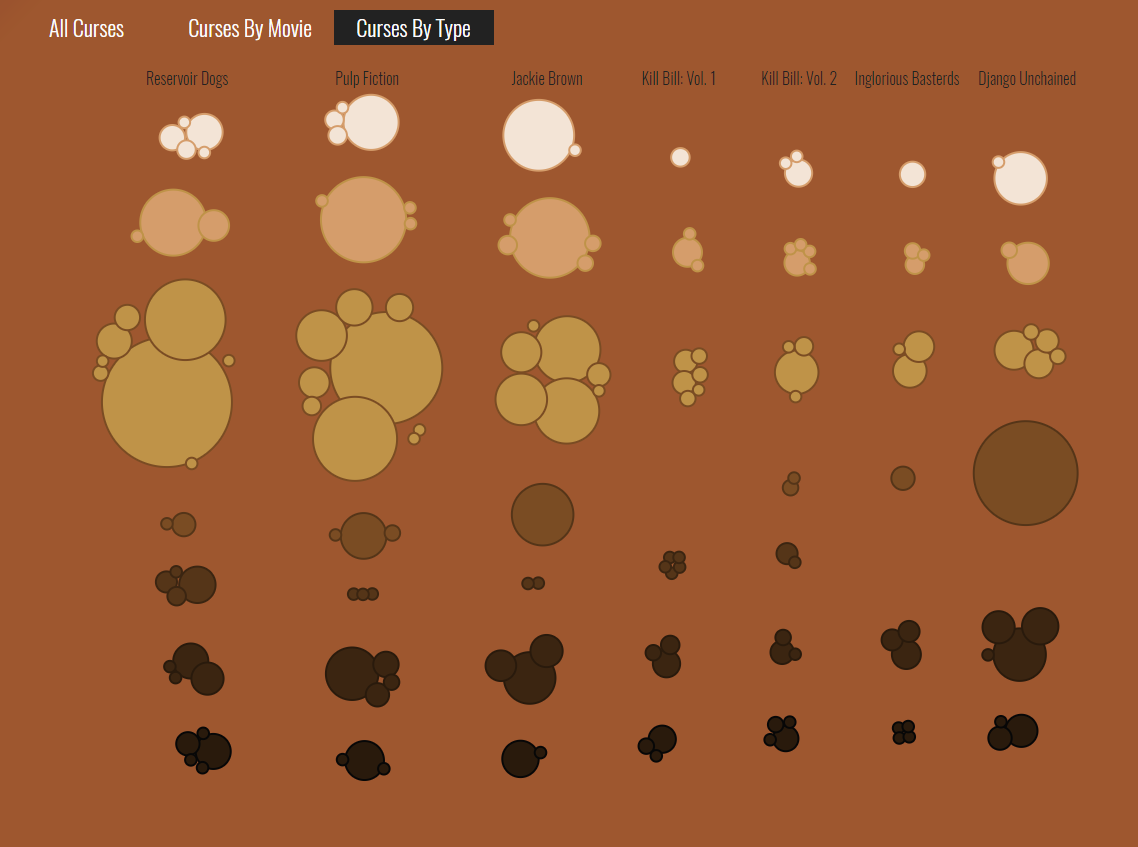


Abb. 3: Bubble Chart – Curses by Type

## Line Graph

Interessant sind vor dem Hintergrund der Filmanalyse auch die Auftretenshäufigkeiten von Schimpfwörtern innerhalb der dramaturgischen Erzählstruktur, sprich der zeitlichen Einbettung. Diesem Aspekt trägt der Line Graph Rechnung, indem er dem Nutzer gestattet, die Menge der Schimpfwörter in verschieden feinkörnig wählbaren Zeitabschnitten zu untersuchen. Unterhalb der Zeitachse wird der ausgewählte Abschnitt vergrößert und die Wörter sind individuell durch hovern auf den Strichen zu ihrem jeweiligen Zeitpunkt zu sehen. Zugleich können mehrere bis alle Filme auf einmal im Graphen dargestellt und simultan verglichen werden. Hierfür wurden alle Laufzeiten normalisiert.

Diese Visualisierung wurde so gewählt, als dass sie sich der Auflösung bis ins kleinste Detail widmet und so eine Minute-zu-Minute-Analyse ermöglicht.

