



Hochschule für Angewandte Wissenschaften Hamburg  
*Hamburg University of Applied Sciences*

# Ausarbeitung Projekt 1

**Bastian Probst**

**Durchführung eines Votings zur Erstellung von  
Benutzerprofilen**

*Fakultät Technik und Informatik  
Studiendepartment Informatik*

*Faculty of Engineering and Computer Science  
Department of Computer Science*

Bastian Probst

**Durchführung eines Votings zur Erstellung von  
Benutzerprofilen**

Ausarbeitung Projekt 1

im Studiengang Master Informatik  
am Department Informatik  
der Fakultät Technik und Informatik  
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuer: Prof. Dr. Kai von Luck

Eingereicht am: 27. Mai 2015

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Architektur</b>	<b>1</b>
2.1	Registrierung . . . . .	2
2.2	Login . . . . .	3
2.3	Web Frontend . . . . .	3
2.4	Persistenz . . . . .	4
2.5	Votingsystem . . . . .	5
2.6	Datenverarbeitung . . . . .	5
2.6.1	Datenzusammenfassung . . . . .	5
2.6.2	Benutzermodelle . . . . .	5
2.7	Vorhersagesystem . . . . .	6
2.8	Auswertung . . . . .	6
<b>3</b>	<b>Projekt</b>	<b>6</b>
3.1	Aktueller Stand . . . . .	6
3.1.1	Webseite . . . . .	7
3.1.2	Akquirierung . . . . .	7
3.2	Umsetzung . . . . .	7
3.2.1	Wordpress . . . . .	7
3.2.2	Registrierung . . . . .	8
3.2.3	Videodarstellung . . . . .	8
3.2.4	Yop Poll . . . . .	8
3.3	Datenanalyse . . . . .	9
<b>4</b>	<b>Fazit</b>	<b>9</b>
4.1	Bisheriger Verlauf . . . . .	10
4.2	Kommende Tätigkeiten . . . . .	10
4.2.1	Zum zweiten Voting . . . . .	10
4.2.2	Programmierung und Personalisierung . . . . .	10
4.2.3	Auswertung . . . . .	10
<b>5</b>	<b>Anhang</b>	<b>10</b>
5.1	Sequenzdiagramm für einen Vote mit Registrierung . . . . .	11
5.2	Bilder der Webseite . . . . .	12

# 1 Einleitung

Immer mehr digitale Medieninhalte werden verfügbar. Um die Benutzer nicht durch dieses beständige Wachstum zu überfordern, ist eine Möglichkeit zum Filtern der Inhalte die Personalisierung [Fan u. a. \(2012\)](#) [Bellekens u. a. \(2009\)](#) [Weiß u. a. \(2008\)](#). In diesem Paper wird der erste Teil eines Experiments beschrieben, welches im Rahmen meiner Masterarbeit durchgeführt wird. Die Teilnehmer bewerten Filme und einige Fragen zur Person. Parallel wird ein System entwickelt, welches kommende Bewertungen, für noch nicht bewertete Filme vorhersagt. Diese werden mit den tatsächlich Bewertungen des zweiten Teils des Experiments verglichen.

Um Benutzermodelle zu erstellen, existieren zwei unterschiedliche Methoden: die Selbstauskunft und die Analyse des Verhaltens der Benutzer [Weiß u. a. \(2008\)](#). Im Kontext dieser Arbeit wird die Methode der Selbstauskunft verwendet.

Es gibt viele unterschiedliche Methoden, um Vorschläge zu generieren. Einige erfordern eine große Menge an Benutzern. Diese sind im Kontext dieser Arbeit nicht weiter berücksichtigt worden. Die Anzahl der akquirierten Teilnehmer ist hierfür nicht ausreichend. Das Experiment wird in Kapitel 3 genauer beschrieben.

