



WatchMe: Entwicklung eines Empfehlungssystems für den Kabelnetzbetreiber LIWEST

DIPLOMARBEIT

zur Erlangung des akademischen Grades

Mag.rer.soc.oec.

im Diplomstudium

WIRTSCHAFTSINFORMATIK

Eingereicht von:
Wolfgang Schneider

Angefertigt am:
Institut für Wirtschaftsinformatik – Data & Knowledge Engineering

Beurteiler:
o. Univ.-Prof. Dr. Michael Schrefl

Mitbetreuung:
Dr. Michael Karlinger

Linz, September 2015

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Die vorliegende Diplomarbeit ist mit dem elektronisch übermittelten Textdokument identisch.

Linz, September 2015

Wolfgang Schneider

Kurzfassung

Um seinen Kunden einen Service zu bieten, der ihn von den Mitbewerbern abhebt, plant der Kabelnetzbetreiber LIWEST, für seine Kunden ein Empfehlungssystem für TV-Sendungen einzuführen.

Im Rahmen dieser Diplomarbeit erfolgt eine Überprüfung verschiedener Filmempfehlungssysteme hinsichtlich ihrer Verwendbarkeit für das LIWEST Empfehlungssystem. Da alle überprüften Empfehlungssysteme entgegen den Anforderungen von LIWEST entweder kostenpflichtig sind, den Austausch von sensiblen Kundendaten mit externen Systemen erfordern oder nicht auf die bei LIWEST verfügbaren Sendungen abgestimmt werden können, kann keines dieser Empfehlungssysteme eingesetzt werden.

Daher wird ein eigenes Empfehlungssystem, welches den Namen WatchMe trägt, entwickelt. Die Entwicklung dieses Empfehlungssystems orientiert sich an bestehenden Ansätzen für Filmempfehlungssysteme und adaptiert Aspekte dieser Ansätze, um den Anforderungen von LIWEST zu entsprechen.

Der entwickelte Prototyp zeigt, dass der Ansatz funktioniert und WatchMe geeignet ist, Empfehlungen für LIWEST Kunden zu generieren. Dies wird abschließend auch anhand eines Anwendungsbeispiels, bei dem die Empfehlungsgenerierung aus Kundensicht dargestellt wird, veranschaulicht. Im September 2015 hat LIWEST die Absicht bekannt gegeben, das in dieser Diplomarbeit prototypisch implementierte Empfehlungssystem WatchMe zur Produktreife zu bringen.

Abstract

In order to offer its customers a service that distinguishes the cable television company LIWEST from competitors, LIWEST wants to introduce a movie recommendation system for its customers.

In this thesis different movie recommendation systems are reviewed with regards to their usefulness for the LIWEST recommendation system. Since all reviewed recommendation systems, either involve charges, require exchange of sensitive customer data with external systems or cannot be attuned to the broadcasts available at LIWEST which is in contrast to the requirements of LIWEST, none of the reviewed systems can be used.

Therefore a new recommendation system, called WatchMe, has been developed. The development of this recommendation system is orientated on existing approaches for movie recommendation systems and adapts aspects of these approaches to meet the requirements of LIWEST.

The developed prototype shows that the approach works and that WatchMe is appropriate for generating recommendations for LIWEST customers. This is also illustrated by an example where the generation of a recommendation is shown from the customer's perspective. In September 2015 LIWEST announced to bring the WatchMe prototype, developed in this diploma thesis, to maturity.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation.....	1
1.2	Ziele und Ergebnisse	1
1.3	Struktur der Diplomarbeit	2
2	Filmempfehlungssysteme	3
2.1	Grundlagen.....	3
2.1.1	Überblick Empfehlungsmethoden	3
2.1.2	Vergleich der Empfehlungsmethoden.....	6
2.2	Überblick Filmempfehlungssysteme	7
2.2.1	Ausgewählte Filmempfehlungssysteme	8
2.2.2	Vergleich der Filmempfehlungssysteme	11
2.3	Überblick Filmempfehlungsansätze	12
2.3.1	Untersuchte Ansätze	13
2.3.2	Vergleich der Ansätze.....	18
2.4	Fazit	23
3	Empfehlungssystem WatchMe	25
3.1	Verwendete Merkmale.....	25
3.2	Schnellempfehlung	28
3.3	Vollständige Empfehlung	29
4	Realisierung WatchMe	33
4.1	Prozesse	33
4.1.1	Authentifizierung durchführen	34
4.1.2	Sendungsbewertung durchführen	35
4.1.3	Benutzerverwaltung durchführen	37
4.1.4	Empfehlungsgenerierung durchführen.....	37
4.1.5	Sendungsdaten-Import	39
4.1.6	Benutzerdaten-Import	40
4.2	Datenstrukturen	42
4.3	Komponenten	46
4.4	Implementierung WatchMe	47
4.4.1	Implementierung der Authentifizierung.....	48
4.4.2	Implementierung der Sendungsbewertung.....	49
4.4.3	Implementierung der Benutzerverwaltung.....	49
4.4.4	Implementierung der Empfehlungsgenerierung	50
4.4.5	Implementierung des Sendungsdaten-Imports	52
4.4.6	Implementierung des Benutzerdaten-Imports.....	55
5	Anwendungsbeispiel	57
5.1	Authentifizierung.....	58
5.2	Empfehlungsgenerierung	58
6	Zusammenfassung.....	63
	Literaturverzeichnis	65
	Anhang.....	69

1 Einleitung

Dieses Kapitel erläutert die Problemstellung der vorliegenden Diplomarbeit. Anschließend werden die Ziele der Diplomarbeit angeführt und zum Abschluss wird ein Überblick über den Inhalt der folgenden Kapitel gegeben.

1.1 Motivation

Die Firma LIWEST ist mit ca. 120.000 Kunden der derzeit zweitgrößte Kabelnetzbetreiber in Österreich. In den letzten Jahren ist das Angebot an Fernsehsendern stark gestiegen. Die Vielfalt an Fernsehsendern erschwert es dem Benutzer, den Überblick darüber zu behalten, welche Sendungen für ihn interessant sein könnten und welche nicht.

Im Rahmen dieser Arbeit soll ein Filmempfehlungssystem entwickelt werden, um dem Kunden für ihn interessante Sendungen vorzuschlagen und ihm somit mühsames Durchforsten von Fernsehzeitschriften zu ersparen. Das Anbieten eines Filmempfehlungssystems ist ein Service, das LIWEST seinen Kunden bietet und das es LIWEST ermöglicht, sich von der steigenden Anzahl der Mitbewerber abzuheben.

1.2 Ziele und Ergebnisse

Das Ziel dieser Diplomarbeit ist die Konzeption und prototypische Implementierung eines Filmempfehlungssystems, das auf folgende Anforderungen von LIWEST abgestimmt ist:

System ist kostenlos: Da das System als kostenloses Zusatzservice angeboten werden soll, muss die Entwicklung ohne Kosten für ein Fremdsystem erfolgen.

System tauscht keine sensiblen Daten aus: Um den Datenschutz für LIWEST Kunden zu gewährleisten, darf das System keinerlei sensible Daten an externe Services übertragen. Als sensible Daten werden alle Daten, die Rückschlüsse auf die Identität der Kunden zulassen, angesehen.

System schlägt Live-TV-Sendungen vor: Da alle derzeit von LIWEST ausgestrahlten Sendungen Live-TV-Sendungen sind, muss das Empfehlungssystem diese auch vorschlagen können.

System ist auf eigene Sendungen anpassbar: LIWEST Kunden können unterschiedliche Fernhepakete buchen. Daher muss das Empfehlungssystem für jeden Kunden genau jene Sender berücksichtigen, die dieser empfangen kann.

System spricht junge Altersgruppe an: Da LIWEST in der Altersgruppe bis 29 Jahre seine Marktanteile erhöhen möchte, soll mit diesem Empfehlungssystem auch vermehrt diese Zielgruppe angesprochen werden.

Um das Ziel dieser Diplomarbeit zu erreichen wurde zuerst überprüft, ob bestehende Empfehlungssysteme den Anforderungen an das Empfehlungssystem für LIWEST entsprechen und somit für die Empfehlungsgenerierung herangezogen werden können. Da kein entsprechendes Empfehlungssystem gefunden werden konnte, wurde ein eigenes Empfehlungssystem mit dem Namen WatchMe entwickelt. Als Benutzeroberfläche wurde eine Android App entwickelt.

1.3 Struktur der Diplomarbeit

Im Kapitel *Filmempfehlungssysteme* wird auf die Grundlagen von Empfehlungssystemen im Allgemeinen und Filmempfehlungssystemen im Speziellen eingegangen, bevor bestehende Filmempfehlungssysteme hinsichtlich ihrer Verwendbarkeit für das LIWEST Empfehlungssystem überprüft werden.

In den nächsten Kapiteln wird auf das im Rahmen dieser Diplomarbeit entwickelte Empfehlungssystem WatchMe eingegangen. Im Kapitel *Empfehlungssystem WatchMe* wird der Ansatz zur Empfehlungsgenerierung konzeptuell beschrieben. Im Kapitel *Realisierung WatchMe* wird auf die Realisierung des Empfehlungssystems eingegangen. Dabei werden zuerst die abzubildenden Prozesse, die Datenstrukturen und die verwendeten Komponenten beschrieben, danach wird auf die Implementierung des Empfehlungssystems eingegangen. Anhand eines Anwendungsbeispiels wird anschließend die Anwendung von WatchMe aus Benutzersicht gezeigt.

Eine Zusammenfassung der Ergebnisse der Diplomarbeit und mögliche zukünftige Erweiterungen von WatchMe werden abschließend im Kapitel *Zusammenfassung* erläutert.

2 Filmempfehlungssysteme

In diesem Kapitel wird zuerst auf die Grundlagen von Empfehlungssystemen im Allgemeinen und von Filmempfehlungssystemen im Speziellen eingegangen. Danach werden bestehende Filmempfehlungssysteme hinsichtlich ihrer Verwendbarkeit für WatchMe untersucht.

2.1 Grundlagen

Empfehlungssysteme schlagen Benutzern mit Hilfe von speziellen Methoden Empfehlungselemente vor, die für den Benutzer vermeintlich interessant sind. Als Empfehlungselement bezeichnet man ein Element, das dem Benutzer von Systemen vorgeschlagen wird. Empfehlungselemente unterscheiden sich von System zu System. Bei Netflix sind Empfehlungselemente Filme, bei Amazon sind es die verkauften Produkte und bei dem Musikstreamingdienst Pandora sind es Musiktitel (Ricci, Rokach, Shapira, & Kantor, 2011).

Für die Generierung von Empfehlungen können unterschiedliche Methoden eingesetzt werden (vgl. Jannach, Zanker, Felfernig, & Friedrich, 2011):

- Collaborative-Recommendation
- Content-Based-Recommendation
- Knowledge-Based-Recommendation
- Hybrid-Recommendation

2.1.1 Überblick Empfehlungsmethoden

In diesem Abschnitt werden die unterschiedlichen Methoden der Empfehlungsgenerierung beschrieben. Anschließend werden im Abschnitt 2.1.2 diese Methoden miteinander verglichen.

2.1.1.1 Collaborative-Recommendation

Bei Collaborative-Recommendation werden einem Benutzer Empfehlungselemente vorgeschlagen, welche von ähnlichen Benutzern in der Vergangenheit gekauft oder positiv bewertet wurden. Die Ermittlung der Ähnlichkeit zwischen Benutzern erfolgt üblicherweise über die Bewertungen von Empfehlungselementen, die sowohl der Benutzer als auch der zu vergleichende Benutzer bewertet haben. Dadurch ist Collaborative-Recommendation von den Eigenschaften der betrachteten Empfehlungselemente weitgehend unabhängig (Klahold, 2009).

Bei Bewertungen wird zwischen impliziten und expliziten Bewertungen unterschieden. Bei expliziten Bewertungen werden die Empfehlungselemente von den Benutzern auf einer vorgegebenen Skala bewertet. Wie eng oder weit diese Skala sein soll kann nicht verallgemeinert werden, sondern hängt von den Elementen, die bewertet werden, ab. Der Nachteil von expliziten Bewertungen ist der Zusatzaufwand seitens der Benutzer. Kann man die Benutzer nicht dazu bewegen zu bewerten, wird die Qualität der Empfehlungen darunter leiden (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Implizite Bewertungen werden hingegen vom System automatisiert aufgezeichnet. So kann bei einem Onlinehändler das Kaufen eines Filmes automatisch als positive Bewertung betrachtet werden. Der Nachteil bei impliziten Bewertungen ist, dass nie mit Sicherheit gesagt werden kann, ob die Einschätzung korrekt ist. Beispielsweise bedeutet die Tatsache, dass ein Benutzer einen Film kauft noch nicht, dass er diesen auch positiv bewerten würde (Jannach, Zanker, Felfernig, & Friedrich, 2011).

2.1.1.2 Content-Based-Recommendation

Bei der Content-Based-Recommendation empfiehlt das System Empfehlungselemente, welche ähnlich zu denen sind, die der Benutzer in der Vergangenheit gekauft oder positiv bewertet hat. Die Empfehlungen hängen dabei sehr stark von den Eigenschaften der Empfehlungselemente ab (Klahold, 2009).

Um eine Ähnlichkeit zwischen Empfehlungselementen zu ermitteln müssen für jedes Empfehlungselement die Eigenschaften, die für den Vergleich mit anderen Empfehlungselementen relevant sind, bekannt sein. Welche Eigenschaften dies sind hängt davon ab, um welche Art von Empfehlungselement es sich in dem konkreten Fall handelt. Bei Büchern kann dies z. B. der Autor oder das Genre sein. Für TV-Sendungen können dies unter anderem Schauspieler, Regisseur, Inhaltsangaben und Genre sein (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Wie dann die Ähnlichkeit von zwei Empfehlungselementen ermittelt wird hängt stark von den zu vergleichenden Eigenschaften ab. So wird die Ähnlichkeit von Texten anders ermittelt als die Ähnlichkeiten von Jahreszahlen.

2.1.1.3 Knowledge-Based-Recommendation

Bei der Knowledge-Based-Recommendation muss der Benutzer seine Anforderungen an das Empfehlungselement bekannt geben und das Empfehlungssystem versucht, eine ent-

sprechende Empfehlung zu ermitteln. Kann keine Empfehlung ermittelt werden, muss der Benutzer seine Anforderungen anpassen (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Aufgrund der Art, wie die Empfehlung ermittelt wird, erfolgt nochmals eine Unterscheidung in Constraint-Based und Case-Based Recommendation.

Bei der Constraint-Based Recommendation werden die Empfehlungen aufgrund eines definierten Empfehlungsregelwerks ermittelt. Bei der Case-Based Recommendation dagegen wird für jede Eigenschaft eines Produktes die Ähnlichkeit zu den Benutzeranforderungen ermittelt und unter Berücksichtigung einer Gewichtung der Eigenschaften werden die Empfehlungen generiert (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Knowledge-Based-Recommendation wird häufig für die Empfehlung von Produkten, für die es keine Kaufhistorie gibt oder bei denen es keinen Sinn macht, die Historie zu verwenden, eingesetzt. Dies trifft auf Produkte zu, bei denen sich die Anforderungen zwischen zwei Käufen stark verändern oder die im Allgemeinen sehr selten gekauft werden. Beispiele dafür sind Computer, bei denen sich aufgrund von Weiterentwicklungen die Anforderungen ändern oder der Kauf von einem Eigenheim (Jannach, Zanker, Felfernig, & Friedrich, 2011).

2.1.1.4 Hybrid-Recommendation

Um die Empfehlungsergebnisse zu verbessern und die Schwächen einzelner Empfehlungsmethoden zu verringern, kann eine Kombination von Empfehlungsmethoden für die Generierung der Empfehlungen angewandt werden. Prinzipiell können dabei die Empfehlungsmethoden beliebig kombiniert und sogar unterschiedliche Ansätze der gleichen Methode, also zum Beispiel mehrere Collaborative-Recommendation Ansätze, zur Anwendung kommen. So eine Kombination mehrerer Empfehlungsmethoden nennt man Hybrid-Recommendation (Jannach, Zanker, Felfernig, & Friedrich, 2011).

In einigen Bereichen, wie zum Beispiel bei Filmempfehlungen, ist es eine gängige Methode, Empfehlungen aufgrund einer Hybrid-Recommendation aus Content-Based-Recommendation und Collaborative-Recommendation zu ermitteln.

Für die Realisierung einer Hybrid-Recommendation gibt es unterschiedliche Ansätze. Eine Variante ist es, zuerst unabhängige Ranglisten für die einzelnen Methoden zu erstellen und dann die Ergebnisse der einzelnen Methoden zu gewichten und zu kombinieren. Eine andere Variante besteht darin, dass das Empfehlungssystem unterschiedliche Situationen

unterscheiden kann und je nach Situation die passende Empfehlungsmethode wählt (Burke, 2002).

2.1.2 Vergleich der Empfehlungsmethoden

Die vier Methoden Collaborative-Recommendation, Content-Based-Recommendation, Knowledge-Based-Recommendation und Hybrid-Recommendation haben alle ihre Stärken und Schwächen und daher auch ihre unterschiedlichen Einsatzgebiete.

Da Collaborative-Recommendation kein spezifisches Wissen über die Eigenschaften der Empfehlungselemente benötigt, ist es sehr gut für Bereiche geeignet, in denen wenig über die Eigenschaften der Elemente bekannt ist. Dadurch, dass auf die Eigenschaften der Elemente verzichtet werden kann, erhöht sich auch die Performance des Empfehlungssystems und es lässt sich sehr gut skalieren. Außerdem eignet sich die Collaborative-Recommendation sehr gut dazu, überraschende und unerwartete Elemente vorzuschlagen (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Der Nachteil der Collaborative-Recommendation ist, dass Elemente nur empfohlen werden, wenn diese bereits einmal bewertet worden sind. Im Gegensatz zur Content-Based-Recommendation können neue Empfehlungselemente also nicht vorgeschlagen werden. Dies ist vor allem in Einsatzgebieten, bei denen Empfehlungselemente nur kurze Zeit verfügbar sind, wie zum Beispiel Live-TV, problematisch. Außerdem sind die Empfehlungen abhängig vom Vorhandensein ähnlicher Benutzer. Können keine Benutzer gefunden werden, die ähnlich jenem Benutzer sind, für den die Empfehlung ermittelt werden soll, können auch keine Empfehlungen generiert werden (Ricci, Rokach, Shapira, & Kantor, 2011).