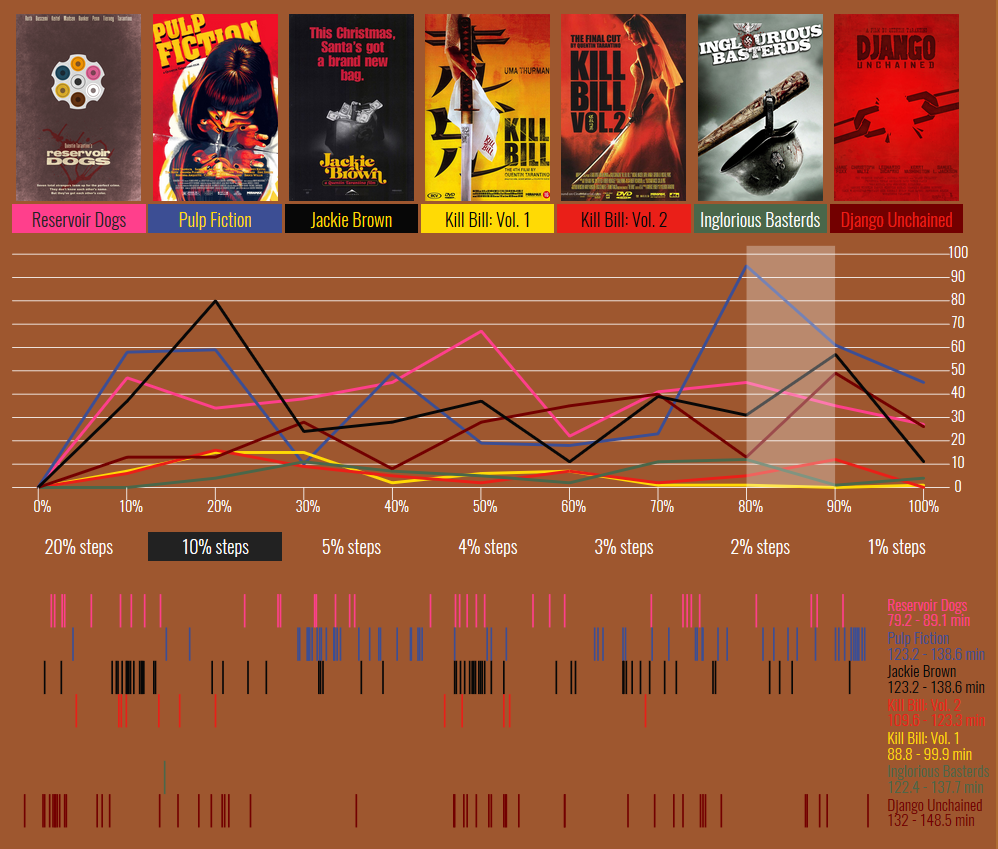


Abb. 4: Line Graph

## Sankey Flow

Der Mehrwert des Sankey Flow besteht darin, dass die Visualisierung es dem Nutzererlaubt, zu erforschen, welche Wörter ganz konkret in welchen Filmen vorkommen und welche davon ausgenommen sind, während die übrigen Darstellungen lediglich einen Überblick über die inklusiven Worte geben.

Es ist möglich, jeweils vom Film ausgehend die enthaltenen und ausgeschlossenen Wörter zu entdecken oder umgekehrt zu sehen, in welchen Filmen ein bestimmtes Wort wie oft gesagt wird. Dadurch sind Rückschlüsse über den Kontext möglich, in dem sich ein Film bewegt oder wie stark hier Vulgärsprache Einzug findet.

Die Breite der Stränge orientiert sich an der Menge der auftretenden Wörter. Zur besseren Unterscheidung und konsistenten Farbcodierung der Filme sind die Balken der Filme sowie deren zugehörige Stränge in den jeweiligen Filmfarben gefärbt.