Die unterschiedlichen Arten von Benutzermodellen und deren Verwendung für Personalisierung wurde im Rahmen des ersten Vortrags des Masterseminars herausgearbeitet [Probst \(2014b\)](#). Im zweiten Vortrag wurde dargestellt, wie solche Benutzermodelle erstellt und für Vorschläge verwendet werden können [Probst \(2014a\)](#). Die theoretischen Grundlagen des Experiments wurden im Hauptseminar vorgestellt [Probst \(2015\)](#).

In dieser Arbeit geht es um die Frage:

„Ist es möglich, durch Selbstauskunft Benutzermodelle zu erstellen, welche zur Vorhersage für Filmbewertung verwendet werden können?“

und um eine praktische Überprüfung dieser Frage. Hierfür ist ein Experiment konzipiert worden, welches diese Frage beantworten soll. In dieser Arbeit wird der erste Teil des Experiments beschrieben. Die folgenden Aufgaben werden aufgeführt. Diese werden im Projekt 2 und im Rahmen der Masterarbeit durchgeführt.

Das 1. Kapitel dieser Ausarbeitung gibt einen Überblick über das Thema und den Inhalt. Im 2. Kapitel wird die Architektur beschrieben, welche im Experiment verwendet wird. Im 3. Kapitel wird das Experiment erklärt. Im 4. Kapitel werden der bisherige Verlauf und die kommenden Tätigkeiten herausgestellt.

## 2 Architektur

In diesem Kapitel wird die Architektur beschrieben, welche für die Umsetzung des Projekts verwendet wird. Es werden die einzelnen Komponenten beschrieben. Dabei handelt es sich um folgende Komponenten:

- eine **Registrierung**
- ein **Login**
- ein **Web Frontend**
- eine **Persistenz**
- ein **Votingsystem**
- eine **Datenverarbeitung**
- eine **Benutzermodellerstellung**
- ein **Vorhersagesystem**
- und eine **Auswertung**.

### 2.1 Registrierung

Damit sich die Benutzer einloggen können, ist eine Registrierung erforderlich. Dies ermöglicht dem System, einen Login zur Verfügung zu stellen, um die Benutzer eindeutig zu identifizieren. Die Registrierung muss folgendes leisten:

- eine **Verifizierung** der Benutzerdaten. Die Benutzer registrieren sich mit einer E-Mailadresse und einem Passwort. Zunächst wird über die Persistenz-Komponente [2.4](#) überprüft, ob ein Benutzer mit dieser Mailadresse vorhanden ist. Ebenfalls muss diese verifiziert werden. Um die Adresse zu Verifizieren, wird eine E-Mail mit einem Verifizierungslink gesendet. Eine valide E-Mailadresse ist unter anderem auch für die kommende Kommunikation mit den Benutzern erforderlich.

- Das **Anlegen und Speichern der Benutzer** im System. Die Benutzerdaten müssen mit Hilfe der Persistierungskomponente in der Datenbank gespeichert werden, um für weitere Logins zur Verfügung zu stehen. Die Loginkomponente benötigt diese Informationen, um den Login zu verifizieren. Sie wird im nächsten Abschnitt beschrieben.

## 2.2 Login

Der Benutzer muss bei der Interaktion mit dem System eindeutig identifiziert werden können. Hierfür ist ein Login erforderlich. Für den Login ist die im vorherigen Kapitel beschriebene Registrierung erforderlich. Die Loginkomponente muss Folgendes leisten:

- ein **Login** mit Abfrage der E-Mail und des Passwortes. Hierbei wird
- eine **Session** erstellt. Diese ermöglicht es, die Aktionen eindeutig dem jeweiligen Benutzer zuzuordnen, ohne dass dieser sich jedes Mal neu identifizieren muss. Zu Sicherheitszwecken wird
- die **Anzahl der Loginversuche von einer IP-Adresse begrenzt**. Dies bietet für den Rahmen dieser Arbeit einen ausreichenden Schutz. Für den Fall, dass ein Benutzer sein Passwort vergessen hat, wird ihm
- eine **Passwortwiederherstellung** ermöglicht. Da der Benutzer bei der Registrierung seine E-Mailadresse angibt, kann er sich an diese Adresse ein neues Passwort zuschicken lassen.