Außerdem kann bei Collaborative-Recommendation Ansätzen der sogenannte Lemming-Effekt beobachtet werden. Dieser tritt vor allem bei impliziter Bewertung auf, kann jedoch auch bei expliziter Bewertung festgestellt werden. Elemente, welche einmal in den Empfehlungslisten auftauchen, werden oft aufgerufen, wodurch sie wieder anderen Benutzern empfohlen werden und noch öfter aufgerufen werden. Dadurch ist es für neue Elemente schwierig in die Liste der Empfehlungen zu rutschen (Klahold, 2009).

Durch die Unabhängigkeit der Content-Based-Recommendation von den Empfehlungen anderer Benutzer eignet sich der Einsatz sehr gut für Bereiche, in denen es oft neue, noch nicht bewertete Elemente gibt. Außerdem kann dem Benutzer sehr leicht dargestellt werden, weshalb Elemente empfohlen werden, was zur Verständlichkeit und somit zur Akzeptanz des

entsprechenden Empfehlungssystems beiträgt (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Der größte Nachteil der Content-Based-Recommendation ist die Spezialisierung. Da bei Content-Based-Recommendation nur auf die Bewertungen des Benutzers geachtet wird ist es unwahrscheinlich, etwas komplett Unerwartetes vorgeschlagen zu bekommen. Bei Musik- und Filmempfehlungen bewegt sich das System z. B. immer in den Genres, deren Alben der Benutzer in der Vergangenheit positiv bewertet hat (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Knowledge-Based-Recommendation ist nicht auf historische Bewertungen angewiesen, da Empfehlungen darauf beruhen, welche Anforderungen der Benutzer derzeit hat und wie die Empfehlungselemente diese Anforderungen erfüllen. Dadurch können auch Empfehlungen für Benutzer generiert werden, zu denen es keine Vergangenheitsdaten gibt (Ricci, Rokach, Shapira, & Kantor, 2011).

Das Hauptproblem bei der Knowledge-Based-Recommendation besteht jedoch in dem hohen Aufwand, welcher beim Aufsetzen und Warten des Empfehlungssystems und beim Hinzufügen von neuen Empfehlungselementen entsteht (Ricci, Rokach, Shapira, & Kantor, 2011).

Die Hybrid-Recommendation Methode wird eingesetzt, um die Nachteile der einzelnen Methoden zu beseitigen bzw. abzuschwächen. Sie kann prinzipiell in allen Bereichen eingesetzt werden, in denen die jeweiligen einzelnen Methoden eingesetzt werden können. Die Nachteile, die bei dem Einsatz von Hybrid-Recommendation entstehen, sind jedoch ein erhöhter Entwicklungsaufwand und negative Auswirkungen auf die Berechnungszeiten der Empfehlungen.

So wäre der Implementierungsaufwand für den Hybrid Algorithmus, der den 1 Mio. Dollar Preis für die Verbesserung des Netflix-Empfehlungsalgorithmus gewonnen hat, so hoch gewesen, dass dieser nie implementiert wurde (Netflix Techblog, 2015).

2.2 Überblick Filmempfehlungssysteme

Um die Frage zu klären, ob die Empfehlungsgenerierung für LIWEST neu konzipiert bzw. implementiert werden muss, wird in diesem Kapitel überprüft, wie gut bestehende Filmempfehlungsservices die Anforderungen und Rahmenbedingungen von LIWEST

erfüllen. Die Kriterien zur Überprüfung ergeben sich aus den im Unterkapitel 1.2 beschriebenen Anforderungen:

- System ist kostenlos
- System tauscht keine sensiblen Daten aus
- System schlägt Live-TV-Sendungen vor
- System ist auf eigene Sendungen anpassbar
- System spricht junge Altersgruppe an

Zusätzlich zu den oben angeführten Anforderungen wird in den Vergleich noch folgendes Kriterium aufgenommen:

- System verfügt über eine API (Application Programming Interface): Da es sich bei den Empfehlungssystemen um externe Systeme handelt wird überprüft, ob eine API vorhanden ist, um die erforderlichen Funktionalitäten integrieren zu können. Dieses Kriterium ist ein K.-o.-Kriterium. Bietet ein System keine API an, werden die weiteren Kriterien nicht beachtet.

Empfehlungssysteme, welche nicht kostenlos sind, werden trotzdem anhand der restlichen Kriterien verglichen, um einen Marktüberblick zu erhalten und um zu sehen, ob die restlichen Kriterien bezüglich der Empfehlungsgenerierung erfüllt werden. Daher wird dieses Kriterium nicht als K.-o.-Kriterium gesehen.

Das Kriterium *System spricht junge Altersgruppe an* ist nur für Empfehlungssysteme relevant, von denen auch die Benutzeroberfläche und nicht nur das Service zur Empfehlungsgenerierung verwendet wird.

2.2.1 Ausgewählte Filmempfehlungssysteme

Im Folgenden werden neun Filmempfehlungssysteme kurz beschrieben und anschließend im Abschnitt 2.2.2 anhand der zuvor beschriebenen Kriterien verglichen.

2.2.1.1 Tastekid

Die Empfehlungsgenerierung von Tastekid (Tastekid, 2015) basiert auf einer Liste mit Sendungen, welche dem Benutzer gefallen. Es können Empfehlungen für Filme und Serien ermittelt werden, jedoch keine Empfehlungen für Live-TV-Sendungen. Tastekid versucht weniger bekannte Titel zu empfehlen, um die Chancen zu erhöhen, dem Benutzer etwas Unbekanntes nahe zu bringen. Die Empfehlungsgenerierung von Tastekid kann nicht auf die

von LIWEST Kunden empfangbaren Sendungen angepasst werden. Tastekid bietet eine kostenlose REST-Schnittstelle, welche das Ergebnis im JSON Format zurückliefert. Da an das Empfehlungssystem nur eine Liste mit Sendungen übermittelt wird, werden keine sensiblen Daten ausgetauscht. Da von Tastekid keine Benutzeroberfläche zur Verfügung gestellt wird, wird das Kriterium *System spricht junge Altersgruppe an* nicht bewertet.

2.2.1.2 Jinni

Jinni (Jinni, 2015) bietet neben kostenlosen Empfehlungen für Benutzer direkt auf ihrer Webseite (ohne API) auch kostenpflichtige Lösungen für Unternehmen an. Dabei können die Sendungen, welche empfohlen werden sollen, komplett auf die für LIWEST Kunden empfangbaren Sendungen angepasst werden. Es können sowohl Video-on-Demand (VoD) Sendungen als auch Sendungen aus dem Live-TV berücksichtigt werden. Für die Empfehlungen werden Sendungen vom Jinni-Team auch noch mit Schlagworten versehen, um noch bessere Empfehlungsergebnisse zu erzielen. Zusätzlich können explizite und implizite Bewertungen sowie VoD-Käufe und das Zappingverhalten in die Empfehlungsgenerierung miteinbezogen werden, sofern diese Daten bekannt sind. Eine genauere Beschreibung der API ist nicht öffentlich zugänglich. Auch über die Preise und die Möglichkeiten der Eingliederung in die jeweilige IT-Architektur ist nichts bekannt. Somit kann auch nicht festgestellt werden, ob für die Empfehlung der Austausch sensibler Daten mit externen Servern benötigt wird. Da von Jinni keine Benutzeroberfläche zur Verfügung gestellt wird, wird das Kriterium *System spricht junge Altersgruppe an* nicht bewertet.

2.2.1.3 Rottentomatoes

Rottentomatoes (Rotten Tomatoes, 2015) ist mit ca. 25 Millionen Besuchern pro Monat eine der beliebtesten Filmbewertungsseiten im Internet. Über die kostenlose API können neben Film- und Schauspielerinformationen auch Kritiker- und Benutzerbewertungen zu Filmen abgerufen werden. Basierend auf einem übermittelten Filmtitel können Filmempfehlungen generiert werden. Rottentomatoes kann keine Filmempfehlungen für Live-TV-Sendungen generieren. Die Filmempfehlungen können auch nicht auf die von LIWEST Kunden empfangbaren Sendungen abgestimmt werden. Über die API werden nur Informationen über Sendungen und Schauspieler ausgetauscht, es werden also keine sensiblen Daten übertragen. Da von Rottentomatoes keine Benutzeroberfläche zur Verfügung gestellt wird, wird das Kriterium *System spricht junge Altersgruppe an* nicht bewertet.

2.2.1.4 *The Filter*

The Filter (The Filter, 2015) ist ein Empfehlungssystem, welches Empfehlungen für unterschiedlichste Empfehlungselemente generieren kann. Ursprünglich entstanden ist es aus einem Empfehlungssystem für Musiktitel, wird inzwischen aber auch für Filmempfehlungen, Produktempfehlungen, Eventempfehlungen und andere Bereiche verwendet. Der Filmempfehlungsalgorithmus verwendet sowohl historische Nutzungsdaten als auch globale Trends. Die Empfehlungen für Filme und Serien können auf die von LIWEST Kunden empfangbaren Sendungen angepasst werden. Es wird jedoch nicht angegeben, ob auch Live-TV-Sendungen vorgeschlagen werden können. Das Empfehlungssystem kann in die jeweils bestehende IT-Architektur integriert werden und kommt somit ohne Austausch sensibler Daten mit externen Servern aus. Das Empfehlungssystem ist kostenpflichtig, wobei auf der Webseite keinerlei Angaben über die Kosten gemacht werden. Da von „The Filter“ keine Benutzeroberfläche zur Verfügung gestellt wird, wird das Kriterium *System spricht junge Altersgruppe an* nicht bewertet.

2.2.1.5 *Reddiscover*

Red Bee Media (RedBee Media, 2015) ist ein zur Ericsson Gruppe gehörendes Unternehmen, welches weltweit als Mediendienstleister operiert. Red Bee Media bietet neben dem Ankauf von TV Metadaten auch das kostenpflichtige Empfehlungssystem Reddiscover an. Bei diesem Empfehlungssystem können Live-TV-Sendungen empfohlen werden und die Empfehlungen können auf die von LIWEST Kunden empfangbaren Sendungen abgestimmt werden.

Für die Sendungsempfehlungen werden in einem nicht näher spezifizierten Algorithmus Sendungsinformationen und Reviews von mehreren Filmseiten (z. B. IMDB) kombiniert. Über die API werden auch Benutzerinformationen und somit sensible Daten übermittelt. Da von Reddiscover keine Benutzeroberfläche zur Verfügung gestellt wird, wird das Kriterium *System spricht junge Altersgruppe an* nicht bewertet. Red Bee Media bietet jedoch auch Zusatzprodukte und Dienstleistungen für die Entwicklung von Benutzeroberflächen an.

2.2.1.6 *Filmaster*

Filmaster (Filmaster, 2015) ist ein Unternehmen, welches sich auf die Entwicklung von Empfehlungssystemen spezialisiert hat. Das kostenpflichtige Filmempfehlungssystem generiert neben Empfehlungen für Live-TV-Sendungen auch Empfehlungen für aktuelle Kinofilme. Die Empfehlungen können auf die von LIWEST Kunden empfangbaren Sendungen abgestimmt werden. Für die Empfehlungsgenerierung werden Daten über

Benutzerverhalten, Sendungsbewertungen des Benutzers und Daten des Facebook-Profiles des Benutzers an externe Server übertragen und auch dort gespeichert. Da von Filmaster keine Benutzeroberfläche zur Verfügung gestellt wird, wird das Kriterium *System spricht junge Altersgruppe an* nicht bewertet.

2.2.1.7 Contentwise

Contentwise (Contentwise, 2015) ist ein kostenpflichtiges Empfehlungssystem des italienischen Unternehmens Contentwise. Contentwise ermöglicht die Generierung von Empfehlungen für Live-TV-Sendungen, wobei die Empfehlungen auf die von LIWEST Kunden empfangbaren Sendern abgestimmt werden können. Über die Art und Weise der Einbindung von Contentwise in die jeweilige IT-Architektur ist ebenso wie über die Kosten des Empfehlungssystems nichts bekannt. Da es keine Informationen über die Einbindung in die IT-Architektur gibt, ist auch unbekannt, ob sensible Daten an externe Server übertragen werden. Contentwise bietet neben der Empfehlungsgenerierung auch Benutzeroberflächen für unterschiedliche Geräte, unter anderem iOS und Android Smartphones. Diese Benutzeroberflächen können auch junge Benutzergruppen ansprechen.

2.2.1.8 Whattorent

Whattorent (What to rent, 2015) ist ein kostenloses Filmempfehlungssystem, welches Filme für unterschiedliche Stimmungen empfiehlt. Whattorent bietet keine API und wird daher nicht näher betrachtet.

2.2.1.9 MovieLens

MovieLens (MovieLens, 2015) ist ein Empfehlungssystem für Filmempfehlungen und wurde 1997 von GroupLens Research der Universität von Minnesota entwickelt. MovieLens bietet keine API und wird daher nicht näher betrachtet.

2.2.2 Vergleich der Filmempfehlungssysteme

Tabelle 1 gibt einen Überblick über die untersuchten Filmempfehlungssysteme hinsichtlich der Erfüllung der Kriterien. Jene Kriterien, die ein System nicht erfüllt, sind rot hinterlegt.

Es ist ersichtlich, dass es einige Filmempfehlungssysteme (Jinni, Reddiscover, Filmaster und Contentwise) gibt, welche die inhaltlichen Anforderungen, die an das System gestellt werden, erfüllen. Diese Systeme können Live-TV-Sendungen empfehlen und die Empfehlungen auf die von LIWEST Kunden empfangbaren Sendungen abstimmen. Diese Empfehlungssysteme

sind jedoch alle kostenpflichtig und können somit aufgrund der Einschränkung, nur kostenlose Systeme zu verwenden, nicht eingesetzt werden.

Die beiden kostenlosen Empfehlungssysteme (Tastekid und Rottentomatoes) können wiederum keine Empfehlungen für Live-TV-Sendungen generieren, weshalb auch diese Systeme nicht für die Empfehlungsgenerierung bei LIWEST eingesetzt werden können.

Service	Kriterien					
	System ist kostenlos	System tauscht keine sensiblen Daten aus	System schlägt Live-TV-Sendungen vor	System ist auf eigene Sendungen anpassbar	System spricht junge Altersgruppe an	System verfügt über eine API
Tastekid	x	x	-	-	o	x
Jinni	-	?	x	x	o	x
Rottentomatoes	x	x	-	-	o	x
The Filter	-	x	?	x	o	x
Reddiscover	-	-	x	x	o	x
Filmaster	-	-	x	x	o	x
Contentwise	-	?	x	x	x	x
Whattorent						-
MovieLens						-

Tabelle 1: Übersicht Empfehlungssysteme

Legende:

x	System erfüllt die Anforderung
-	System erfüllt die Anforderung nicht
?	Es ist nicht bekannt ob die Anforderung erfüllt wird.
o	Kriterium ist nicht zutreffend

2.3 Überblick Filmempfehlungsansätze

Da keines der untersuchten Systeme für WatchMe verwendet werden kann, werden Ansätze von Filmempfehlungssystemen verglichen, um festzustellen, ob bzw. inwiefern einer dieser Ansätze für die Empfehlungsgenerierung bei LIWEST implementiert werden kann.

Bei den Ansätzen für Filmempfehlungen ist primär wichtig, welche Empfehlungsmethode (siehe auch Abschnitt 2.1.1) verwendet wird.

Weiters sind die verwendeten Merkmale, d. h. jene Eigenschaften, aufgrund derer die Empfehlungen generiert werden, interessant.

Außerdem ist bei einem Ansatz zur Empfehlungsgenerierung noch interessant, welche Ähnlichkeitsmaße zur Berechnung der Ähnlichkeiten zwischen Empfehlungselementen verwendet werden.

Daher werden die Ansätze hinsichtlich der folgenden Kriterien untersucht:

- Verwendete Methoden
- Verwendete Merkmale
- Verwendete Ähnlichkeitsmaße

2.3.1 Untersuchte Ansätze

Bei dem Vergleich der Ansätze wurden sowohl jene Ansätze, die den in Abschnitt 2.2.1 verglichenen Filmempfehlungssystemen zugrundeliegen, als auch solche, die im Rahmen einer Literaturanalyse ermittelt wurden, berücksichtigt.

Allerdings sind von den im Abschnitt 2.2.1 angeführten Filmempfehlungssystemen nur zu ContentWise Informationen über den Ansatz zur Empfehlungsgenerierung bekannt, sodass nur dieses eine der im Abschnitt 2.2.1 beschriebenen Systeme für den Vergleich herangezogen werden kann. Vier neue Ansätze wurden über die durchgeführte Literaturanalyse ermittelt.

2.3.1.1 Ricci et al.

Der Ansatz zur Empfehlungsgenerierung von ContentWise wird von Ricci et al. (Ricci, Rokach, Shapira, & Kantor, 2011) beschrieben. Bei ContentWise kommen zwei unterschiedliche Collaborative-Recommendation Algorithmen sowie ein Content-Based-Recommendation Algorithmus zum Einsatz. Für die Empfehlung von VoD-Sendungen kommen alle drei Algorithmen zum Einsatz, wohingegen für Live-TV-Sendungen nur der Content-Based Algorithmus verwendet wird. Die Collaborative-Recommendation Algorithmen können nicht für Live-TV-Sendungen eingesetzt werden, da bei Collaborative-Recommendation neue Sendungen erst empfohlen werden, wenn sie bereits bewertet wurden. Da dies bei Live-TV mit vielen neuen Sendungen, welche immer nur für kurze Zeit (während der Ausstrahlung) verfügbar sind, keinen Sinn macht, wird hier auf den Content-Based Algorithmus vertraut (Ricci, Rokach, Shapira, & Kantor, 2011).

Die beiden Collaborative-Recommendation Algorithmen verwenden unterschiedliche Ansätze, um die zu empfehlenden Filme zu ermitteln. Der erste Algorithmus ermittelt ähnliche Filme aufgrund von Benutzerbewertungen. Filme werden also als ähnlich bezeichnet, sobald die Benutzer die Filme ähnlich bewerten. Der zweite Collaborative-Recommendation Algorithmus versucht für den Benutzer, für welchen die Empfehlung generiert werden soll, zu ermitteln, welche Merkmale für seine Bewertungen relevant sind. Aufgrund dieser Merkmale werden ähnliche Benutzer gesucht und Filme vorgeschlagen, die von diesen Benutzern positiv bewertet wurden (Ricci, Rokach, Shapira, & Kantor, 2011).

Zur Generierung der Empfehlungen werden folgende Merkmale verwendet: Filmtitel, Genres, Schauspieler, Sendungsbeschreibung, implizite Bewertungen und Benutzerpräferenzen.