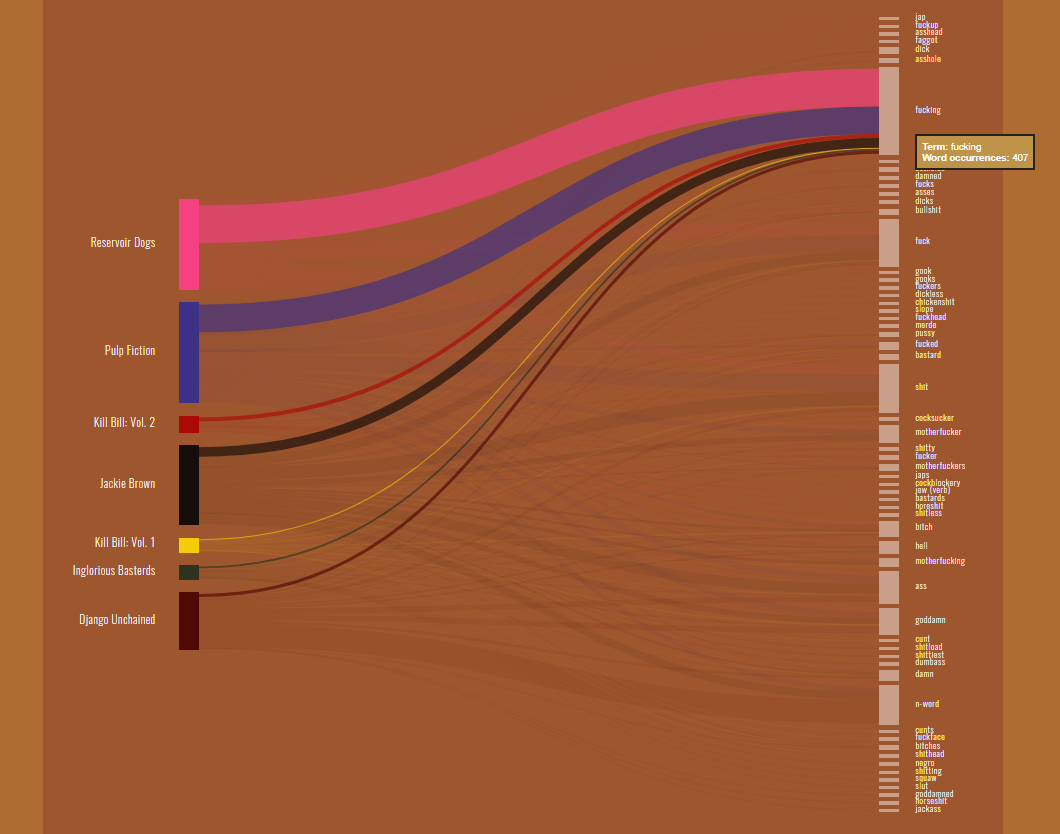


Abb. 5: Sankey Flow

## BarChart

Die Bar-Chart-Visualisierung bietet eine Veranschaulichung des Curse-Death-Verhältnisses in den ausgewählten Filmen. Neben der Darstellung eines Grouped Bar Charts (Gruppiertes Säulendiagramm) steht zusätzlich ein Stacked Bar Chart (gestapeltes Säulendiagramm) zur Verfügung. Die höhenproportionale Darstellung der Balken entlang der vertikalen Größenachse erlaubt dem Nutzer eine ungefähre Vorstellung der Häufigkeitswerte (Bsp. Siehe Abb. : in Reservoir Dogs kommen ~425 cursewords vor).

Abb. 6: Bar Chart

(1) Grouped Bar Chart: Für jede Datenkategorie (bzw. jeden Film) existieren entlang der horizontalen Rubrikenachse zwei gruppierte, unterschiedlich gefärbte Balken, welche für die beiden Datenreihen „curses“ und „deaths“ stehen. Innerhalb jeder Gruppierung wird die Balken-Reihenfolge beibehalten. Diese Darstellung eignet sich hier für den Vergleich von einzelnen Werten untereinander, sowohl bezogen auf nur einen Film, als auch Kategorie-übergreifend. (2) Stacked Bar Chart: Die einzelnen Teilbalken einer Kategorie werden zu einem Säulendiagramm übereinandergestapelt, wodurch sich die Höhe des neu entstandenen Balkens vergrößert.

Folgende Explorative Möglichkeiten stehen dem Nutzer hier zur Verfügung: Mithilfe von zwei Buttons kann der User zwischen den beiden genannten Diagramm-Varianten wählen. Die Balken positionieren sich dann je nach Einstellung innerhalb ihrer Kategorie nebeneinander als Balken-Gruppe oder jeweils übereinander als ein einziger gestapelter Balken, der im Vergleich mit den anderen Kategorien gesetzt wird. Außerdem erscheint beim Hovern der Balken ein Tooltip, welcher genaue Werte bezüglich der Auftretenshäufigkeiten der beiden Typ-Kategorien „curses“ und „deaths“ angibt. Das erlaubt es dem Nutzer, sich einen Überblick über die Häufigkeitsverhältnisse von Toden zu Schimpfwörtern innerhalb eines Films und zwischen Filmen im Vergleich zu schaffen.