## 2.3 Web Frontend

Um mit dem System interagieren zu können, wird ein Webfrontend benötigt. Dies ist die Schnittstelle zwischen dem System und dem Benutzer. Folgende Eigenschaften sind erforderlich:

- eine **Registrierungsmaske** zum Eingeben der E-Mailadresse und des gewünschten Passwortes. Auf diese Weise kann der Benutzer mit der Registrierungskomponente kommunizieren und seine Logindaten hinterlegen. Diese Informationen sind für sowohl für die Registrierung als auch die Login Komponente erforderlich.
- ein **Anmeldebildschirm** zur Kommunikation mit der Loginkomponente für die Identifizierung am System. Der Benutzer identifiziert sich mit seiner E-Mailadresse und seinem Passwort.

- eine **Anzeige von Trailern** aus YouTube, damit der Benutzer die Möglichkeit hat, die Trailer von ihm unbekanntem Filmen zu schauen.
- eine **Votingoberfläche** für die Kommunikation mit der in Kapitel 2.5 beschriebenen Votingkomponente. Hier kann der Benutzer eine Bewertung der Filme und der Kategorien von -3 bis +3 inklusive der Null abgeben. Somit sind sieben unterschiedliche Bewertungen möglich. Eine höhere Anzahl würde nur einen minimalen Mehrwert liefern. Ebenfalls sind die Fragen zur Person wie Alter, Geschlecht, usw. auf diesem Wege realisiert.
- eine **Anzeige der Informationen** über das Experiment. Hierbei werden folgende Fragen beantwortet:
  - Wer führt dieses Experiment durch?
  - Zu welchem Zweck wird das Experiment durchgeführt?
  - Wie ist der zeitliche Ablauf?
  - Was wird von den Teilnehmern erwartet?
- eine **Feedback-Anzeige** nach dem Voting. Wenn eine Bewertung erfolgreich durchgeführt wurde, gibt die Votingkomponente ein Feedback. Dem Benutzer wird direkt nach dem Voten angezeigt, dass er geantwortet hat. Auch beim späteren Betrachten des Votes wird er darüber informiert, dass er schon geantwortet hat. Somit ist dem Benutzer ersichtlich, welche Filme schon bewertet wurden.
- eine **Navigation**, um durch das System navigieren zu können, wie zum Aufrufen der Informationsseite und der einzelnen Votes.

### 2.4 Persistenz

Für einen schnellen Zugriff und die Persistenz der Daten wird eine Persistenz-Komponente benötigt. Diese speichert

- die **Teilnehmer des Votings** mit ihrer E-Mailadresse und Passwort,
- die **Bewertungen der Filme** mit Zeitpunkt, Wert und dem Teilnehmer,
- die **Bewertungen der Kategorien** mit Zeitpunkt, Wert und dem Teilnehmer und
- die **Angaben zur Person** mit dem jeweiligen Benutzer.

Auf Datenbankebene werden diese Tabellen und Spalten benötigt:

- **Benutzer:** Benutzer ID, E-Mailadresse, Passwort.
- **Fragen** (Filme/Kategorien): Frage ID, Fragetext.
- **Antworten:** Antwort ID, Antworttexte.
- **Verbindungstabelle von Fragen zu Antworten:** Frage ID, Antwort ID
- **Tabelle der Votings:** Antwort ID, Benutzer ID, Zeitpunkt)

## 2.5 Votingsystem

Für die Verarbeitung der Bewertungen wird ein Votingsystem benötigt. Dieses bietet die Möglichkeit, Fragen zu erstellen. Für diese müssen vorgefertigte Antworten zu definieren sein oder eine freie Eingabe ermöglicht werden. Die Benutzer können die Fragen beantworten. Es kann die Anzahl der möglichen Antworten festgelegt und überprüft werden.

## 2.6 Datenverarbeitung

Bei der Datenverarbeitung müssen zum einen ähnliche Filme zusammengefasst werden, zum anderen müssen Benutzermodelle erstellt werden.