Bei ContentWise werden keine durch Benutzer direkt vergebenen Bewertungen verwendet. Stattdessen werden die Bewertungen aufgrund der durch den Benutzer konsumierten Sendungen automatisiert erstellt. Für Live-TV-Sendungen wird die Bewertung aufgrund der Sehdauer einer Sendung ermittelt. Schaut ein Benutzer die ganze Sendung an, so wird eine Bewertung von 5 angenommen, schaut er die Sendung mindestens 5 Minuten, dann wird eine Bewertung von 3 vergeben. Für VoD-Filme wird unabhängig von der Sehdauer eine Bewertung von 4 festgelegt, da angenommen wird, dass ein Benutzer nur dann bereit ist, für einen Film zu zahlen, wenn ihm der Film auch gefällt (Ricci, Rokach, Shapira, & Kantor, 2011).

Ein Problem bei der Empfehlungsgenerierung stellt die fehlende Information über den tatsächlichen Benutzer zum Zeitpunkt der Empfehlungsgenerierung dar. Da es bei Contentwise keine Authentifizierung gibt, ist der eigentliche Benutzer unbekannt. Um trotzdem auf den Benutzer schließen zu können, wird angenommen, dass Hausfrauen morgens fernsehen, Kinder nachmittags, abends die ganze Familie und die Erwachsenen in der Nacht (Ricci, Rokach, Shapira, & Kantor, 2011).

Bei dem Content-Based-Recommendation Algorithmus wird die Ähnlichkeit für alle Merkmale mittels Cosinus-Ähnlichkeit berechnet. Die Cosinus-Ähnlichkeit ist vor allem für Text- bzw. Dokumentenvergleiche sehr gut geeignet, da sie die Häufigkeit der Vorkommnisse der einzelnen Wörter bei der Berechnung berücksichtigt (Jannach, Zanker, Felfernig, & Friedrich, 2011). Die bei den beiden Collaborative-Recommendation Algorithmen verwendeten Ähnlichkeitsmaße sind nicht bekannt.

2.3.1.2 *Klahold*

Klahold (Klahold, 2009) beschreibt das Filmempfehlungssystem MovieLens. Als Methode für die Filmempfehlung setzt MovieLens auf ein Hybridverfahren aus Content-Based-Recommendation und Collaborative-Recommendation (Klahold, 2009). Eine genauere Beschreibung, warum diese Methoden verwendet werden, ist nicht verfügbar.

Für den Content-Based-Recommendation Ansatz werden sowohl die Filmbeschreibungen, die beteiligten Personen als auch die Genres miteinbezogen. Die dafür benötigten Filmmetadaten wie z. B. Genres und beteiligte Personen werden über die IMDB-Datenbank¹ bezogen. Zusätzlich können die Benutzer den Filmen Eigenschaften zuweisen, welche auch in die Ähnlichkeitsberechnung miteinbezogen werden (Klahold, 2009).

Beim Collaborative-Recommendation Algorithmus werden für die Ermittlung ähnlicher Benutzer nur die Benutzerbewertungen verwendet (Klahold, 2009). Die Ähnlichkeitsberechnungen für den Content-Based-Recommendation Algorithmus erfolgen für alle Merkmale mit einer Abwandlung des TF-IDF² Maßes (Xia & Chai, 2011). Das TF-IDF Maß ermittelt die Ähnlichkeit zweier Elemente aufgrund der Häufigkeit des Vorkommens einzelner Merkmalsausprägungen innerhalb der Elemente. Die Ähnlichkeit für jede Merkmalsausprägung wird zusätzlich aufgrund ihrer Relevanz gewichtet. Merkmalsausprägungen, welche über alle Elemente hinweg eine hohe Häufigkeit aufweisen, werden dabei als weniger relevant betrachtet als Merkmalsausprägungen, welche eine geringere Häufigkeit aufweisen.

Die Berechnungen für den Collaborative-Recommendation Ansatz werden mittels Pearson-Korrelationskoeffizient durchgeführt (Good, et al., 1999). Der Pearson-Korrelationskoeffizient misst den Grad des linearen Zusammenhangs zwischen zwei Merkmalen (Chok, 2010).

2.3.1.3 *Gupta, Jain und Song*

Gupta, Jain und Song (Gupta, Jain, & Song, 2008) beschreiben einen Ansatz, in dem für die Filmempfehlungsgenerierung Informationen aus Facebook-Profilen gesammelt werden.

Durch den Vergleich folgender Profilinformatoren werden ähnliche Benutzer ermittelt: Lieblingsfilme, Lieblingsbücher, Lieblingsmusik, Geburtstag, Wohnort, Name,

¹ <http://www.imdb.com/>

² Term Frequency – Inverse Document Frequency

Beziehungsstatus, Geschlecht, Aktivitäten und politische Ansicht (Gupta, Jain, & Song, 2008).

Für die Empfehlungsgenerierung werden folgende drei Collaborative-Recommendation Ansätze verwendet und deren Ergebnisse miteinander verglichen: Support Vector Machine (SVM) (Cristianini & Shawe-Taylor, 2000), Adapted Ranking SVM (Gupta, Jain, & Song, 2008) und K-Means Clustering (Klahold, 2009).

Beim SVM Ansatz erfolgt für jeden Film eine Unterteilung in Benutzer, die den Film mögen, und Benutzer, die den Film nicht mögen. Danach wird versucht, den Benutzer, für den die Empfehlung generiert wird, anhand seiner Profilinformationen einer der beiden Gruppen zuzuordnen. Wird der Benutzer jener Gruppe von Personen zugeordnet, die den Film mögen, wird ihm der Film empfohlen (Gupta, Jain, & Song, 2008).

Der Adapted Ranking SVM Ansatz ist eine Erweiterung des SVM Ansatzes. Dieser Ansatz ermöglicht es, die zu empfehlenden Filme in eine Rangfolge zu bringen (Gupta, Jain, & Song, 2008).

Beim K-Means-Clustering Ansatz wird versucht, Stereotypen bei jenen Benutzern zu ermitteln, welche einen bestimmten Film mögen. Darauf basierend wird überprüft, welchem Stereotyp der Benutzer, für den der Filmvorschlag erstellt werden soll, am ähnlichsten ist. Filme, die dieser Stereotyp hoch bewertet, werden dann dem Benutzer vorgeschlagen (Gupta, Jain, & Song, 2008).

Ein Vergleich der drei Ansätze ergab, dass der SVM Ansatz und der K-Means-Clustering Ansatz für eine Empfehlungsgenerierung annähernd gleich gut geeignet sind, wohingegen der Adapted Ranking SVM Ansatz weniger gut geeignet ist (Gupta, Jain, & Song, 2008).

2.3.1.4 Soares und Viana

Soares und Viana beschreiben ein Filmempfehlungssystem, welches Benutzer bei der Auswahl von Live-TV-Sendungen und auch VoD-Sendungen unterstützt (Soares & Viana, 2014).

Das Empfehlungssystem verwendet ein Hybridverfahren aus einem Collaborative-Recommendation Algorithmus und einem Content-Based-Recommendation Algorithmus.

Folgende Merkmale werden für die Empfehlungsgenerierung herangezogen: die Sendungsbewertung, der Titel, die Beschreibung und die Sendungsdauer. Zusätzlich werden für VoD-

Sendungen das Genre, die beteiligten Personen, das Produktionsdatum und das Produktionsland verwendet. Falls der Benutzer für eine Sendung keine Bewertung abgibt, wird die Bewertung von 1-5 aufgrund der vorhandenen Daten über die Sehdauer der Sendung berechnet (Soares & Viana, 2014).

Für den Collaborative-Recommendation Ansatz werden mittels Cosinus-Ähnlichkeit (Jannach, Zanker, Felfernig, & Friedrich, 2011) ähnliche Benutzer ermittelt. Danach werden Sendungen, die von diesen Benutzern positiv bewertet wurden, vorgeschlagen.

Für den Content-Based-Recommendation Ansatz werden je nach Merkmal unterschiedliche Ähnlichkeitsmaße verwendet. Für Merkmale, bei denen die Reihung der Merkmalsausprägungen eine Rolle spielt, wie z. B. bei der Liste der beteiligten Personen, wird das Inverse-Rank-Measure (Soares & Viana, 2014) verwendet. Bei diesem Ähnlichkeitsmaß wird nicht nur das Auftreten von Ausprägungen zur Berechnung der Ähnlichkeit herangezogen, sondern auch deren Reihenfolge. Dabei erhalten vorgereihte Ausprägungen eine höhere Gewichtung als nachgereihte. Für alle anderen Merkmale wird die Cosinus-Ähnlichkeit als Ähnlichkeitsmaß verwendet (Soares & Viana, 2014).

2.3.1.5 Yu et al.

Yu et al. beschreiben drei Ansätze, um Filmempfehlungen für Personengruppen zu ermitteln. Dabei wird davon ausgegangen, dass Sendungen oft von mehreren Personen gemeinsam geschaut werden. Es werden daher Ansätze für Gruppenempfehlungen vorgestellt. Diese Ansätze befassen sich hauptsächlich mit der Profilerstellung für die Personengruppen und weniger mit der eigentlichen Empfehlungsgenerierung (Yu, Zhou, Hao, & Gu, 2006).

Im ersten Ansatz wird für jede Gruppe ein eigenes Gruppenprofil mit Informationen über bevorzugte Genres und Sendungsformate angelegt. Auch Sendungsbewertungen werden direkt in diesem Profil gespeichert. Die generierten Empfehlungen werden dann auch diesem Gruppenprofil zugeordnet (Yu, Zhou, Hao, & Gu, 2006).

Im zweiten Ansatz gibt es kein eigenes Gruppenprofil. Vielmehr wird aufgrund des Profils jedes einzelnen Gruppenmitgliedes eine Liste von Filmempfehlungen ermittelt und anschließend alle diese Empfehlungslisten zusammengefasst, um einen Sendevorschlag für die gesamte Gruppe zu ermitteln (Yu, Zhou, Hao, & Gu, 2006).

Im dritten Ansatz wird aus den einzelnen Benutzerprofilen, durch das System automatisiert, ein neues Gruppenprofil ermittelt und für dieses automatisiert erstellte Profil wird dann die Sendungsempfehlung ermittelt (Yu, Zhou, Hao, & Gu, 2006).

Auch wenn der erste Ansatz der wohl genaueste und einfachste ist, haben die Ansätze zwei und drei den Vorteil, dass sie Vorschläge für dynamisch zusammengestellte Gruppen generieren können. Dies ist vor allem dann von Vorteil, wenn sich die Gruppenzusammensetzung öfters ändert. Da im ersten Ansatz eigene Profile für jede Gruppe angelegt werden müssen, müssen für diese Profile zuerst auch eine gewisse Anzahl an Bewertungen abgegeben werden, um sinnvolle Empfehlungen generieren zu können. Da Ansätze zwei und drei auf die bestehenden Bewertungen zurückgreifen können, kommen diese Ansätze ohne zusätzliche Bewertungen aus (Yu, Zhou, Hao, & Gu, 2006).

Da Yu et al. den Fokus auf die Profilgenerierung legen, sind für keinen der drei Ansätze Details zur Empfehlungsgenerierung bekannt. Für den dritten Ansatz haben Yu et al. allerdings noch folgende Details der verwendeten Ähnlichkeitsmaße bekannt gegeben.

In das Gesamtprofil werden Eigenschaften, welche in den Einzelprofilen ähnlich bewertet wurden, übernommen. Für den Vergleich der Merkmale wird die Ähnlichkeit mit Hilfe einer abgewandelten Berechnung nach Dalal berechnet. Dabei werden die Eigenschaften mit -1 (= negativ), 0 (= neutral) und 1 (= positiv) bewertet und die Ähnlichkeit der Bewertungen ermittelt (Yu, Zhou, Hao, & Gu, 2006).

Für die Empfehlungsgenerierung werden als Merkmale die Sendungsbeschreibung, das Genre und die Schauspieler verwendet. Die Ähnlichkeitsberechnung zwischen den Sendungen erfolgt hier mittels der Cosinus-Ähnlichkeit (Yu, Zhou, Hao, & Gu, 2006).

2.3.2 Vergleich der Ansätze

Die für die Empfehlungsgenerierung verwendeten Methoden werden in den unterschiedlichen Ansätzen ähnlich eingesetzt (siehe Tabelle 2).

Ansätze	Empfehlungsmethoden			
	LIVE-TV		VoD	
	Content-Based-Recommendation	Collaborative-Recommendation	Content-Based-Recommendation	Collaborative-Recommendation
Ricci et al.	x		x	x
Klahold			x	x
Gupta, Jain und Song				x
Soares und Viana	x	x	x	x
Yu et al.	?	?	?	?

Tabelle 2: Übersicht Empfehlungsmethoden

Bei Ansätzen für Live-TV-Sendungen wird auf Content-Based-Recommendation gesetzt, da Erstausstrahlungen bei Collaborative-Recommendation aufgrund fehlender Bewertungen nicht beurteilt werden können. Einzig bei dem vom Soares und Viana (Soares & Viana, 2014) beschriebenen Hybrid-Recommendation Ansatz wird ungeachtet dessen auch für Live-TV-Sendungen zusätzlich zum Content-Based-Recommendation Ansatz der Collaborative-Recommendation Ansatz eingesetzt.

Bei der Empfehlung von VoD-Sendungen wird in den Ansätzen von Ricci et al. (Ricci, Rokach, Shapira, & Kantor, 2011), Klahold (Klahold, 2009) und Soares und Viana (Soares & Viana, 2014) auf Hybrid-Recommendation Ansätze, welche aus einer Content-Based-Recommendation und einer Collaborative-Recommendation bestehen, gesetzt. Nur der Ansatz von Gupta, Jain und Song (Gupta, Jain, & Song, 2008) verwendet einen reinen Collaborative-Recommendation Ansatz.

Durch die Verwendung von Hybridansätzen für Empfehlungen von VoD-Sendungen wird auf der einen Seite das Problem gelöst, dass durch einen reinen Collaborative-Recommendation Ansatz unbewertete Filme nicht vorgeschlagen werden, zum anderen werden überraschendere Empfehlungen als bei reiner Content-Based-Recommendation erzielt (siehe auch Abschnitt 2.1.2).

Bei dem von Yu et al. vorgestellten Ansatz der Empfehlungsgenerierung für Personengruppen ist die Methode zur Empfehlungsgenerierung nicht bekannt.

Die von den Ansätzen verwendeten Merkmale können prinzipiell in drei Bereiche gegliedert werden:

- Sendungsdaten
- Benutzerpräferenzen
- Benutzerinformationen

Fast alle Ansätze verwenden Sendungsdaten für ihre Empfehlungen (siehe Tabelle 3). Nur der Ansatz von Gupta, Jain und Song (Gupta, Jain, & Song, 2008) lässt diese Daten außen vor und beschränkt sich rein auf Benutzerpräferenzen und Benutzerinformationen, um ähnliche Benutzer zu finden und somit Filme zu empfehlen.

Ansätze	Sendungsdaten						
	Filmtitel	Produktionsland	Beschreibung	Genre	Beteiligte Personen	Produktionsjahr	Dauer
Ricci et al.	x	x	x	x	x	x	x
Klahold			x	x			
Gupta, Jain und Song							
Soares und Viana	x	x	x	x	x	x	x
Yu et al.			x	x	x		

Tabelle 3: Merkmalsübersicht Sendungsdaten

Der von Klahold (Klahold, 2009) beschriebene Ansatz ist der Einzige, der keine Benutzerpräferenzen verwendet (siehe Tabelle 4). Bei allen anderen Ansätzen werden Benutzerpräferenzen verwendet, jedoch unterscheidet sich die Ermittlung dieser Daten bei den einzelnen Ansätzen. Ansätze, bei denen Live-TV-Sendungen über IPTV Technologie ausgestrahlt werden ((Ricci, Rokach, Shapira, & Kantor 2011), (Soares & Viana, 2014)), ermitteln die bevorzugten Genres und die Bewertungen aus dem Sehverhalten der Benutzer. Dazu wird analysiert, welche Sendungen und Kanäle geschaut werden, welche Genres der Kunde bevorzugt und wie lange Sendungen geschaut werden. Im Ansatz von Gupta, Jain und Song (Gupta, Jain, & Song, 2008) werden diese Informationen hingegen explizit beim Benutzer abgefragt.

Ansätze	Benutzerpräferenzen						
	Bevorzugte Genres	Explizite Bewertungen	Implizite Bewertungen	Lieblingsfilme	Lieblingsbücher	Lieblingsmusik	Lieblingsschauspieler
Ricci et al.	x		x				x
Klahold							
Gupta, Jain und Song				x	x	x	
Soares und Viana	x	x	x				
Yu et al.	x						x

Tabelle 4: Merkmalsübersicht Benutzerpräferenzen

Zusätzliche Informationen über den Benutzer werden nur im Ansatz von Gupta, Jain und Song (Gupta, Jain, & Song, 2008) verwendet (siehe Tabelle 5). Im Detail werden in diesem Ansatz das Geburtsdatum, der Wohnort, der Beziehungsstatus, das Geschlecht, politische Ansichten, Aktivitäten und die Freundesliste aus dem Facebook-Profil des Benutzers ausgelesen.

Ansätze	Benutzerinformationen						
	Geburtsdatum	Wohnort	Beziehungsstatus	Geschlecht	Aktivitäten	Freundesliste	Politische Ansichten
Ricci et al.							
Klahold							
Gupta, Jain und Song	x	x	x	x	x	x	x
Soares und Viana							
Yu et al.							

Tabelle 5: Merkmalsübersicht Benutzerinformationen

Bei den verwendeten Ähnlichkeitsmaßen erklären sich die Unterschiede hauptsächlich durch die unterschiedlichen verwendeten Merkmale.

Für die Ähnlichkeitsberechnung von Filmbeschreibungen wird bei den Ansätzen von Ricci et al. (Ricci, Rokach, Shapira, & Kantor 2011), Soares und Viana (Soares & Viana, 2014) und

Yu et al. (Yu, Zhou, Hao, & Gu, 2006) die Cosinus-Ähnlichkeit verwendet (siehe Tabelle 6). Im Ansatz von Klahold (Klahold, 2009) wird dafür die TF-IDF verwendet. Einzig im Ansatz von Gupta, Jain und Song (Gupta, Jain, & Song, 2008) wird dieses Merkmal gar nicht zur Empfehlungsgenerierung herangezogen.