## Movie Ratings

Zur Darstellung der Ratings aus der IMDb und von Rotten Tomatoes wurde ein “liquid fill gauge“ gewählt. Dieses hat die Optik eines runden, sich füllenden Tanks, wobei der finale Füllstand den Wert des jeweiligen Ratings repräsentiert und dieser auch als numerischer Zahlwert in der Mitte angezeigt wird und beim Laden des Features mit hochzählt. Die Füllanzeige ist entsprechend der Rating-Skala in der IMDb und Rotten Tomatoes von 0-10 skaliert. Dem Nutzer wird ein Überblick über die Bewertungen auf den beiden Plattformen gegeben. Das erlaubt den Vergleich der Filmbewertungen gegeneinander („Welcher ist der bestbewertete Film?“), sowie der Bewertungen eines Films auf den verschiedenen Plattformen.

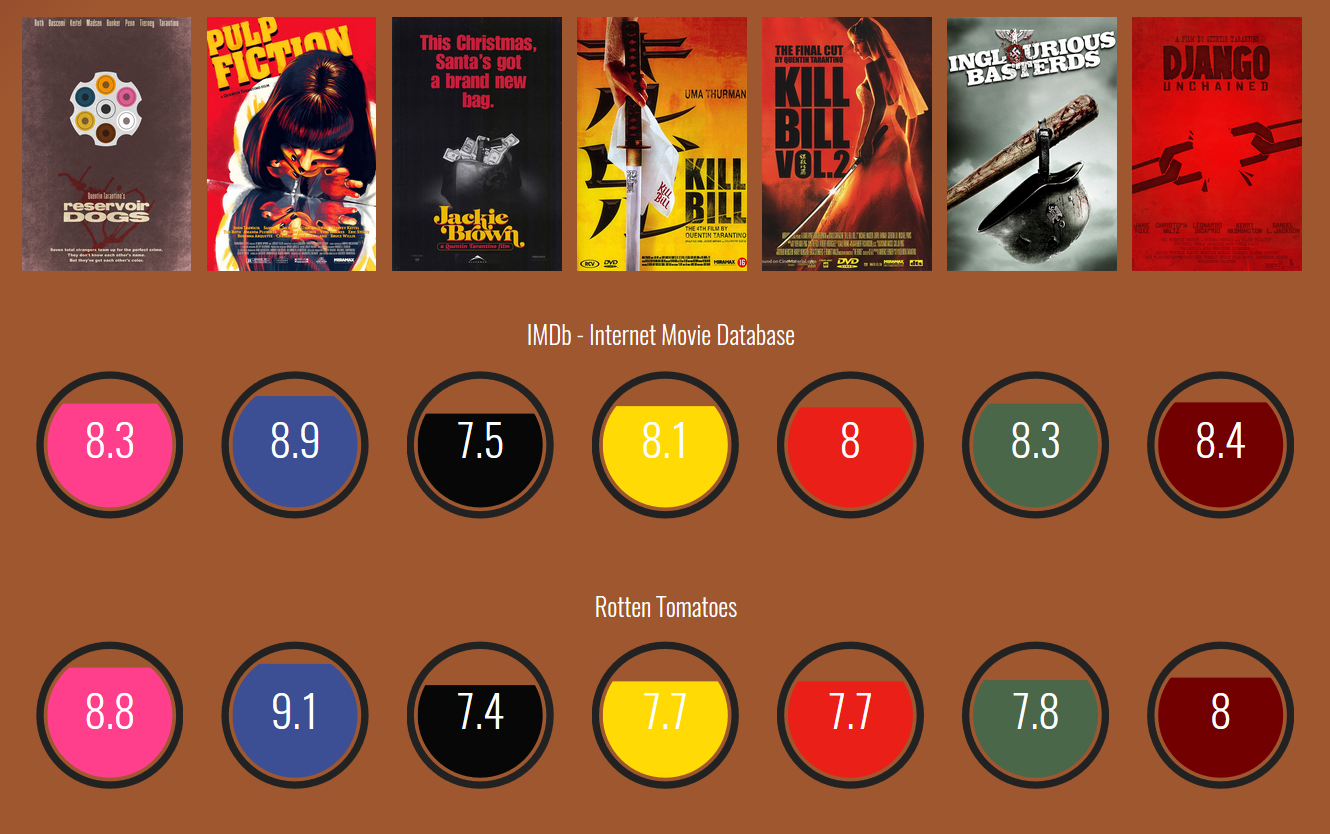
Gestaltungstechnisch greift die runde Kreisform die Optik des Bubblecharts wieder auf und wird mit der ansprechenden Animation in Form des Hochzählens des Ratings und Füllung des Kreises ergänzt. Damit entsteht ein konsistenter Eindruck, der bspw. durch eine Veranschaulichung mit x von 10 Sternen gestört würde. Die Farbwahl der Füllung entspricht den „Farbcodes“ der einzelnen Filme, welche im gesamten Projekt verwendet werden.