### 2.6.1 Datenzusammenfassung

Bei der Datenzusammenfassung müssen sinnvolle Kategorien gebildet werden, in denen die Filme zusammengefasst werden. Nach der Erstellung der Kategorien müssen die Filme den Kategorien zugeordnet werden. Dies geschieht anhand der Eigenschaften. Ein Film kann zu mehreren Kategorien gehören. Die Kategorien werden von dem Benutzer bewertet.

### 2.6.2 Benutzermodelle

Zum Erstellen der Benutzermodelle werden die Votings genauer analysiert. Es wird betrachtet, wie die jeweiligen Kategorien bewertet werden. Ebenfalls müssen die Bewertungen für die Filme im Benutzermodell gespeichert werden. Anhand der Bewertungen für Filme muss berechnet werden, wie der Benutzer tatsächlich Filme einer bestimmten Kategorie bewertet. Das Ergebnis dieser Berechnung wird im Benutzermodell gespeichert.

## 2.7 Vorhersagesystem

Das Vorhersagesystem verwendet die Daten aus den Benutzermodellen und der Datenzusammenfassung, um vorherzusagen, wie die Benutzer weitere Filme bewerten. Hierfür ist erforderlich, dass:

- **mehrere Algorithmen** verwendet werden können,
- die erstellten **Benutzermodelle** verwendet werden,
- das System automatisch **Bewertungen** berechnet und
- **Ausreißer** berücksichtigt werden.

## 2.8 Auswertung

Die unterschiedlichen Algorithmen der Vorhersagen werden miteinander verglichen. Dies wird im Rahmen der Auswertung im Anschluss an das Experiment erarbeitet. Von der Auswertung wird erwartet, dass

- **mehrere Metriken** verwendet werden.
- **unterschiedliche Bewertungen** der Algorithmen angewendet werden.
- eine **Analyse der Gesamtperformance** durchgeführt wird.
- die **Ausreißer** betrachtet werden.
- die **Auswertung reproduzierbar** ist.

# 3 Projekt

In diesem Kapitel werden der aktuelle Stand und besondere Aspekte bei der Umsetzung des Projekts beschrieben.

## 3.1 Aktueller Stand

Im Rahmen des Projektes wurde bisher das erste Voting abgeschlossen. Hierfür wurde eine Webseite erstellt, eine Akquirierung von Teilnehmern vorgenommen und eine erste Analyse der Daten durchgeführt.

### 3.1.1 Webseite

Diese Webseite enthält folgende Seiten:

- eine **Informationsseite** über den Versuchsaufbau.
- eine Seite für die **Angaben zur Person**.
- eine Seite für das **Bewerten der 23 Kategorien**.
- und jeweils eine Seite für die fünf Kategorien und deren zehn **zu bewertende Filme**.

### 3.1.2 Akquirierung

Um Teilnehmer für das Voting zu akquirieren, wurden in einer Erstsemester-Vorlesung, in einem Kurs des NextMedia Studiengangs und mit einem Aufruf über den Mailverteiler der HAW um Teilnahme gebeten. Einige der Teilnehmer sind aus dem privaten Umfeld. Über die Teilnehmer ist bekannt, dass

- 319 sich angemeldet haben.
- 233 mit dem Voting angefangen haben.
- 90 alles bewertet haben.
- 15 Teilnehmern noch 1-5 Votes fehlen.
- 135 männlich und 87 weiblich sind.
- 26 unter 21, 161 zwischen 21-30, 23 zwischen 31-40, 11 zwischen 41-50, einer zwischen 51-60, zwei zwischen 61-70 und keiner über 70 Jahre alt sind.
- 86 mehrere Filme pro Woche, 87 einen pro Woche und 50 seltener Filme schauen.

## 3.2 Umsetzung

Für die Umsetzung wurde eine Wordpress Installation mit diversen Plugins verwendet.