Für Merkmale, bei denen nur das Vorhandensein einer Ausprägung geprüft wird (z. B. Genres), wird bei dem von Klahold (Klahold, 2009) beschriebenen Ansatz auf die Pearson-Korrelation (Good, et al., 1999) zurückgegriffen. In den anderen Ansätzen wird auch dafür die Cosinus-Ähnlichkeit verwendet.

Um die Reihenfolge der Nennung der beteiligten Personen zu berücksichtigen, wird bei dem Ansatz von Soares und Viana (Soares & Viana, 2014) für die Ähnlichkeitsberechnung anhand der beteiligten Personen das Inverse Rank Measure verwendet.

Im Ansatz von Yu et al. (Yu, Zhou, Hao, & Gu, 2006) wird die Ähnlichkeit von Merkmalen, die Benutzerpräferenzen wiedergeben, wie z. B. die Beliebtheit von Genres, mit der Dalal's Distance berechnet.

Die von Gupta, Jain und Song (Gupta, Jain, & Song, 2008) beschriebenen Ansätze verwenden anstelle der Ähnlichkeitsmaße drei Machine Learning Algorithmen, um die Ähnlichkeit zwischen den Benutzern zu ermitteln.

Ansätze	Ähnlichkeitsberechnung			Machine Learning
	Ähnlichkeits-/Distanzmaß			
	Cosinus-Ähnlichkeit bzw. TF-IDF	Pearson-Korrelation	Dalal's Distance	Inverse Rank Measure
Ricci et al.	x			
Klahold	x	x		
Gupta, Jain und Song				x
Soares und Viana	x			x
Yu et al.	x		x	

Tabelle 6: Ähnlichkeitsberechnung

2.4 Fazit

Der Vergleich der Filmempfehlungssysteme im Unterkapitel 2.2 zeigt, dass es Filmempfehlungssysteme gibt, welche die inhaltlichen Anforderungen von LIWEST erfüllen. Diese sind jedoch kostenpflichtig und kommen somit nicht für die Empfehlungsgenerierung bei LIWEST in Frage. Daher wurden im Unterkapitel 2.3 verschiedene Ansätze von Filmempfehlungssystemen verglichen, um festzustellen, ob bzw. inwiefern einer dieser Ansätze für die Empfehlungsgenerierung bei LIWEST implementiert werden kann.

Von den untersuchten Ansätzen sind nur die Ansätze von Ricci et al. (Ricci, Rokach, Shapira & Kantor, 2011) und von Soares und Viana (Soares & Viana, 2014) für die Empfehlungsgenerierung von Live-TV-Sendungen geeignet.

Beide Ansätze haben jedoch Nachteile, wodurch eine direkte Implementierung der jeweiligen Ansätze als nicht sinnvoll betrachtet wird. Insbesondere verwenden beide Ansätze zur Empfehlungsgenerierung für Live-TV-Sendungen nur einen Teil der verfügbaren Sendungsdaten und auch keine Informationen über die Lieblingsschauspieler und Lieblingsfilme der Benutzer.

3 Empfehlungssystem WatchMe

Aufgrund der Ergebnisse des Vergleichs der Filmempfehlungssysteme wurde die Entscheidung getroffen, dass es für eine Empfehlungsgenerierung, die den LIWEST Anforderungen entspricht, am sinnvollsten ist, einen eigenen Ansatz zu entwickeln.

Da LIWEST ausschließlich Live-TV-Sendungen anbietet, wird, so wie auch im Ansatz von Ricci et al. (Ricci, Rokach, Shapira, & Kantor, 2011), ein Content-Based Ansatz verfolgt.

Im Folgenden werden die für die Empfehlungsgenerierung verwendeten Merkmale und die jeweilige Art der Ähnlichkeitsberechnung im Unterkapitel 3.1 beschrieben. Im Unterkapitel 3.2 wird die sogenannte Schnellempfehlung erörtert. Diese kommt zum Einsatz, wenn zum Zeitpunkt der Empfehlungsgenerierung noch nicht genügend Informationen über Benutzervorlieben zur Verfügung stehen. Im Unterkapitel 3.3 wird die sogenannte vollständige Empfehlungsgenerierung beschrieben, die zum Einsatz kommt, wenn bereits genügend Informationen über Benutzervorlieben vorliegen.

3.1 Verwendete Merkmale

Die für die Empfehlungen herangezogenen Merkmale setzen sich aus Sendungsinformationen, Benutzerpräferenzen und Benutzerinformationen zusammen. Im Gegensatz zu IPTV-Lösungen kann bei LIWEST für Benutzerpräferenzen nicht auf das Sehverhalten der Benutzer zurückgegriffen werden, um bevorzugte Genres sowie implizite Bewertungen zu ermitteln. Daher werden die Genrepräferenzen und die Bewertungen zu Sendungen direkt vom Benutzer abgefragt. Zusätzlich werden Sendungsbewertungen über die Verbindung zum Facebook-Profil des Benutzers ermittelt.

Folgende Merkmale werden für die Empfehlungsgenerierung herangezogen:

Sendungsbeschreibung: Die Ähnlichkeitsberechnung für die Sendungsbeschreibung wird mittels Cosinus-Ähnlichkeit ermittelt. Die Cosinus-Ähnlichkeit liefert vor allem bei Text- bzw. Dokumentenvergleichen sehr gute Ergebnisse, da sie für den Vergleich die Häufigkeit der Vorkommnisse der einzelnen Wörter berücksichtigt. Die Cosinus-Ähnlichkeit kann Werte zwischen 0 (keine Ähnlichkeit) und 1 (identisch) annehmen (Jannach, Zanker, Felfernig, & Friedrich, 2011).

Um nur die tatsächlich für den Inhalt relevanten Wörter zu vergleichen, werden von den Sendungsbeschreibungen zuvor noch Stoppwörter herausgefiltert.

Genre: Die Ähnlichkeit von zwei Sendungen bezüglich ihrer Genres wird mittels Pearson-Korrelationskoeffizienten ermittelt. Der Pearson-Korrelationskoeffizient misst den Grad des linearen Zusammenhangs zwischen zwei Merkmalen. Der kann Werte zwischen -1 (stark negativer Zusammenhang) und 1 (stark positiver Zusammenhang) annehmen (Chok, 2010).

Für den Vergleich der Genres wird für die zu vergleichenden Sendungen für alle Genres angegeben, ob die jeweilige Sendung zu diesem Genre gehört (= 1) oder nicht dazu gehört (= 0). Die beiden daraus entstehenden Vektoren werden mittels Pearson-Korrelationskoeffizienten miteinander verglichen (Klahold, 2009).

Beteiligte Personen: Auch der Vergleich anhand der an einer Sendung beteiligten Personen wird durch die Berechnung des Pearson-Korrelationskoeffizienten durchgeführt. Alle an der Sendung beteiligten Personen werden mit 1 gekennzeichnet, alle nicht beteiligten mit 0. Dabei macht es keinen Unterschied, in welcher Funktion (Schauspieler, Regisseur) die Person beteiligt ist.

Produktionsjahr: Um die Ähnlichkeit von zwei Sendungen bezüglich ihres Produktionsjahres zu vergleichen, wird die absolute Differenz der jeweiligen Produktionsjahre herangezogen.

Produktionsländer: Die Ähnlichkeitsberechnung für die Produktionsländer erfolgt wie für die Genres und die beteiligten Personen mittels Pearson-Korrelationskoeffizienten.

Dauer: Bei der Ähnlichkeitsberechnung der Dauer wird die absolute Differenz der Sendungsdauer in Minuten zur Berechnung herangezogen.

Bewertungen: Für die Ähnlichkeitsberechnung aufgrund der Bewertungen wird für jede Sendung die Durchschnittsbewertung aus den Bewertungen aller Benutzer für diese Sendung ermittelt und die absolute Differenz der durchschnittlichen Sendungsbewertungen ermittelt.

Lieblingsschauspieler: Sendungen können sich auch anhand der Anzahl der mitspielenden Stars ähnlich sein. So gibt es Sendungen mit sehr vielen Stars (z. B. „Departed“, „Ocean’s Eleven“), Sendungen ganz ohne Stars (z. B. „Das Hochzeitsvideo“) und Sendungen, in denen nur wenige Stars mitspielen (z. B. „My Own Private Idaho“, „Before Sunrise“).

Ob ein Schauspieler ein Star ist, wird für WatchMe anhand der Benutzer, die diesen Schauspieler als Lieblingsschauspieler angegeben haben, definiert.

Bei dem Vergleich von zwei Sendungen bezüglich der Lieblingsschauspieler geht es daher nicht um bestimmte Personen an sich, sondern darum, ob in einem Film viele beliebte Schauspieler oder wenige bzw. keine beliebten Schauspieler mitspielen. Um zwei Sendungen anhand der Anzahl der beteiligten Lieblingsschauspieler zu vergleichen, wird für beide Sendungen für die beliebtesten drei Schauspieler die Anzahl der Nennungen als Lieblingsschauspieler ermittelt. Die Summe der Differenzen aus der Anzahl der jeweiligen Nennungen zwischen den beliebtesten Schauspielern, den zweitbeliebtesten Schauspielern und den drittb Liebtesten Schauspielern ergibt das Ergebnis.

In Tabelle 7 wird der Vergleich der beliebtesten Schauspieler der Filme „Mr. & Mrs. Smith“ und „Date Doctor“ dargestellt, wobei die jeweilige Anzahl der Nennungen für dieses Beispiel frei gewählt wurde. In diesem Beispiel ist Angelina Jolie die beliebteste Schauspielerin des Filmes „Mr. & Mrs. Smith“ mit 700 Nennungen. Bei „Date Doctor“ ist es Will Smith mit 500 Nennungen – die Differenz ergibt 200. Die Differenz der zweitbeliebtesten Schauspieler Brad Pitt und Kevin James mit 650 bzw. 500 Nennungen ergibt 150 und die Differenz der drittb Liebtesten Schauspieler Vince Vaughn und Eva Mendes ergibt 350 – in Summe ergibt dies eine Differenz von 700.

Sendung	Beliebtester Schauspieler	Zweitbeliebtester Schauspieler	Drittb Liebtestester Schauspieler	Summe
Mr. & Mrs. Smith	Angelina Jolie 700	Brad Pitt 650	Vince Vaughn 100	
Date Doctor	Will Smith 500	Kevin James 500	Eva Mendes 450	
Differenz	200	150	350	700

Tabelle 7: Vergleich beliebteste Schauspieler

Folgende Merkmale, die in den im Abschnitt 2.3.1 beschriebenen Ansätzen verwendet werden, werden bei WatchMe nicht berücksichtigt: Sendungstitel, Lieblingsbücher, Lieblingsmusik, Wohnort, Beziehungsstatus, Geschlecht, Aktivitäten, politische Ansichten.

Der Sendungstitel wird nicht berücksichtigt, da ein zusätzlicher Vergleich von Sendungen über den Titel nicht zu einer verbesserten Ergebnisliste führt. Thematisch ähnliche Sendungen werden über die Sendungsbeschreibung ermittelt und Fortsetzungen werden sowohl über die Ähnlichkeit der Beschreibung als auch über die Ähnlichkeit bei den beteiligten Personen als ähnliche Sendungen erkannt.

Alle anderen Merkmale werden nicht verwendet, da sie zwar für die Ermittlung ähnlicher Benutzer für Collaborative-Recommendation Ansätze eingesetzt werden, jedoch für die Empfehlungsgenerierung bei einem Content-Based-Recommendation Ansatz keine Relevanz besitzen.

3.2 Schnellempfehlung

Da für eine Empfehlungsgenerierung durch eine Content-Based-Recommendation Sendungsbewertungen des Benutzers vorhanden sein müssen, wird die vollständige Empfehlungsgenerierung mittels Content-Based Ansatz erst durchgeführt, wenn eine Mindestanzahl von Sendungsbewertungen vorliegt. Bis diese Anzahl erreicht ist, wird eine Schnellempfehlung durchgeführt. Diese Mindestanzahl von Bewertungen ist konfigurierbar.

Für die Schnellempfehlung werden die Genres der potentiellen Sendungen mit den Genrepräferenzen des Benutzers abgeglichen. Potentielle Sendungen sind jene Sendungen, die zum gewählten Zeitpunkt ausgestrahlt werden und deren Restlaufzeit einem vom Benutzer konfigurierbaren Mindestwert entspricht. Dieser Mindestwert wird in Prozent angegeben, da ansonsten kurze Sendungen, wie zum Beispiel Nachrichtensendungen, nie vorgeschlagen werden würden.

Die Genrepräferenzen muss der Benutzer eingeben, bevor er eine Empfehlung ermitteln lassen kann. Genrepräferenzen werden je Sendungskategorie angegeben. Genres, die der Benutzer mag, werden mit 1 bewertet, Genres, die der Benutzer nicht mag, mit -1 und alle anderen Genres mit 0.

Für die Schnellempfehlungsgenerierung wird für jede Sendung der Empfehlungswert E mit der in Abbildung 1 dargestellten Formel berechnet. Es wird die Summe der Präferenzwerte p für das jeweilige Genre g über alle Genres G der Sendung s berechnet. Anschließend wird diese Summe durch die Anzahl der Genres dieser Sendung n dividiert.

$$E_s = \frac{\sum_{g \in G} p_g}{n}$$

Abbildung 1: Berechnung Schnellempfehlung

Vorgeschlagen werden dann die drei Sendungen mit den höchsten Empfehlungswerten. Haben mehrere Sendungen den gleichen Empfehlungswert, werden diese anhand der prozentuellen Restlaufzeit absteigend gereiht.

3.3 Vollständige Empfehlung

Bei dem für die Empfehlungsgenerierung verwendeten Content-Based-Recommendation Ansatz wird die Ähnlichkeit zwischen allen potentiellen Sendungen und allen positiv bewerteten Sendungen, das sind Sendungen mit einer Bewertung von 4 oder 5, berechnet.

Um die Ähnlichkeit zwischen zwei Sendungen zu berechnen, wird zuerst die Ähnlichkeit für jedes einzelne Merkmal mit Hilfe der jeweiligen Ähnlichkeitsfunktion berechnet.

Abbildung 2 zeigt dies exemplarisch am Beispiel der Genre der bereits positiv bewerteten Sendung „22 Bullets“ und der potentiellen Sendungen „South Park“, „3 Days to kill“ und „Windtalkers“. Die Ähnlichkeit der Genres wird mit dem Pearson-Korrelationskoeffizient (Chok, 2010) berechnet.

Der Film „22 Bullets“ ist den Genres Action, Krimi und Thriller zugeordnet, die Serie „South Park“ den Genres Animation und Komödie. Der Vergleich der beiden Sendungen mittels Pearson-Korrelation ergibt einen Wert von -0,0387492. „3 Days to Kill“ ist den Genres Action, Drama, Krimi und Thriller zugeordnet, was im Vergleich zu „22 Bullets“ einen Wert von 0,736111 ergibt. Der Vergleich mit „Windtalkers“, welcher den Genres Action, Drama, Geschichte und Kriegsfilm zugeordnet ist, ergibt den Wert von 0,208333.

Der Film „3 Days to Kill“ ist hinsichtlich der Genres dem Film „22 Bullets“ also am ähnlichsten, die Serie „South Park“ ist „22 Bullets“ am wenigsten ähnlich.

Genre je Sendung

	Action	Animation	Drama	Geschichte	Komödie	Kriegsfilm	Krimi	Sci-Fi	Sport	Thriller
22 Bullets	1	0	0	0	0	0	1	0	0	1
South Park	0	1	0	0	1	0	0	0	0	0
3 Days to Kill	1	0	1	0	0	0	1	0	0	1
Windtalkers	1	0	1	1	0	1	0	0	0	0



Ergebnis Pearson-Korrelation

	South Park	3 Days to Kill	Windtalkers
22 Bullets	-0,0387492	0,736111	0,208333

Abbildung 2: Ähnlichkeitsermittlung Genre

Nachdem für ein Merkmal die Ähnlichkeitswerte ermittelt wurden, werden diese Werte mit Hilfe der Min/Max-Normalisierung (Cleve & Lämmel, 2014) normalisiert. Dadurch werden die Werte in den Wertebereich von 0 bis 1 transformiert. Dies ermöglicht einen späteren Vergleich bzw. eine Kombination der Werte.

Das Ergebnis der Normalisierung im Beispiel der Ähnlichkeitsberechnung zwischen dem Film „22 Bullets“ und den Sendungen „South Park“, „3 Days to Kill“ und „Windtalkers“ von zuvor wird in Tabelle 8 dargestellt. Die Min/Max-Normalisierung ergibt für „South Park“ einen Wert von 0, für „3 Days to Kill“ einen Wert von 1 und für „Windtalkers“ einen Wert von 0,319.

Positiv bewertete Sendungen	Potentielle Sendungen		
	South Park	3 Days to Kill	Windtalkers
22 Bullets	0	1	0,319

Tabelle 8: Normalisierte Ähnlichkeitswerte vom Vergleich der Genres

Um die einzelnen Merkmale unterschiedlich gewichten zu können, wird für jedes Merkmal ein Faktor festgelegt, mit dem die normalisierten Werte multipliziert werden. Durch die Gewichtung der Merkmale wird festgelegt, wie stark das jeweilige Merkmal die Ähnlichkeitsberechnung beeinflusst.

In diesem Beispiel wird für alle Faktoren der Wert 1 festgelegt. Tabelle 9 gibt einen Überblick über die normalisierten und gewichteten Werte aller Merkmale im Vergleich zwischen „3 Days to Kill“ und „22 Bullets“. Dabei ist ersichtlich, dass für die Merkmale Genres, Lieblingsschauspieler, Produktionsland und Bewertungen der Höchstwert von 1 erzielt wird, die Dauer mit 0,999, das Produktionsjahr mit 0,96 und die Beschreibung mit 0,81 ebenfalls sehr hohe Werte erzielen, wohingegen der Vergleich der beteiligten Personen beider Filme lediglich einen Wert von 0,186 erzielt. In Summe ergibt dies den Wert von 6,82. Bei einem maximal möglichen Wert von 8 kann dies als hohe Ähnlichkeit angesehen werden.