Abb. 7: Movie Ratings

# Gestaltungskonzept

Ein wichtiger Punkt stellt die Interaktivität der Visualisierungen dar, mit der die Nutzer den Datensatz erkunden können. Dies wurde durch eine Reihe von gestalterischen Elementen unterstützt, die im Folgenden näher vorgestellt werden.

Um das Konzept schnell verständlich zu machen, wird auf ein einheitliches Layout geachtet. Die Farbgestaltung der gesamten Oberfläche deckt sich mit den Kolorierungen des Logos – Jules Winnfields Kopf aus „Pulp Fiction“ – und ist in Braun-, Beige-, Weinrot-, und Schwärztönen gehalten. Für die Farbgebung der spezifischen filmbezogenen Elemente hat man sich am jeweils zugehörigen Filmplakat orientiert, damit der Nutzer diese beim Verwenden von entsprechenden Tools leicht mit dem ausgewählten Film assoziieren kann.

Ein Tab-Layout sorgt für einen übersichtlich gegliederten Aufbau. Mit Hilfe von Buttons kann zwischen den 5 verschiedenen Ansichten der einzelnen Implementierungen gewechselt werden. Innerhalb der Tabcontents findet der Nutzer kurze Texte, welche die jeweilige Visualisierung und die explorativen Möglichkeiten beschreiben.

Fährt der Nutzer mit der Maus über Elemente, wird dies für ihn erkenntlich gemacht, indem das gerade fixierte Element durch andersfarbliche Markierungen, eine farbliche Umrandung oder Vergrößerung der Elemente hervorgehoben wird. Des Weiteren werden beim Hovern von entsprechenden Elemente (Bubbles, Lines, Balken, …) Tooltips erzeugt, welche detaillierte Informationen, wie die exakte Anzahl der Schimpfwörter und Tode, oder die genaue Filmzeit, zu dem Element anzeigen. Auch bei den Tooltips wurde auf ein einheitliches Layout geachtet (siehe dazu vorangehende Abbildungen).

# Projektmanagement

Grundlegende Struktur und Konzept wurden gemeinsam erarbeitet, im Anschluss wurde die Hauptverantwortung für einzelne Features auf die Gruppenmitglieder aufgeteilt.

Um gemeinsam an diesem Projekt zu arbeiten nutzten wir die den Online-Dienst GitHub und arbeiteten in mehreren Branches, sodass parallel in verschiedenen Visualisierungen gearbeitet werden konnte. In den wöchentlichen Sitzungen tauschten wir uns über den Fortschritt aus, unterstützten uns gegenseitig bei Problemen und setzten die nächsten Ziele fest. Bei Schwierigkeiten oder kleineren Aufgaben wurden diese mit Hilfe der anderen Gruppenmitglieder erarbeitet. Hervorzuheben ist hierbei die große Hilfsbereitschaft von Louis Ritzkowski, der in allen Bereichen bei eben diesen geholfen hat.

**Hauptverantwortung**

|  |  |
| --- | --- |
| Bubble Chart | Louis Ritzkowski |
| Line Graph | Louis Ritzkowski |
| Sankey Flow | Teresa Then, Alicia Fritsche |
| Bar Chart | Teresa Then, Alicia Fritsche |
| Movie Ratings | Amanda Breuer |
| Grafiken, Farbgestaltung, Design,  Layout, Konzept | alle |