### 3.2.1 Wordpress

Die Webseite ist eine Wordpress Installation. Die Darstellung des Textes und die Menüführung sind Standardfunktionen. Somit konnte die Informationsseite problemlos erstellt werden. Wordpress sieht im Standard keine versteckten Seiten vor. Daher sind eine Privatisierung und eine Registrierung erforderlich.

### 3.2.2 Registrierung

Um die geforderte Privatisierung und Registrierung zu ermöglichen, wurden folgende zwei Plugins verwendet:

- Das Plugin **Register** ermöglicht es, dass die Teilnehmer sich mit ihrer E-Mailadresse registrieren und ihr Passwort wiederherstellen können. Der Login ist eine Standardfunktion.
- Mit dem Plugin **Privat only** wird verhindert, dass nicht eingeloggte Benutzer auf Seiten zugreifen können. Ebenfalls wird eine Seite zur Verfügung gestellt, welche einen neuen Benutzer darauf hinweist, dass dieser sich registrieren muss.

### 3.2.3 Videodarstellung

Die Standarddarstellung von Youtube-Videos hat Defizite in der Performance. Daher ist das Plugin „Youtube Lite“ verwendet worden. Dieses ermöglicht einen schnelleren Seitenaufbau, da die Videos erst bei der Betrachtung vollständig geladen werden.

### 3.2.4 Yop Poll

Zur Durchführung des Votings wurde das Plugin „Yop Poll“ verwendet. Dieses Plugin stellt ein Votingsystem bereit. Die Fragen und Antwortmöglichkeiten können dabei individuell definiert werden. Auf diese Weise war es möglich, die Bewertungen mit ihren sieben Ausprägungen und Freitextantworten zu definieren. Um doppeltes Abstimmen zu verhindern, bietet „Yop Poll“ die Möglichkeit, die Anzahl der Votes zu beschränken. Damit dem Benutzer angezeigt wird, welche Votings bereits durchgeführt wurden, musste eine Abfrage im Sourcecode ergänzt werden:

```
1 if( $this->hasPollsLeft()){
2 $template .= '<div id="yop-poll-container-error-' . $poll_id
3 . $unique_id . '" class="yop-poll-container-error"></div>';
4 }
5 else {
6 $template .= '<div id="yop-poll-container-error-' . $poll_id .
7 $unique_id . '" class="yop-poll-container-error">
8 Dies hast du schon gevotet!</div>';
9 }
```

Hierfür mussten folgende Funktionen zum Plugin hinzugefügt werden:

```
1 private function hasPollsLeft(){
2 global $current_user;
3 $voter = array();
4 $voter['poll_id'] = $this->poll['id'];
5 $voter['user_id'] = $current_user->ID;
6 $voter['user_type'] = 'wordpress';
7 return $this->user_have_votes_to_vote( $voter );
8 }
```

```
1 public function user_have_votes_to_vote( $voter ) {
2 $poll_options = $this->poll_options;
3 if ( $voter['user_id'] > 0 ){
4 if ( 'yes' == $poll_options['limit_number_of_votes_per_user'] ){
5 if ( $this->get_voter_number_of_votes( $voter ) >= $poll_options['number_of_votes_
6 return false;
7 }
8 }
9 }
10 return true;
11 }
```

### 3.3 Datenanalyse

Nicht alle Teilnehmer haben zuende ge votet. Einige haben sich lediglich angemeldet, andere haben angefangen und nicht alles bewertet. Zur Überprüfung wurde manuell für die Benutzer eine Abfrage durchgeführt, welche der Bewertungen sie noch nicht durchgeführt haben. Nach dem Ablauf der Frist für das Bewerten wurden alle Teilnehmer angeschrieben, die angefangen hatten und noch nicht fertig waren. Ihnen wurde mitgeteilt, welche Filme ihnen noch fehlten. Dies führte dazu, dass einige der Teilnehmer das Voting noch zuende durchgeführt haben. Der Stand der Votings befindet sich im Anhang.