Sendungen	Merkmale								Summe
	Genres	Lieblingsschauspieler	Dauer	Produktionsjahr	Produktionsland	Beteiligte Personen	Beschreibung	Bewertungen	
3 Days to Kill (vgl. mit 22 Bullets)	1	1	0,999	0,96	1	0,186	0,81	1	6,82

Tabelle 9: Gewichtete Werte aller Merkmale

Um einen Wert zu erhalten, der auch bei dem Hinzufügen von weiteren Merkmalen vergleichbar bleibt, wird mit der in Abbildung 3 dargestellten Formel für den Vergleich jeder potentiellen Sendung p mit jeder positiv bewerteten Sendung s der Empfehlungswert E berechnet. Der Empfehlungswert gibt die prozentuelle Erreichung der maximal erreichbaren Werte an. Dazu wird zuerst die Summe der normalisierten Ähnlichkeitswerte v je Merkmal m über alle Merkmale M berechnet und mit 100 multipliziert und danach durch die Summe der Faktoren f je Merkmal über alle Merkmale dividiert.

$$E_{p,s} = \frac{\sum_{m \in M} v_m * 100}{\sum_{m \in M} (f_m * 1)}$$

Abbildung 3: Berechnung prozentueller Erreichung der maximal erreichbaren Werte

Tabelle 10 zeigt beispielhaft die prozentuelle Ähnlichkeit zwischen sieben potentiellen Sendungen und sieben von einem Benutzer bereits positiv bewerteten Sendungen. Die Ähnlichkeit zwischen der potentiellen Sendung „3 Days to Kill“ und der vom Benutzer positiv bewerteten Sendung „22 Bullets“ beträgt beispielsweise 85,25 % und ist somit als hoch einzustufen.

Positiv bewertete Sendungen	Potentielle Sendungen						
	Die Unglaublichen	Pfarrer Braun	4 Frauen & die Liebe	CSI Miami	South Park	3 Days to Kill	Windtalkers
Dracula	47,77 %	48,25 %	46,54 %	46,74 %	58,16 %	49,37 %	58,58 %
Argo	49,81 %	50,22 %	59,11 %	49,91 %	62,34 %	58,80 %	61,29 %
Ocean's Eleven	37,85 %	44,23 %	36,83 %	46,84 %	48,46 %	46,95 %	58,46 %
Hawaii 5-0	55,88 %	55,99 %	62,27 %	57,12 %	63,07 %	56,87 %	73,17 %
Face Off	45,31 %	50,42 %	43,91 %	50,12 %	55,51 %	56,41 %	77,64 %
How I met your mother	51,26 %	52,57 %	51,79 %	50,07 %	78,32 %	49,55 %	58,00 %
22 Bullets	49,97 %	57,36 %	50,47 %	57,06 %	49,64 %	85,25 %	48,24 %

Tabelle 10: Übersicht Ähnlichkeiten von Sendungen

Das Ergebnis der vollständigen Empfehlung stellen jene drei potentiellen Sendungen dar, die die größte prozentuelle Ähnlichkeit zu einer der Sendungen haben, die vom Benutzer bereits positiv bewertet wurden.

Im Beispiel von zuvor sind das die Sendungen „Windtalkers“, „South Park“ und „3 Days to Kill“.

4 Realisierung WatchMe

In diesem Kapitel wird ein Überblick über die Realisierung von WatchMe gegeben. Es werden die von WatchMe implementierten Prozesse, dargestellt als BPMN³-Diagramme, die verwendeten Datenstrukturen, dargestellt als UML⁴-Diagramme, sowie die einzelnen Komponenten und das Zusammenspiel zwischen den Komponenten beschrieben.

4.1 Prozesse

Um die Empfehlungsgenerierung durchführen zu können, müssen folgende Funktionalitäten zur Verfügung gestellt werden. Der Benutzer muss sich authentifizieren können, damit bekannt ist, für welchen Benutzer eine Empfehlung generiert werden soll. Um Informationen über die Präferenzen des Benutzers zu erhalten, muss dieser seine Benutzerdaten verwalten können, und um Sendungsbewertungen zu bekommen, muss der Benutzer die Möglichkeit erhalten, Sendungen zu bewerten. Außerdem müssen die den Empfehlungen zugrunde liegenden Sendungsdaten zur Verfügung stehen. Der Zugriff auf die Sendungsdaten könnte ad hoc bei jeder Empfehlungsgenerierung erfolgen. Da diese Daten jedoch von externen Anbietern bezogen werden müssen, kann ein Zugriff zur Laufzeit nicht immer sichergestellt werden und die Antwortzeiten können durch den zusätzlichen externen Zugriff erhöht werden. Daher werden die Sendungsdaten lokal redundant gespeichert.

Bei den Funktionalitäten wird zwischen solchen, die der Benutzer anstoßen kann (benutzergesteuerte Funktionalitäten), und solchen, die vom System automatisiert durchgeführt werden (automatisierte Funktionalitäten), unterschieden.

Die benutzergesteuerten Funktionalitäten werden direkt vom Benutzer angestoßen und müssen daher über eine entsprechende Benutzeroberfläche zur Verfügung gestellt werden. Die automatisierten Funktionalitäten werden nicht vom Benutzer angestoßen, sondern automatisiert durch WatchMe.

³ <http://www.bpmn.org/>

⁴ <http://www.uml.org/>

Funktionalität	Typ
Authentifizierung	Benutzergesteuert
Sendungsbewertung	Benutzergesteuert
Benutzerverwaltung	Benutzergesteuert
Empfehlungsgenerierung	Benutzergesteuert
Sendungsdaten-Import	Automatisiert
Benutzerdaten-Import	Automatisiert

Tabelle 11: WatchMe Funktionalitäten

4.1.1 Authentifizierung durchführen

In Abbildung 4 wird der Prozess der Authentifizierung des Benutzers dargestellt. Der Benutzer kann sich mit den Logindaten seines Facebook-Accounts authentifizieren. Dadurch ist bekannt, für wen Empfehlungen generiert werden sollen. Versucht der Benutzer sich mit falschen Logindaten zu authentifizieren, wird eine Fehlermeldung ausgegeben.

Bei korrekten Logindaten wird überprüft, ob es bereits eine Zuordnung des Benutzers zu einem LIWEST Kunden gibt. Falls noch keine Zuordnung vorhanden ist wird der Benutzer aufgefordert, seine Kundennummer und sein Kundenkennwort einzugeben, woraufhin die Benutzerzuordnung durchgeführt wird. Die Überprüfung, ob zu den Benutzerdaten eine Zuordnung vorhanden ist, erfolgt nicht bereits bei der Installation der App, sondern bei jedem Start, da sich an einem Gerät unterschiedliche Benutzer anmelden können und für jeden Benutzer die Zuordnung geprüft werden muss.

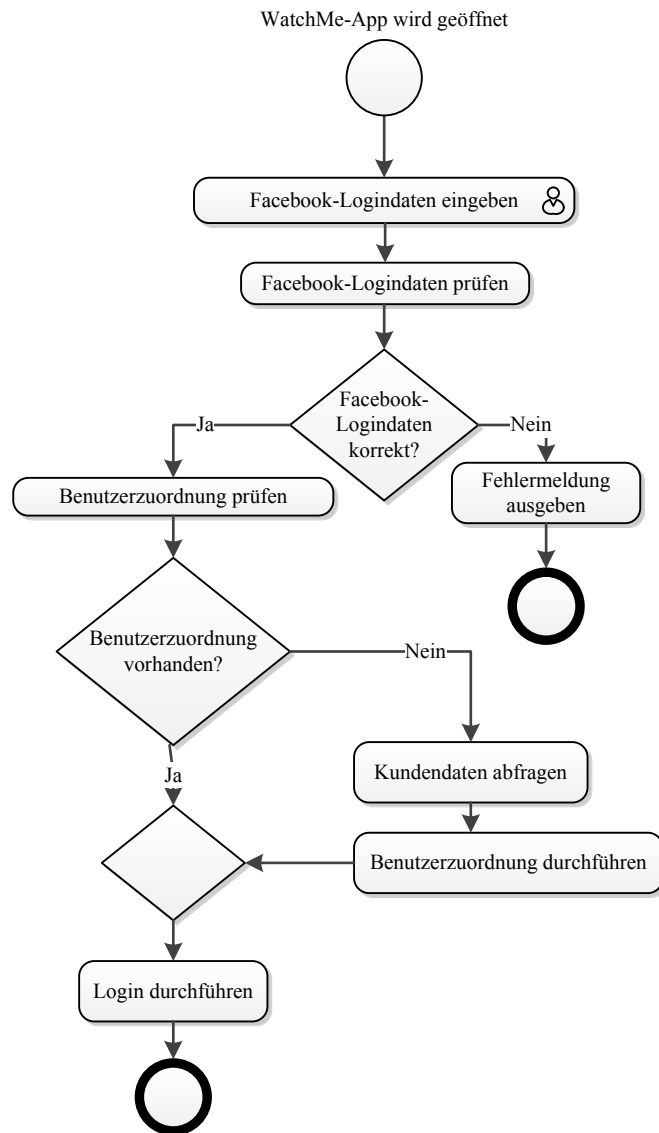


Abbildung 4: Prozess Authentifizierung durchführen

4.1.2 Sendungsbewertung durchführen

In Abbildung 5 wird der Prozess zum Durchführen der Sendungsbewertung dargestellt. Die Sendungsbewertung ermöglicht es dem Benutzer, beliebige Sendungen auf einer Skala von 1 (Sendung hat dem Benutzer gar nicht gefallen) bis 5 (Sendung hat dem Benutzer sehr gut gefallen) zu bewerten.

Benutzer möchte Sendungen bewerten

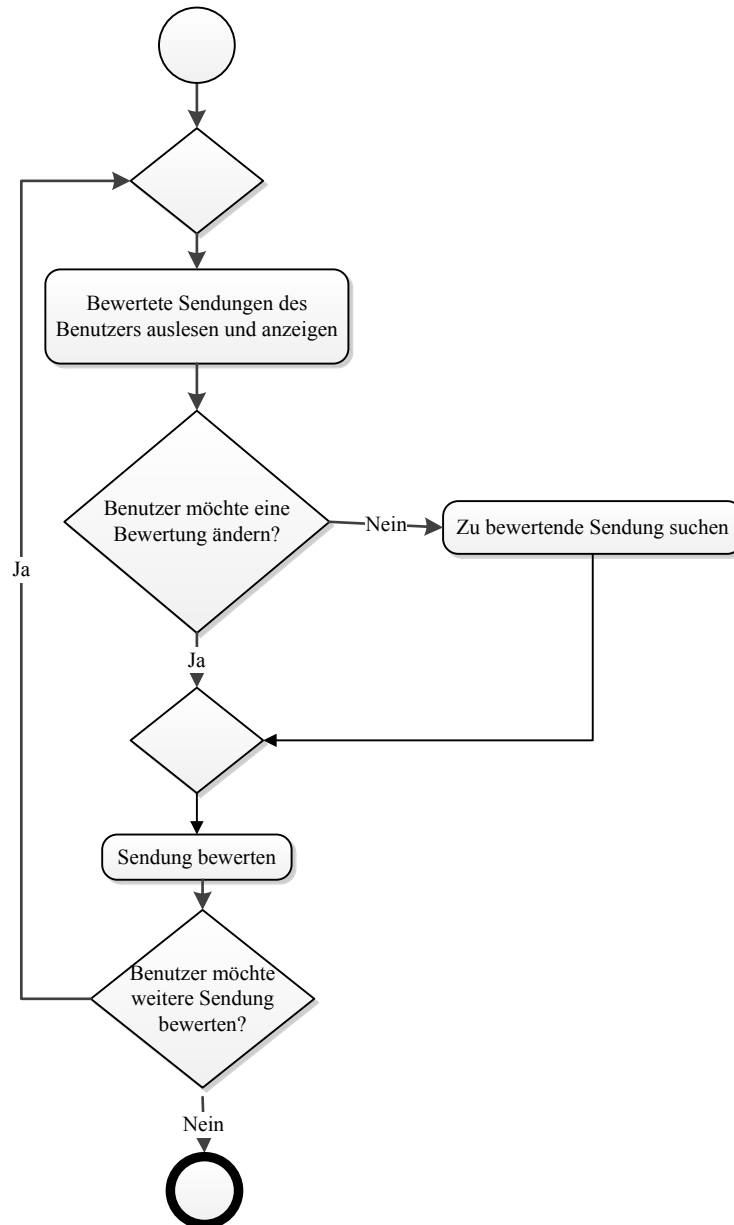


Abbildung 5: Prozess Sendungsbewertung durchführen

Um die Sendungsbewertung über die WatchMe-App durchzuführen, werden alle bereits bewerteten Sendungen ausgelesen und angezeigt. Der Benutzer kann die Bewertungen dieser Sendungen ändern oder über die Suchfunktion Sendungen, die er bewerten möchte, suchen.

Diese Schritte wiederholt der Benutzer so oft, bis er keine weiteren Bewertungen mehr vornehmen möchte.

4.1.3 Benutzerverwaltung durchführen

Der in Abbildung 6 dargestellte Prozess der Benutzerverwaltung ermöglicht es dem Benutzer folgende Informationen, welche für die Empfehlungsgenerierung relevant sind, zu pflegen:

Mindestanzahl der Bewertungen: Die Mindestanzahl der Bewertungen gibt an, wie viele Sendungen bewertet sein müssen, damit eine vollständige Empfehlung durchgeführt wird.

Minimale Restlaufzeit: Die minimale Restlaufzeit gibt an, wie lange die prozentuelle Restlaufzeit einer Sendung zum Empfehlungszeitpunkt sein muss, damit sie bei der Empfehlungsgenerierung berücksichtigt wird.

Individuelle Beliebtheit der Genres je Kategorie: Die individuelle Beliebtheit der Genres je Kategorie gibt für jede Kategorie an, welche Genres der Benutzer mag und welche er nicht mag.

Nach der Bearbeitung der Benutzerinformationen werden diese wieder in der Datenbank gespeichert.

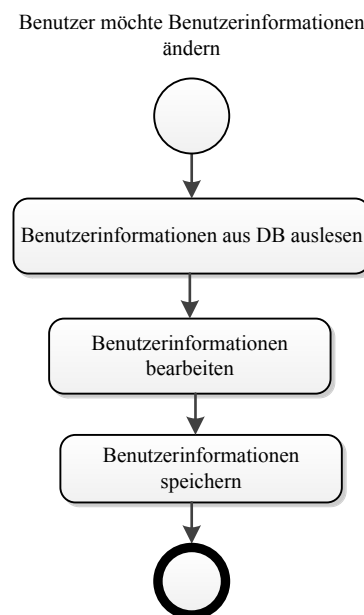


Abbildung 6: Prozess Benutzerverwaltung durchführen

4.1.4 Empfehlungsgenerierung durchführen

In Abbildung 7 ist der Prozess der Empfehlungsgenerierung dargestellt. Um eine Empfehlung generieren zu lassen, muss der authentifizierte Benutzer den gewünschten Zeitpunkt

auswählen. Das System prüft anhand der Anzahl der bisher bewerteten Sendungen, ob eine vollständige Empfehlung oder eine Schnellempfehlung durchgeführt wird.

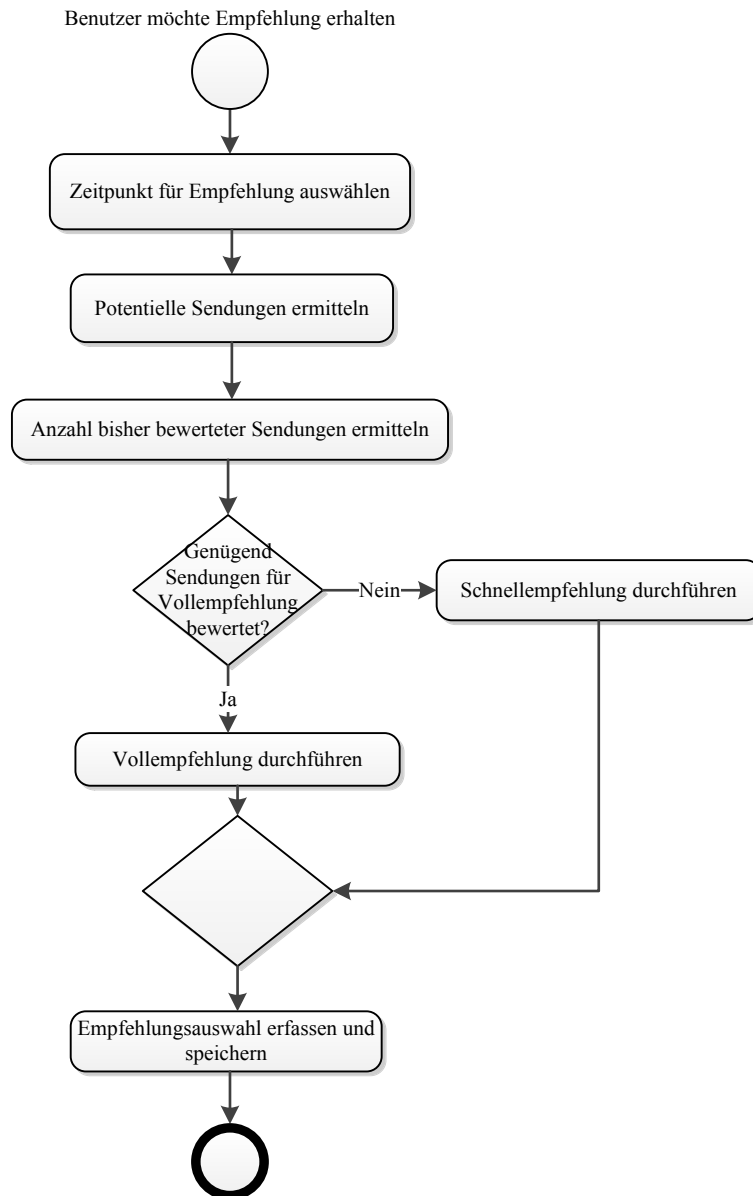


Abbildung 7: Prozess Empfehlungsgenerierung durchführen

Die Mindestanzahl der Bewertungen für eine vollständige Empfehlung kann der Benutzer in seinen Benutzereinstellungen festlegen. Für die Schnellempfehlung werden die Genres der potentiellen Sendungen mit der individuellen Beliebtheit der Genres je Kategorie verglichen und jene Sendungen, deren Genres am ehesten den Vorlieben des Benutzers entsprechen, vorgeschlagen. Bei der vollständigen Empfehlung werden alle potentiellen Sendungen mit

den durch den Benutzer bereits positiv bewerteten Sendungen verglichen. Die Sendungen, die den positiv bewerteten Sendungen am ähnlichsten sind, werden dem Benutzer als Empfehlung vorgeschlagen. Der genaue Ablauf der vollständigen Empfehlungsgenerierung ist im Unterkapitel 3.3 beschrieben.