## **4 Fazit**

### **4.1 Bisheriger Verlauf**

Das erste Voting ist erfolgreich durchlaufen. Eine erste Auswertung hat stattgefunden. Von den 300 anfänglichen Teilnehmern haben 90 alle Filme bewertet und Fragen beantwortet.

### **4.2 Kommende Tätigkeiten**

Im weiteren Verlauf muss ein zweites Voting, die Programmierung der Personalisierung und eine Auswertung der Ergebnisse durchgeführt werden.

#### **4.2.1 Zum zweiten Voting**

Beim zweiten Voting werden die Benutzer weitere zehn Filme bewerten. Diese Bewertungen werden mit denen vom System erstellten verglichen.

#### **4.2.2 Programmierung und Personalisierung**

Es muss ein System entwickelt werden, welches es ermöglicht, anhand der gewonnenen Daten Benutzermodelle zu erstellen. Ebenfalls muss die Verwendung unterschiedlicher Algorithmen ermöglicht werden. Mit diesen Benutzermodellen und Algorithmen werden die Bewertungen für die zehn Filme des zweiten Votings berechnet. Diese müssen mit den tatsächlichen Votings verglichen werden.

#### **4.2.3 Auswertung**

Nach der Durchführung des zweiten Votings und der Berechnung der Bewertungen, werden die Ergebnisse ausgewertet. Es werden unterschiedliche Metriken verwendet und die Aspekte berücksichtigt wie Ausreißer, untypisches Verhalten beim Abstimmen und das Votingverhalten allgemein.

## 5 Anhang

### 5.1 Sequenzdiagramm für einen Vote mit Registrierung

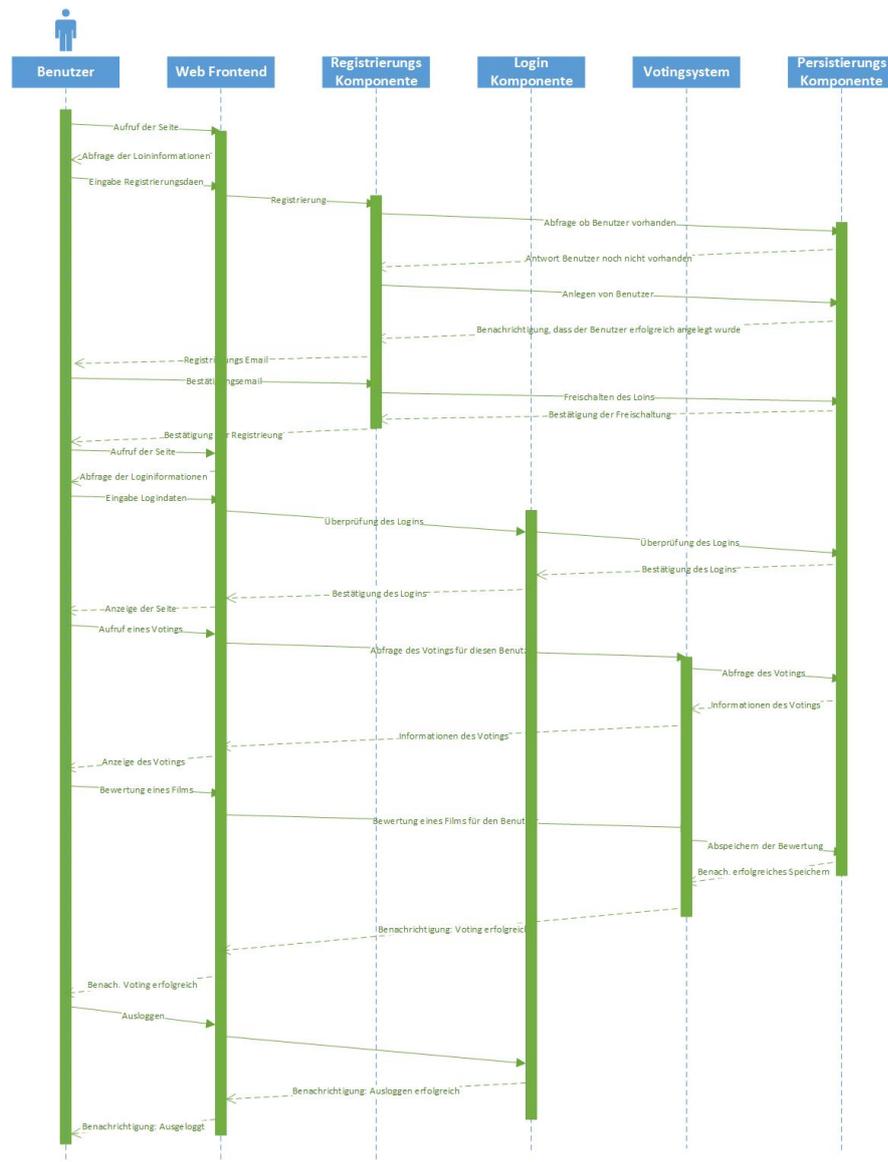


Abbildung 5.1: Sequenzdiagramm für einen Vote mit Registrierung

## 5.2 Bilder der Webseite

Bitte melde dich an, oder verwende den Link unten, um dich zu registrieren.

E-Mail

Passwort

Angemeldet bleiben

[Registrieren](#) | [Passwort vergessen?](#)

Abbildung 5.2: Login und Registrierung



## ALLGEMEINE FRAGEN

*dies hast du schon gevotet!*

Wie alt bist du?

- unter 21
- 21-30
- 31-40
- 41-50
- 51-60
- 61-70
- über 70

ABSENDEN

*dies hast du schon gevotet!*

Bist du männlich oder weiblich?

- männlich
- weiblich

ABSENDEN

Abbildung 5.3: Menü und allgemeine Fragen

- 1
- 2
- 3 (sehr gut)

VOTE

John Wick (2014)



Wie findest du den Film: John Wick?

- 3 (sehr schlecht)
- 2
- 1
- 0 (neutral)

Abbildung 5.4: Filmbewertung

# Literaturverzeichnis

- [Bellekens u. a. 2009] BELLEKENS, Pieter ; HOUBEN, Geert-Jan ; AROYO, Lora ; SCHAAP, Krijn ; KAPTEIN, Annelies: User Model Elicitation and Enrichment for Context-sensitive Personalization in a Multiplatform Tv Environment. In: *Proceedings of the Seventh European Conference on European Interactive Television Conference*. New York, NY, USA : ACM, 2009 (EuroITV '09), S. 119–128. – URL <http://doi.acm.org/10.1145/1542084.1542106>. – ISBN 978-1-60558-340-2
- [Fan u. a. 2012] FAN, Xiangyu ; MOSTAFA, Javed ; MANE, Ketan ; SUGIMOTO, Cassidy: Personalization is Not a Panacea: Balancing Serendipity and Personalization in Medical News Content Delivery. In: *Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium*. New York, NY, USA : ACM, 2012 (IHI '12), S. 709–714. – URL <http://doi.acm.org/10.1145/2110363.2110445>. – ISBN 978-1-4503-0781-9
- [Probst 2014a] PROBST, Bastian: Personalisierte Vorschläge für Medieninhalte. (2014)
- [Probst 2014b] PROBST, Bastian: Personalisierung von interaktiven Second Screen Anwendungen mithilfe von Benutzermodellen. (2014)
- [Probst 2015] PROBST, Bastian: Personalisierte Vorschläge für Medieninhalte. (2015)
- [Weiß u. a. 2008] WEISS, Diana ; SCHEUERER, Johannes ; WENLEDER, Michael ; ERK, Alexander ; GÜLBAHAR, Mark ; LINNHOF-POPIEN, Claudia: A User Profile-based Personalization System for Digital Multimedia Content. In: *Proceedings of the 3rd International Conference on Digital Interactive Media in Entertainment and Arts*. New York, NY, USA : ACM, 2008 (DIMEA '08), S. 281–288. – URL <http://doi.acm.org/10.1145/1413634.1413687>. – ISBN 978-1-60558-248-1