Die Empfehlungsgenerierung liefert bis zu drei Empfehlungen. Der Benutzer hat die Möglichkeit bekannt zu geben, welche dieser empfohlenen Sendungen er beabsichtigt tatsächlich anzusehen. Die vom Benutzer angegebene Sendung wird vom System mit dem Wert 3 bewertet. Durch diese implizite Bewertung wird die Qualität zukünftiger Empfehlungen erhöht.

4.1.5 Sendungsdaten-Import

Abbildung 8 zeigt den Prozess zum Importieren von Sendungsdaten, bei dem alle für die Empfehlungsgenerierung benötigten Daten der Sendungen importiert werden.

Konkret handelt es sich dabei in erster Linie um das eigentliche Fernsehprogramm, also die Angabe darüber, welche Sendung zu welchem Zeitpunkt auf welchem Sender ausgestrahlt wird. Des Weiteren zählen Detailinformationen wie die Sendungsbeschreibungen, die an einer Sendung beteiligten Personen, die der Sendung zugeordneten Genres, die Produktionsländer, die Sendungsdauer sowie das Produktionsjahr zu den Sendungsdaten.

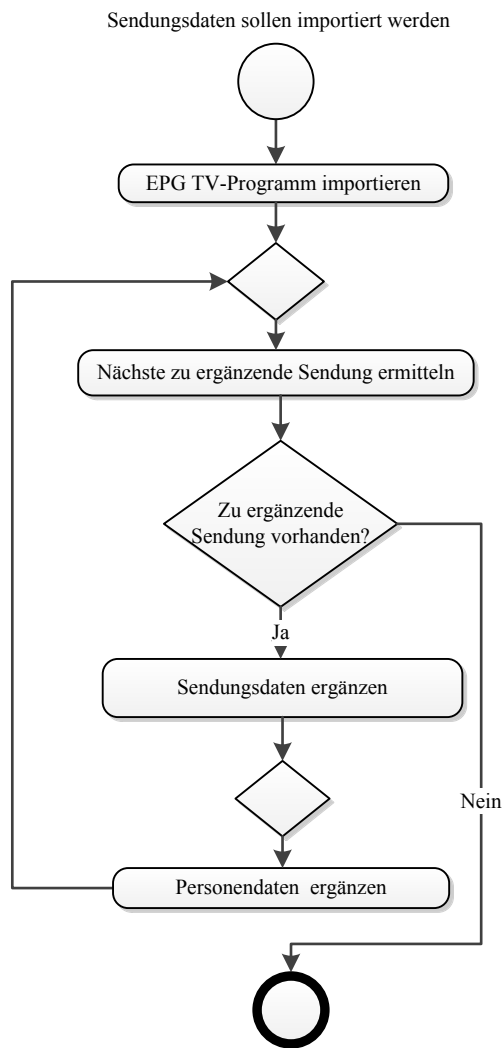


Abbildung 8: Prozess Sendungsdaten-Import durchführen

Als Ausgangspunkt werden hierfür Daten aus einem Electronic Program Guide (EPG), wie beispielsweise jene auf epgdata.com, verwendet. Da EPG-Daten häufig nur rudimentäre Informationen zu Sendungen zur Verfügung stellen, werden diese Daten mit jenen von Filmdatenbanken, wie z. B. „The Movie Database“⁵ (TMBD), ergänzt.

4.1.6 Benutzerdaten-Import

In Abbildung 9 wird der Prozess des Benutzerdaten-Imports dargestellt. Da sehr viele Benutzer bereits Informationen über Lieblingssendungen, Sendungsbewertungen und Lieblingsschauspieler in ihren Facebook-Profilen pflegen, werden diese Informationen importiert.

⁵ <https://www.themoviedb.org>

Da der Import der Benutzerdaten über das Facebook-Profil die Wartezeit des Benutzers nach dem Login erhöht, kann die Häufigkeit des Imports konfiguriert werden. Die Benutzerdaten werden in der Folge erst dann wieder importiert, wenn der Aktualisierungszeitpunkt erreicht ist.

Um über die importierten Lieblingssendungen und Lieblingsschauspieler zusätzliche Informationen zu erhalten, werden diese Daten mit Daten aus einer Filmdatenbank ergänzt. Anschließend werden die Bewertungen in die Datenbank übertragen. Hat ein Benutzer eine Sendung auf Facebook als gesehen markiert, wird eine implizite Bewertung von 3 angenommen. Wurde eine Sendung auf Facebook als Lieblingssendung markiert, erhält die Sendung die Bewertung 5. Wenn ein Benutzer eine Sendung in Facebook bewertet hat, wird diese Bewertung direkt übernommen, da Facebook, so wie auch WatchMe, eine Bewertungsskala von 1 bis 5 verwendet. Als Bewertungsdatum wird das Datum, an dem die Bewertung in Facebook abgegeben wurde, eingetragen. Außerdem werden die vom Benutzer auf Facebook als Lieblingsschauspieler markierten Schauspieler ebenfalls in der Datenbank vermerkt.

Benutzerdaten sollen importiert werden

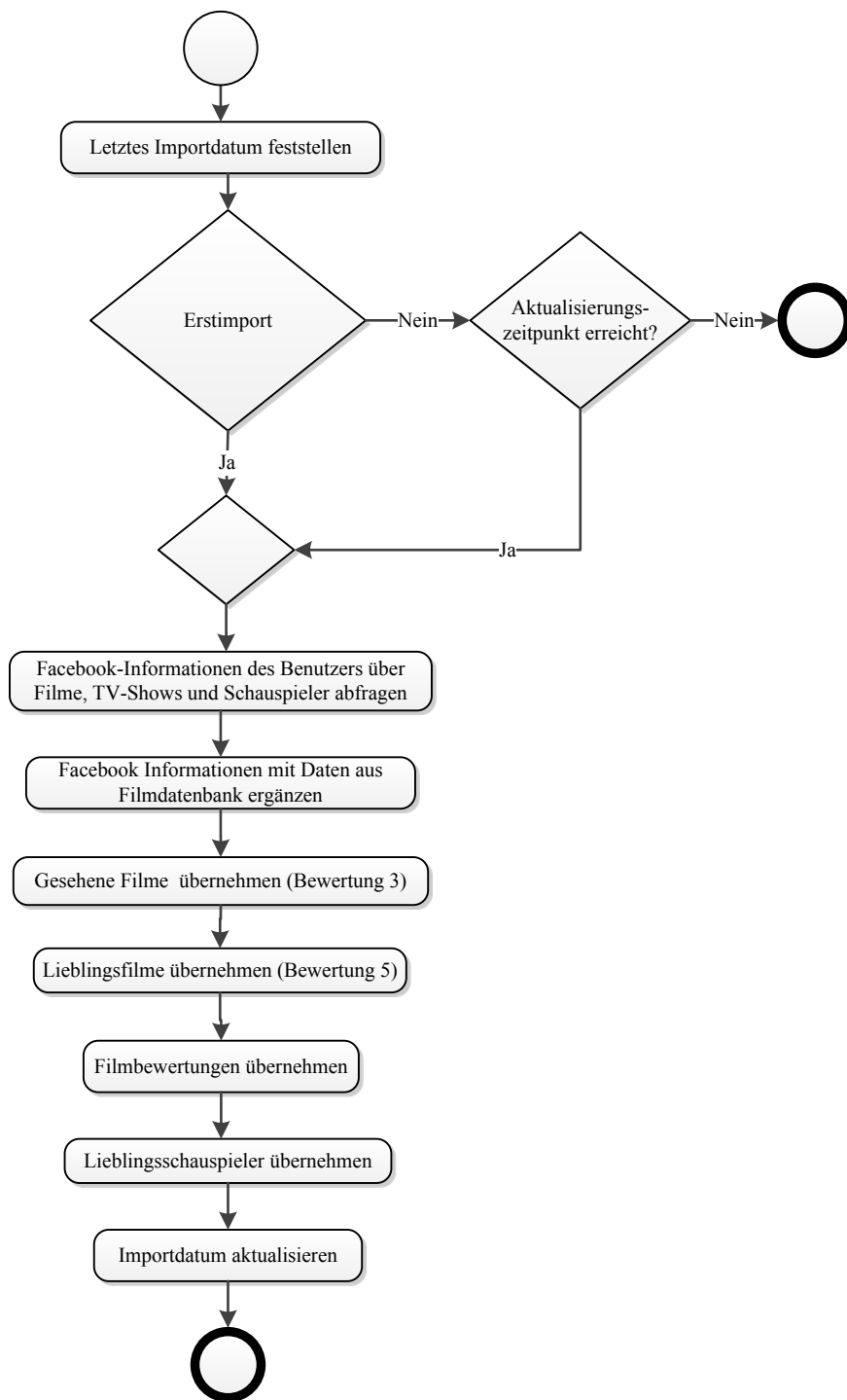


Abbildung 9: Prozess Benutzerdaten-Import durchführen

4.2 Datenstrukturen

In diesem Unterkapitel werden die für die Prozesse benötigten Daten erläutert und das zugrunde liegende Datenmodell (siehe Abbildung 10) dargestellt.

Für das Empfehlungssystem werden drei Arten von Personen unterschieden:

- Kunden (Customer)
- Benutzer (User)
- Beteiligte Person (Participant)

Eine Person kann gleichzeitig Kunde, Benutzer und beteiligte Person sein. Von jeder Person sind Vor- und Nachname, Geschlecht und Geburtsdatum bekannt.

Für Kunden ist zusätzlich noch die LIWEST Kundennummer bekannt und welche Sender (Station) der Kunde empfangen kann. Jeder Kunde kann zumindest einen Sender empfangen. Welche Sender der Kunde genau empfangen kann hängt von den bei LIWEST gebuchten Produkten ab. Die empfangbaren Sender sind für die Empfehlungen relevant, da nur Sendungen auf Sendern, die der Kunde auch empfangen kann, vorgeschlagen werden sollen. Für jeden Sender sind der Name und die Sprache, in der der Sender ausstrahlt, bekannt.

Der LIWEST Kunde repräsentiert immer einen gesamten Haushalt. Da ein Haushalt jedoch aus mehreren Haushaltsmitgliedern bestehen kann und Filmempfehlungen für die unterschiedlichen Haushaltsmitglieder verschieden ausfallen sollen, können jedem Kunden mehrere Benutzer zugeordnet werden. Die Filmempfehlungen werden für den Benutzer ermittelt, nicht für den Kunden. Die Zuordnung eines Benutzers zum LIWEST Kunden erfolgt bei der erstmaligen Authentifizierung des Benutzers (siehe auch Abschnitt 4.1.1). Der Benutzer ist immer genau einem Kunden zugeordnet.

Zu jedem Benutzer sind sein Facebook-Benutzername, das Datum der letzten Aktualisierung der Benutzerdaten, seine Lieblingsschauspieler sowie seine Einstellungen für die Empfehlungsgenerierung bekannt. Diese Einstellungen sind die Mindestanzahl an bewerteten Sendungen, um eine vollständige Empfehlungsgenerierung durchführen zu können, sowie die minimale Restlaufzeit einer Sendung, damit diese bei der Empfehlung berücksichtigt wird.

Zu den Sendungsbewertungen (Rating) sind die Bewertung (rating) und der Zeitpunkt der Bewertungsabgabe bekannt. Ein Benutzer kann beliebig viele Sendungen (Content) bewerten, jedoch kann es zu jeder Sendung von einem Benutzer nur eine Bewertung geben. Ändert ein Benutzer die Bewertung einer Sendung, so wird das Bewertungsdatum überschrieben.

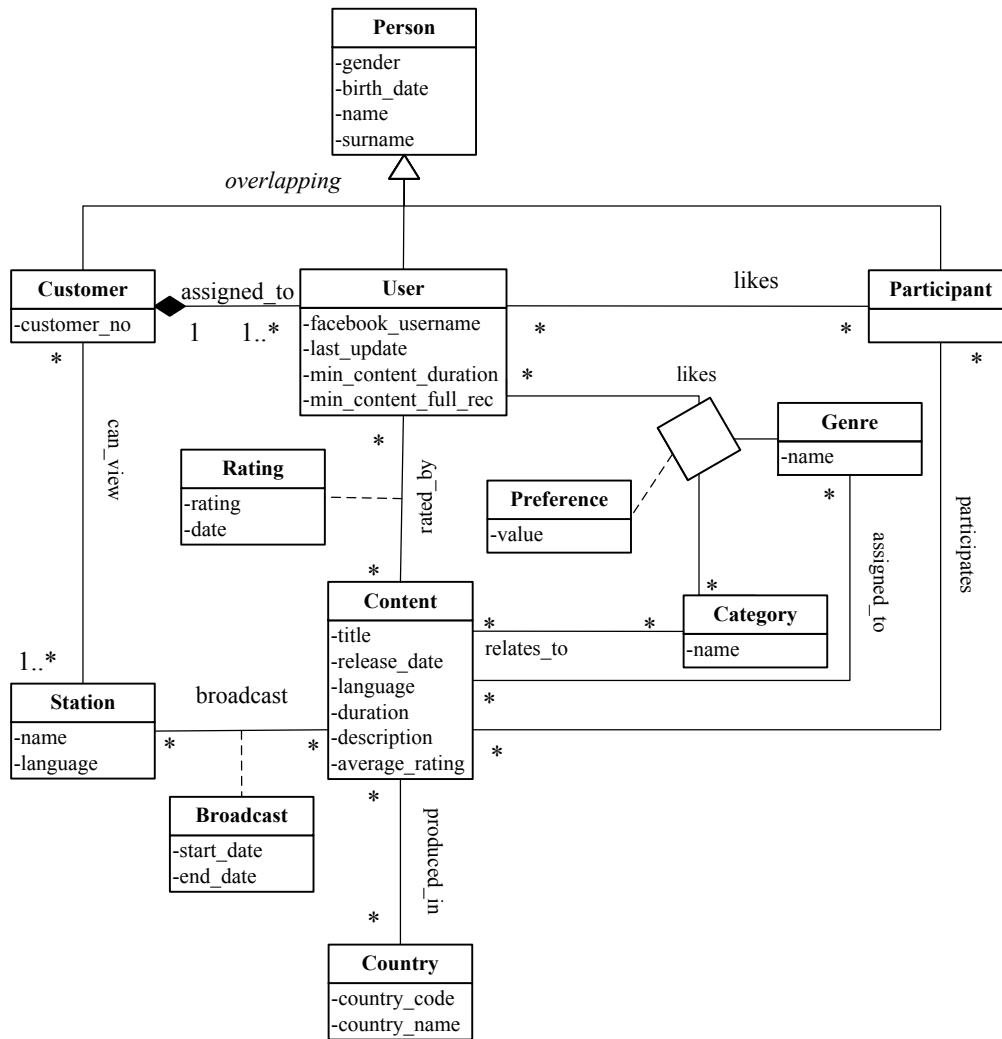


Abbildung 10: Datenmodell

Jede Sendung kann beliebig vielen Genres (Genre) und beliebig vielen Sendungskategorien (Category) zugeordnet sein. Genres unterteilen Sendungen nach ihrem Inhalt. Beispiele für Genres sind Action, Fußball, Thriller und Western. Sendungskategorien dagegen sagen etwas über die Art der Sendung aus. Es gibt folgende Sendungskategorien: Information, Musik, Sport, Entertainment, Kinder. Das gleiche Genre kann Sendungen unterschiedlicher Sendungskategorien zugeordnet sein. So ist eine Fußball Live-Übertragung zur Sendungskategorie „Sport“ und dem Genre „Fußball“ zugeordnet, wohingegen die Fußballdokumentation „Deutschland. Ein Sommermärchen“ der Sendungskategorie „Information“ und dem Genre „Fußball“ zugeordnet ist.

Zu jedem Benutzer ist die individuelle Beliebtheit der Genres je Kategorie bekannt (Preference). Ein Benutzer kann ein Genre in einer bestimmten Kategorie mögen (value = 1), nicht mögen (value = -1) oder dem Genre neutral gegenüber stehen (value = 0).

Für Sendungen sind neben den zugeordneten Genres und Sendungskategorien folgende Attribute bekannt: Titel, Erscheinungsjahr, Sprache, Dauer, Beschreibung und Produktionsländer. Außerdem sind für Sendungen auch die beteiligten Personen bekannt.

Strahlt ein Sender eine Sendung aus (Broadcast) ist der genaue Ausstrahlungszeitraum (start_date und end_date) bekannt. Diese Information ist relevant, da sich der Endzeitpunkt aufgrund von Werbeunterbrechungen nicht direkt aus der Sendungsdauer und der Startzeit ergibt.

Wie in Abbildung 11 dargestellt, können Sendungen zudem in Filme (Movie) und Serien (Series) unterteilt werden. Filme sind als Einzelstücke konzipierte Sendungen. Im Unterschied dazu sind Serien eine inhaltlich eng zusammengehörende Abfolge von Sendungen. Beispiele für Serien sind Sitcoms wie „How I met your mother“ oder Natursendungen wie „Universum“.

Eine Serie kann aus mehreren Staffeln (Season) bestehen, wobei jede Staffel aus mindestens einer Serienepisode besteht. Die Episode ist die Sendung, die dann tatsächlich ausgestrahlt wird. Zu jeder Serie ist der Serienname bekannt.

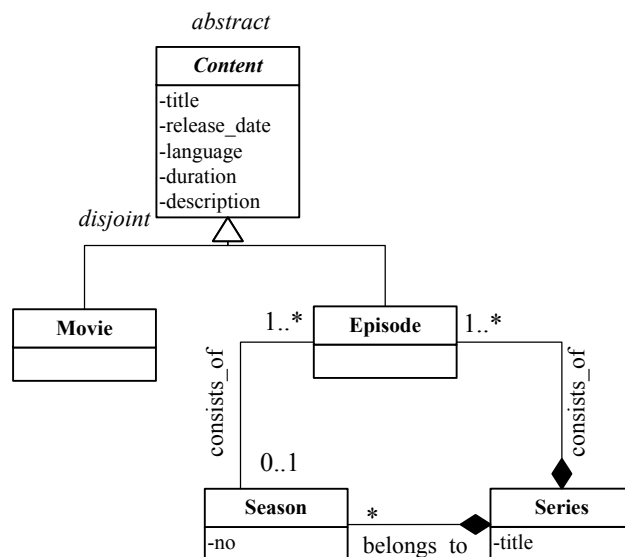


Abbildung 11: Datenmodell Film/Serie

4.3 Komponenten

Um die in Unterkapitel 4.1 erläuterten Prozesse zu realisieren, wurde die in Abbildung 12 dargestellte Architektur der Komponenten gewählt.

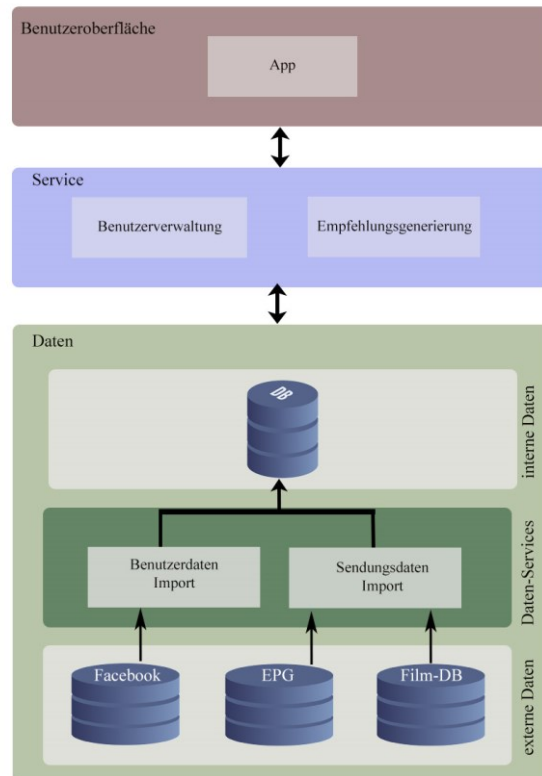


Abbildung 12: Architekturübersicht

Die Komponenten von WatchMe sind in eine 3-schichtige Architektur gegliedert. Die erste Schicht beinhaltet die Benutzeroberfläche, die zweite Schicht die Logik in Form von Services und die dritte Schicht die Daten und die benötigte Importfunktionalität.

Die Benutzeroberfläche dient dazu, dem Benutzer alle benötigten Informationen für die im Unterkapitel 4.1 beschriebenen Funktionalitäten darzustellen und ihm die Durchführung der angebotenen Funktionalitäten zu ermöglichen.

Bei den Services wird zwischen demjenigen zur Benutzerverwaltung und dem zur Empfehlungsgenerierung unterschieden. Die Services zur Benutzerverwaltung ermöglichen die Verwaltung der Benutzerdaten und der abgegebenen Sendungsbewertungen.

Das Service für die Empfehlungsgenerierung beinhaltet die gesamte Logik zur Empfehlungsgenerierung. Das Service bestimmt anhand der Benutzerdaten, ob eine Schnellempfehlung oder eine vollständige Empfehlung durchgeführt werden soll und ermittelt

die für die Empfehlungsgenerierung relevanten Sendungen. Das Service führt auch die Ähnlichkeitsberechnung zwischen Sendungen durch und ermittelt die Empfehlungen.

Die für die Empfehlungsgenerierung benötigten EPG-Daten und zusätzlichen Sendungs-
informationen werden über externe Anbieter bezogen. Um die Antwortzeiten zur Laufzeit zu
minimieren, werden diese Daten lokal redundant gespeichert, wofür der Sendungsdaten-
Import zuständig ist. Der Sendungsdaten-Import importiert die EPG-Daten, ergänzt diese mit
den Sendungsinformationen aus einer Filmdatenbank und speichert die Daten in einer
internen Datenbank.

Der Benutzerdaten-Import ist für den Import bereits bestehender Informationen über
Lieblingssendungen, Sendungsbewertungen und Schauspieler aus den Facebook-Profilen der
Benutzer zuständig.

4.4 Implementierung WatchMe

In diesem Unterkapitel wird zuerst auf grundlegende Implementierungsentscheidungen
eingegangen und danach wird die Implementierung der im Unterkapitel 4.1 erläuterten
Funktionalitäten beschrieben.

App: Um die Funktionalitäten von WatchMe spontan vor dem Fernsehgerät nutzen zu
können, dient als Benutzeroberfläche eine App. Dass Smartphones bzw. Tablets während des
TV Konsums bereits stark genutzt werden, bestätigt eine Studie von Deloitte aus dem Jahr
2015, laut der 79 % der Personen im Alter von 14 bis 18 Jahren und 62 % der Personen im
Alter von 19 bis 24 Jahren in Deutschland während des Fernsehens ihr Smartphone
verwenden (Deloitte, 2015).

Die WatchMe-App wird für die Android Plattform entwickelt (Android, 2015). Android als
Entwicklungsplattform wurde gewählt, da sich die Entwicklung mittels Java und Eclipse sehr
gut in die bestehende Software-Architektur von LIWEST einfügt und bestehendes Know-how
in der Android Entwicklung bei langjährigen Softwarepartnern vorhanden ist.

Bei der Entwicklung einer Android App muss bestimmt werden, welchen API Level die App
unterstützen soll. Der API Level bestimmt die Funktionen, die in der entsprechenden Android
Version zur Verfügung stehen. Jede Android Version unterstützt genau einen API Level. Da
API Level jedoch abwärtskompatibel sind, werden implizit auch alle vorhergehenden API
Versionen unterstützt (Android SDK, 2015).

Durch die Festlegung des API Levels wird also auf der einen Seite der mögliche Funktionsumfang für die App festgelegt, auf der anderen Seite aber auch die Reichweite der App, also die Anzahl an Benutzern, die jene Android Version installiert haben, die mit diesem API Level kompatibel ist.

Die WatchMe-App wird für den API Level 16, welcher ab Android 4.1 (Jelly Bean) verfügbar ist, entwickelt, da er auf der einen Seite eine sehr hohe Anzahl von Benutzern erreicht und auf der anderen Seite eine hohe Anzahl an Funktionalitäten zur Verfügung stellt.

Laut einem Vergleich der Android Versionen von Google aus Juni 2015 nutzen 89 % aller Android Benutzer eine Android Version, welche mit dem API Level 16 kompatibel ist (Android Platform Versions, 2015). Im Hinblick auf die unterstützten Funktionalitäten sind die beiden folgenden Faktoren wesentlich:

- 1) Ab dem API Level 14 werden die Android-Versionen für Mobiltelefone und Tablets zusammengefasst und somit kann eine App auf beiden Gerätetypen verwendet werden (Android API 4.0, 2015).
- 2) Ab API Level 16 (Android API 4.1, 2015) steht „Project Butter“ zur Verfügung, wodurch eine flüssigere Bedienung von Apps möglich ist.

Datenbank: Da die dem Empfehlungssystem WatchMe zugrunde liegenden Daten strukturiert sind, werden diese in einer relationalen Datenbank abgelegt. Die Tatsache, dass bei LIWEST bereits Oracle Datenbanken im Einsatz sind, hat dazu geführt, als Datenbanksystem Oracle in der Version 12c auszuwählen.

Java: Die Entscheidung, Java als Programmiersprache für den Sendungsdaten-Import und für die Web-Services zu wählen, ergibt sich aus der Entscheidung, dass für die Web-Services ein Servlet Container eingesetzt wird. Für diesen Zweck wurde Apache Tomcat⁶ gewählt, da sich bei LIWEST bereits mehrere Tomcat Instanzen in Betrieb befinden.

4.4.1 Implementierung der Authentifizierung

Die Authentifizierung des Benutzers erfolgt durch den Facebook-Account des Benutzers. Die Login-Möglichkeit mittels Facebook-Account wird inzwischen von vielen Webseiten und Apps angeboten und ist daher bei Kunden bekannt und wird von ihnen akzeptiert.

⁶ <http://tomcat.apache.org/>

Damit sich der Benutzer mit seinen Facebook-Logindaten an der App authentifizieren kann, wird das direkt von Facebook zur Verfügung gestellte Facebook-SDK für Android⁷ verwendet. Die Korrektheit der Logindaten wird direkt von Facebook überprüft, die WatchMe-App hat keinerlei Kenntnisse über die Logindaten.

Wurde der Benutzer noch keinem LIWEST Kunden zugeordnet, wird der Kunde aufgefordert, seine Kundennummer und sein Kundenpasswort einzugeben. Dadurch wird die Benutzerzuordnung durchgeführt. Für die prototypische Implementierung von WatchMe im Rahmen dieser Diplomarbeit wurde für die Benutzerzuordnung keine GUI entwickelt.

4.4.2 Implementierung der Sendungsbewertung

Technisch gesehen ist die Sendungsbewertung eine relativ einfache Funktionalität, da es nur darum geht, per Web-Service Sendungsdaten abzufragen, Bewertungen anzuzeigen und die geänderten Werte wieder über das entsprechende Web-Service in die Datenbank zu speichern.

Um nicht zu viele Daten austauschen zu müssen, werden zuerst nur die bewerteten Sendungen des aktuellen Monats angezeigt. Dieser Zeitraum kann in der App auf das aktuelle Jahr bzw. auf die Anzeige aller Bewertungen ausgedehnt werden.

Da die durchschnittliche Sendungsbewertung für den Vergleich der Sendungen herangezogen wird, wird bei jeder Bewertung einer Sendung der Durchschnittswert aller Bewertungen dieser Sendung ermittelt und bei der Sendung in der Datenbank hinterlegt. Somit muss der Wert nicht bei jedem Sendungsvergleich neu ermittelt werden.

4.4.3 Implementierung der Benutzerverwaltung

Bei der Benutzerverwaltung werden die Werte der benutzerspezifischen Einstellungen per Web-Service aufgerufen, angezeigt und die geänderten Werte wieder über ein Web-Service in die Datenbank geschrieben. Folgende benutzerspezifischen Einstellungen können geändert werden:

Mindestanzahl der Bewertungen: Dieser Wert gibt die minimale Anzahl an Benutzerbewertungen an, die vorliegen müssen, um eine vollständige Empfehlungsgenerierung durchzuführen. Dafür wird zumindest eine Bewertung benötigt, weswegen nur Werte ≥ 1 eingetragen werden können. Standardmäßig ist die Mindestanzahl der Bewertungen auf 30 gesetzt.

⁷ <https://developers.facebook.com/docs/android>

Minimale Restlaufzeit: Für die minimale Restlaufzeit ist der Standardwert auf 0 gesetzt. Da dieser Wert einen Prozentsatz angibt, kann er auf einen beliebigen Wert von 0 bis 100 geändert werden.

Individuelle Beliebtheit der Genres je Kategorie: Der Benutzer kann die Beliebtheit von Genres in Kategorien wie folgt numerisch angeben:

- 1: Der Benutzer mag das Genre nicht
- 0: Dem Benutzer ist das Genre egal
- 1: Der Benutzer mag das Genre

Für die prototypische Implementierung von WatchMe im Rahmen dieser Diplomarbeit wurde für die Benutzerverwaltung keine GUI entwickelt.

4.4.4 Implementierung der Empfehlungsgenerierung

Die Empfehlungsgenerierung ist als Web-Service, welchem die Benutzer-ID und der gewünschte Zeitpunkt übergeben werden, implementiert. Das Web-Service gibt bis zu drei Empfehlungen zurück.

Das Web-Service überprüft zuerst, ob die Voraussetzungen für eine vollständige Empfehlung bereits gegeben sind. Falls dies nicht der Fall ist, wird die Schnellempfehlung durchgeführt und das Ergebnis zurückgeliefert.

Für die vollständige Empfehlung werden zuerst alle potentiellen Sendungen und alle positiv bewerteten Sendungen in einer eigenen Liste abgespeichert.

Danach wird die Ähnlichkeit zwischen jeder potentiellen Sendung und jeder positiv bewerteten Sendung ermittelt.

Für die Berechnung des Pearson-Korrelationskoeffizienten (Chok, 2010), welcher für die Merkmale Genre, beteiligte Personen und Produktionsländer verwendet wird, wird eine Funktion aus der Math3-Bibliothek von Apache⁸ verwendet.

Für die Ähnlichkeitsberechnung der Sendungsbeschreibung wird die Cosinus-Ähnlichkeit verwendet, da diese durch die Berücksichtigung der Häufigkeit der Vorkommnisse der einzelnen Wörter vor allem für Textvergleiche sehr gut geeignet ist (Jannach, Zanker, Felfernig, & Friedrich, 2011).

⁸ <http://commons.apache.org/proper/commons-math/>

Bevor die Berechnung der Cosinus-Ähnlichkeit für die Sendungsbeschreibungen erfolgt, werden aus den Sendungsbeschreibungen Stoppwörter entfernt. Dazu wird eine von Herrn Damien Doyle (Doyle, 2015) zur Verfügung gestellte Stoppwortliste verwendet. Die eigentliche Berechnung der Cosinus-Ähnlichkeit erfolgt in einer selbst geschriebenen Methode. Abbildung 13 zeigt die Berechnung der Cosinus-Ähnlichkeit. Aus den Sendungsbeschreibungen wurden die Stoppwörter herausgefiltert und Häufigkeitsvektoren gebildet. Diese Häufigkeitsvektoren werden nun in die Formel für die Cosinus-Ähnlichkeit eingesetzt und diese somit berechnet.

```
for (int i = 0; i < movVector1.length; i++) {
    dotProduct += movVector1[i] * movVector2[i]; //a.b
    magnitude1 += Math.pow(movVector1[i], 2); // (a^2)
    magnitude2 += Math.pow(movVector2[i], 2); // (b^2)
}

magnitude1 = Math.sqrt(magnitude1); //sqrt (a^2)
magnitude2 = Math.sqrt(magnitude2); //sqrt (b^2)
cosineSimilarity = dotProduct / (magnitude1 * magnitude2);
```

Abbildung 13: Codebeispiel Cosinus-Ähnlichkeit

Für die Merkmale Produktionsjahr und Dauer wird die Ähnlichkeitsberechnung durch die Ermittlung der absoluten Differenz der jeweiligen Werte durchgeführt. Dazu wird die Funktion `abs()` der Klasse `java.lang.Math` verwendet.

Die Ähnlichkeitsberechnung der Lieblingsschauspieler erfolgt in einer selbstgeschriebenen Methode. Zuerst wird durch ein SQL-Statement für die zu vergleichenden Sendungen die jeweilige Anzahl der Nennungen als Lieblingsschauspieler für die drei beliebtesten Schauspieler ermittelt. Danach wird das Ergebnis durch die Berechnung der Summe der Differenzen der Anzahl der Nennungen zwischen den beliebtesten Schauspielern, den zweitbeliebtesten Schauspielern und den drittbeliebtesten Schauspielern ermittelt.

Der Vergleich der Sendungsbewertungen erfolgt über die absolute Differenz der Durchschnittsbewertungen der zu vergleichenden Sendungen. Da die Durchschnittsbewertung bei jeder Sendung hinterlegt ist (siehe auch Abschnitt 4.4.2), muss dieser Werte nicht mehr berechnet werden.

Nachdem alle Ähnlichkeitsberechnungen durchgeführt worden sind, werden die einzelnen Werte mittels Min/Max-Normalisierung, welche mittels einer selbstgeschriebenen Methode durchgeführt wird, normalisiert und das Ergebnis generiert. Aus diesem werden die drei Sendungen mit dem höchsten Empfehlungswert ermittelt und zurückgegeben.

4.4.5 Implementierung des Sendungsdaten-Imports

Der Sendungsdaten-Import erfolgt mittels eigens geschriebenem JAVA-Programm, welches periodisch automatisiert ausgeführt wird.

Um die benötigten Sendungsdaten zu importieren, werden zwei unterschiedliche Datenquellen verwendet:

- epgdata.com (EPG Data, 2015) für EPG-Daten
- themoviedatabase.org (The Movie Database, 2015) für die über das Fernsehprogramm hinausgehenden Informationen

Daten von epgdata.com: Die EPG-Daten werden von epgdata.com als XML-Dateien zur Verfügung gestellt und können mittels HTTP-Request automatisiert abgefragt werden. Für das Fernsehprogramm gibt es für jeden Kalendertag eine eigene Datei. Zusätzlich gibt es jeweils eine Stammdatendatei für Genres, Sendungskategorien und Fernsehsender. In diesen Dateien ist die Zuordnung der Namen von Genres, Sendungskategorien und Fernsehsendern zu den im Fernsehprogramm verwendeten IDs hinterlegt.

Abbildung 14 zeigt einen Ausschnitt für das Fernsehprogramm vom 29.08.2014. Für jede Sendung gibt es ein `<data>`-Element. Die Informationen zu einer Sendung und der Ausstrahlung der Sendung befinden sich in den Unterelementen. In Tabelle 12 werden die einzelnen Unterelemente des `<data>`-Elements beschrieben.

```

<?xml version="1.0" encoding="utf-8" standalone="no"?>
<pack>
  <data>
    <d2>56</d2>
    <d4>2014-08-29 05:05:00</d4>
    <d5>2014-08-29 06:20:00</d5>
    <d10>400</d10>
    <d19>Stöckl.</d19>
    <d21>Einmal wöchentlich lädt Barbara Stöckl bis zu fünf
    prominente Persönlichkeiten und Menschen mit interessanter
    Lebensgeschichte in ihre Talkshow "STÖCKL.". Die Themen geben
    die Gäste selbst vor. Sie sprechen über ihr Leben, ihre
    Erfahrungen, Haltungen, Zukunftsvisionen und über
    gesellschaftlich relevante Themen.</d21>
    <d25>452</d25>
    <d32>A</d32>
    <d33>2014</d33>
    <d34>Barbara Stöckl</d34>
    <d35>Hans Sigl - Sven Hannawald - Thomas Glavinic</d35>
    <d36></d36>
    <d37></d37>
  </data>
  <data>
    .
    .
    .
  </data>
  .
  .
  .
</pack>

```

Abbildung 14: XML-Beispiel Fernsehprogramm

Element	Beschreibung
<d2>	ID des Fernsehsender
<d4>	Startzeitpunkt der Ausstrahlung
<d5>	Endzeitpunkt der Ausstrahlung
<d10>	ID der Sendungskategorie
<d19>	Titel der Sendung
<d21>	Sendungsbeschreibung
<d25>	ID des Genres
<d32>	Produktionsland
<d33>	Produktionsjahr
<d34>	Moderator
<d35>	Studiogäste
<d36>	Regisseur
<d37>	Schauspieler

Tabelle 12: Elementbeschreibung Fernsehprogramm

Die IDs der Fernsehsender, der Sendungskategorien und der Genres werden beim Import durch einen Abgleich mit den jeweiligen Stammdaten aufgelöst. In dem in Abbildung 14

gezeigten Beispiel handelt es sich um den Fernsehsender „3SAT“, die Sendungskategorie „Entertainment“ und das Genre „Talkshow“.

In den folgenden Elementen können mehrere Werte eingetragen sein: ID der Sendungskategorie, ID des Genres, Produktionsland, Moderator, Studiogäste, Regisseur, Schauspieler. Falls mehrere Werte eingetragen sind, sind diese durch ein Trennzeichen getrennt und werden beim Import aufgeteilt.

Die XML-Dateien werden unter Verwendung der Jackson-API⁹ in Java Objekte umgewandelt und anschließend mit Daten der The Movie Database (TMDB) ergänzt.

Daten von themoviedatabase.org: Da in den EPG-Daten nur rudimentäre Sendungsinformationen enthalten sind, werden diese durch Daten der TMDB ergänzt. Die Daten werden über eine kostenlose REST-Schnittstelle, welche das Ergebnis im JSON Format zurückliefert, zur Verfügung gestellt. Für die Implementierung von WatchMe wird ein Java-Wrapper von Herrn Holger Brandl (Holger Brandl API, 2015) verwendet, um die REST-Schnittstelle von TMDB aufzurufen, da der Wrapper sowohl den Aufruf der Schnittstelle erleichtert als auch die Ergebnisse direkt in Java Objekten returniert. Durch den Abgleich mit TMDB sollen die Produktionsländer, die Genres und die beteiligten Personen der EPG-Daten ergänzt werden.

Um die EPG-Daten einer Sendung ergänzen zu können, wird diese anhand des Sendungstitels und des Produktionsjahres in der TMDB gesucht. Andere Parameter für die Suche können nicht übergeben werden. Als Ergebnis werden die gefundenen Sendungen mit allen Sendungsdaten zurückgegeben. Sollte die Zuordnung nicht eindeutig sein, werden die gefundenen Sendungen noch über den Namen des Regisseurs eingegrenzt. Bei den im Rahmen der Implementierung testweise importieren Daten konnten dadurch Falschzuordnungen gänzlich verhindert werden.

Da sich die Genrenamen zwischen den EPG-Daten und den Daten der TMDB unterscheiden, wird eine Zuordnung zwischen den unterschiedlichen Genres geschaffen. So werden z. B. die Genres „Adventure“ und „Abenteuer“ beide dem Genre „Abenteuer“ zugeordnet. Außerdem werden durch diese Zuordnung auch zusammengesetzte Genres aufgeteilt. So wird das Genre „Action & Adventure“ den Genres „Action“ und „Abenteuer“ zugeordnet.

⁹ <https://github.com/FasterXML/jackson-dataformat-xml>

4.4.6 Implementierung des Benutzerdaten-Imports

Für den Zugriff auf Informationen des Facebook-Profiles des Benutzers wird das Facebook-SDK für Android¹⁰ verwendet. Bei jeder Abfrage auf Informationen des Facebook-Profiles durch die WatchMe-App wird vom Facebook-SDK überprüft, ob die App die benötigten Berechtigungen hat. Falls die Berechtigungen noch nicht erteilt wurden, wird der Benutzer durch das Facebook-SDK aufgefordert, die Berechtigungen zu erteilen.

Informationen aus dem Facebook-Profil können jedoch nur abgefragt werden, nachdem sich der Benutzer an der App mit seinen Facebook-Logindaten authentifiziert hat. Eine Abfrage im Hintergrund, ohne vorheriger Authentifikation des Benutzers mit seinen Facebook-Logindaten, ist also nicht möglich. Die Berechtigungen gelten so lange, bis der Benutzer der App die Berechtigungen über seine Facebook-Profileinstellungen wieder entzieht. Erteilt der Benutzer die Berechtigungen nicht, wird vom Facebook-SDK eine Fehlermeldung an die App zurückgegeben.

Bei jeder Datenabfrage muss über den Parameter „fields“ angegeben werden, welche Daten abgefragt werden. Tabelle 13 zeigt die Daten, die aus dem Facebook-Profil abgefragt werden, mit dem jeweiligen Wert für den Parameter „fields“. Die Lieblingssendungen und die Lieblingsschauspieler werden beide über den Parameterwert „likes“ abgeglichen. Ob es sich bei dem zurückgegebenen Datensatz um Schauspieler oder um Sendungen handelt, wird über die zurückgegebene Kategorie ermittelt. Für gesehene Sendungen wird der Wert „video.watches“ und für bewertete Sendungen der Wert „video.ratings“ übergeben.

Abgefragte Daten	Parameterwert
Gesehene Sendungen	video.watches
Lieblingssendungen	likes
Bewertete Sendungen	video.ratings
Lieblingsschauspieler	likes

Tabelle 13: Benutzerdaten-Import Parameter

Abbildung 15 zeigt einen Beispielcode zum Abfragen von bewerteten Sendungen.

```
Bundle bundle = new Bundle();
bundle.putString("fields", "video.rates");
new Request(s, "me", bundle, HttpMethod.GET, new Request.Callback()
{
    ...
    ...
})
```

Abbildung 15: Beispielcode Abfrage bewertete Sendungen

¹⁰ <https://developers.facebook.com/docs/android>

Abbildung 16 zeigt die von Facebook zurückgegebenen Daten und Tabelle 14 beschreibt die zurückgegebenen Attribute. In dem gezeigten Beispiel wird also die Sendung „22 Bullets“ am 03.03.2015 um 12:36 auf einer 5-teiligen Bewertungsskala mit dem Wert 4 bewertet.

```
{
  "start_time": "2015-03-03T15:12:36+0000",
  "data": {
    "rating": {
      "value": 4,
      "scale": 5
    },
    "movie": {
      "id": "202788869741027",
      "title": "22 Bullets",
    }
  }
}
```

Abbildung 16: Facebook Sendungsbewertung

Attribute	Beschreibung
start_time	Zeitpunkt der Bewertung
rating.value	Höhe der Bewertung
rating.scale	Bewertungsskala
movie.title	Titel der Sendung
movie.id	Facebook-ID der Sendung

Tabelle 14: Beschreibung Elemente Sendungsbewertung

Nach der Abfrage der jeweiligen Daten vom Facebook-Profil des Benutzers werden die Daten über die TMDb ergänzt und danach in der Datenbank gespeichert. Für gesehene Sendungen wird eine Bewertung von 3 gespeichert, für Lieblingssendungen eine Bewertung von 5 und bei bewerteten Sendungen wird die tatsächliche Bewertung übernommen. Lieblingsschauspieler werden in der Datenbank als Lieblingsschauspieler des Benutzers vermerkt.

5 Anwendungsbeispiel

In diesem Kapitel wird anhand eines konkreten Beispiels die Funktionsweise von WatchMe aus Sicht des Benutzers dargestellt.

In dem Beispiel wird für einen bestehenden Benutzer („Benutzer A“) eine Empfehlung für einen bestimmten Zeitpunkt generiert. Folgende Benutzereinstellungen sind bei „Benutzer A“ hinterlegt:

Mindestanzahl der Bewertungen: Die Mindestanzahl der Bewertungen, die für eine vollständige Empfehlungsgenerierung benötigt werden, ist auf 5 gesetzt.

Minimale Restlaufzeit: Die minimale Restlaufzeit, damit eine Sendung empfohlen werden kann, ist auf 20 % gesetzt.

Individuelle Beliebtheit der Genres je Kategorie: Wie in Tabelle 15 dargestellt sind bei „Benutzer A“ folgende Genrepräferenzen hinterlegt: „Benutzer A“ mag bei Entertainmentsendungen die Genres Action, Drama und Thriller, das Genre Horror mag er nicht. Bei Informationssendungen mag er die Genres Nachrichten und Natur nicht und bei Sportsendungen mag er die Genres Fußball und Tennis nicht.

Sendungskategorie	Genre	Beliebtheit
Entertainment	Action	1
Entertainment	Drama	1
Entertainment	Horror	-1
Entertainment	Thriller	1
Information	Nachrichten	-1
Information	Natur	-1
Sport	Fußball	-1
Sport	Tennis	-1

Tabelle 15: Individuelle Beliebtheit der Genres

Zusätzlich zu den Benutzereinstellungen sind die in Tabelle 16 angegebenen Bewertungen von „Benutzer A“ in der Datenbank hinterlegt.

Sendung	Bewertung
Argo	5
Ocean's Eleven	5
Face Off	5
How I met your Mother	5
Hawaii 5-0	5
Dracula	4
22 Bullets	4
Die gefürchteten 4	2
Super Hero Squad	1

Tabelle 16: Sendungsbewertungen

5.1 Authentifizierung

Beim Start der App wird „Benutzer A“ aufgefordert, sich mit seinen Facebook-Logindaten zu authentifizieren (siehe Abbildung 17 und Abbildung 18). Eine alternative Login-Möglichkeit wird nicht angeboten, da die Profildaten des Facebook-Profiles für die Empfehlungsgenerierung genutzt werden (siehe Abschnitt 4.1.6). „Benutzer A“ gibt seine Logindaten ein und authentifiziert sich mit einem Klick auf den Button „Log In“.

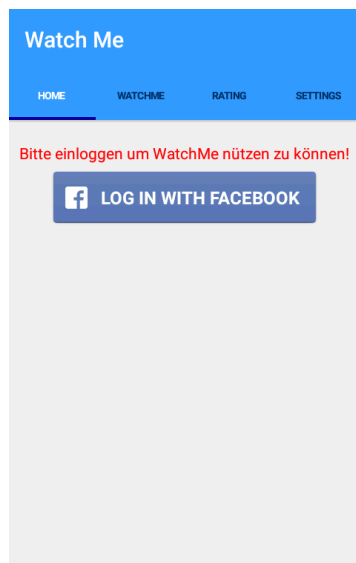


Abbildung 17: WatchMe Startbildschirm

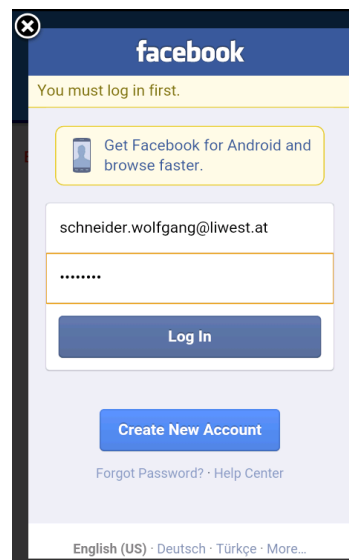


Abbildung 18: WatchMe Facebook-Login

5.2 Empfehlungsgenerierung

Nachdem sich „Benutzer A“ authentifiziert hat, gibt er im Tab „WATCHME“ den Empfehlungszeitpunkt an (siehe Abbildung 19 - Abbildung 21).

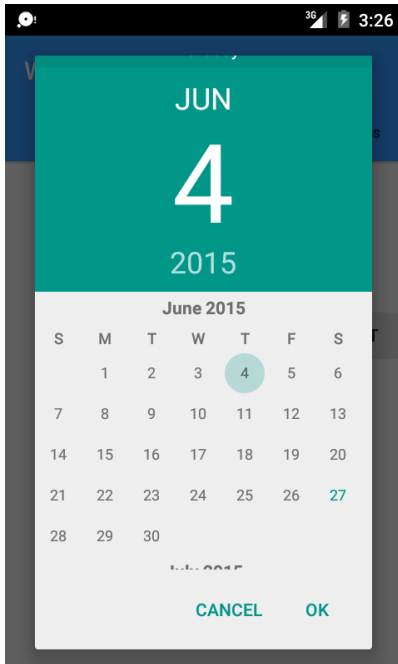


Abbildung 19: WatchMe Datumsauswahl

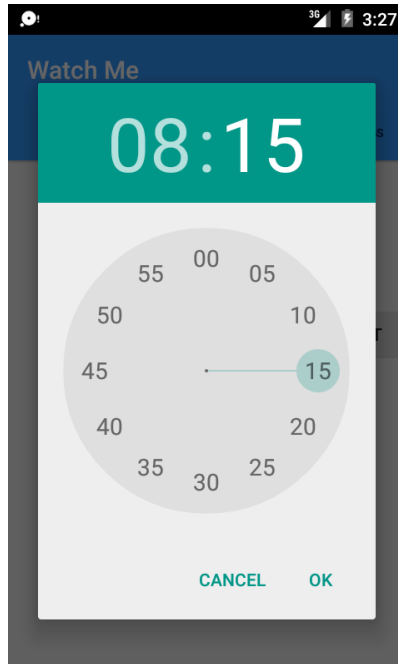


Abbildung 20: WatchMe Uhrzeitauswahl

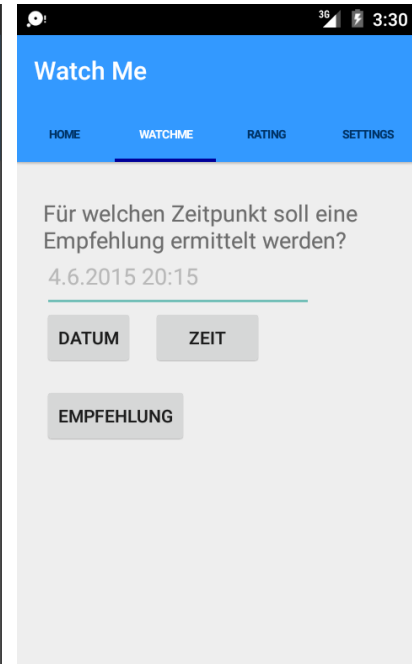


Abbildung 21: WatchMe Empfehlungsgenerierung

Danach startet „Benutzer A“ durch einen Klick auf den Button „EMPFEHLUNG“ die Empfehlungsgenerierung. Da „Benutzer A“ bereits mehr als die angegebene Mindestanzahl von Sendungen bewertet hat, wird eine vollständige Empfehlungsgenerierung durchgeführt.

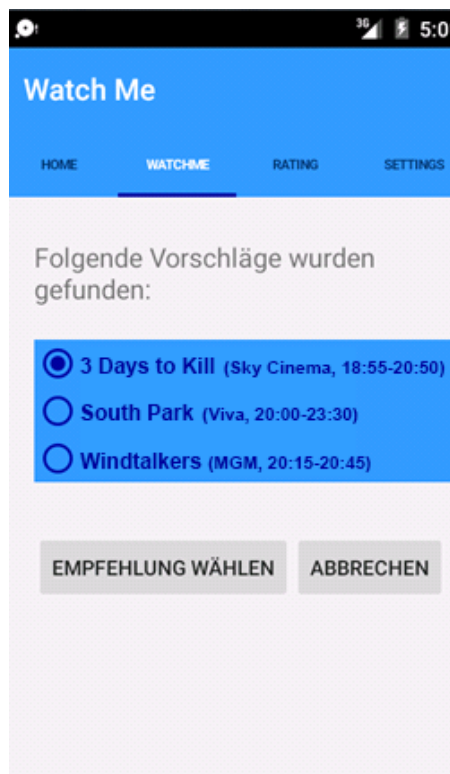


Abbildung 22: Empfehlungsauswahl

„Benutzer A“ werden die in Abbildung 22 dargestellten Sendungen „3 Days to Kill“, „South Park“ und „Windtalkers“ empfohlen. „Benutzer A“ wählt seine favorisierte Sendung „3 Days to Kill“ und bestätigt die Auswahl durch einen Klick auf den Button „EMPFEHLUNG WÄHLEN“. Durch die Auswahl der favorisierten Sendung wird eine implizite Bewertung gespeichert, wodurch sich die Qualität der zukünftigen Empfehlungen verbessert.

Um die Tauglichkeit der Empfehlungsgenerierung in WatchMe zu illustrieren, wird die Empfehlung der Sendungen „3 Days to Kill“, „South Park“ und „Windtalkers“ genauer analysiert.

Die drei vorgeschlagenen Sendungen wurden aus den neun in Tabelle 17 gezeigten Sendungen, die zur Empfehlungsgenerierung zur Verfügung standen, ausgewählt. Es ist anhand der in Tabelle 17 grün hinterlegten Ähnlichkeitswerte deutlich erkennbar, dass die drei vorgeschlagenen Sendungen starke Ähnlichkeit zu den vom Benutzer positiv bewerteten Sendungen „22 Bullets“, „How I met your mother“ und „Windtalkers“ aufweisen. Die in Tabelle 17 gelb hinterlegten Ähnlichkeitswerte zeigen, dass die drei empfohlenen Sendungen nur geringe Ähnlichkeiten zu den vom Benutzer negativ bewerteten Sendungen „Die gefürchteten 4“ und „Super Hero Squad“ aufweisen. Dies verdeutlicht, dass WatchMe tatsächlich solche Sendungen empfiehlt, die dem Geschmack des Benutzers entsprechen.

Bewertete Sendungen		Zur Auswahl stehende Sendungen								
Sendung	Bewertung	Die Unglaublichen	Pfarrer Braun	4 Frauen & die Liebe	CSI Miami	South Park	3 Days to Kill	Windtalkers	Pat Garrett & Billy the Kid	Phileas & Ferb
Dracula	4	47,77 %	48,25 %	46,54 %	46,74 %	58,16 %	49,37 %	58,58 %	46,31 %	45,07 %
Argo	5	49,81 %	50,22 %	59,11 %	49,91 %	62,34 %	58,80 %	61,29 %	45,49 %	49,25 %
Ocean's Eleven	5	37,85 %	44,23 %	36,83 %	46,84 %	48,46 %	46,95 %	58,46 %	34,35 %	35,36 %
Hawaii 5-0	5	55,88 %	55,99 %	62,27 %	57,12 %	63,07 %	56,87 %	73,17 %	46,54 %	49,86 %
Face Off	5	45,31 %	50,42 %	43,91 %	50,12 %	55,51 %	56,41 %	77,64 %	42,43 %	51,34 %
How I met your mother	5	51,26 %	52,57 %	51,79 %	50,07 %	78,32 %	49,55 %	58,00 %	48,85 %	50,10 %
22 Bullets	4	49,97 %	57,36 %	50,47 %	57,06 %	49,64 %	85,25 %	48,24 %	45,74 %	48,99 %
Die gefürchteten 4	2	41,09 %	51,16 %	41,29 %	44,69 %	31,24 %	31,78 %	53,21 %	74,84 %	43,59 %
Super Hero Squad	1	57,84 %	44,81 %	37,12 %	62,30 %	50,12 %	49,13 %	47,15 %	63,42 %	74,58 %

Tabelle 17: Vergleich Empfehlungswerte

Für die Tauglichkeit der Empfehlungsgenerierung in WatchMe spricht des Weiteren der Umstand, dass Sendungen, die sehr ähnlich zu jenen sind, die der Benutzer negativ bewertet hat, nicht von WatchMe empfohlen werden. Im Beispiel zeigt sich dies dadurch, dass die beiden Sendungen „Pat Garrett & Billy the Kid“ und „Phileas & Ferb“, die entsprechend der grau hinterlegten Ähnlichkeitswerte in Tabelle 17 starke Ähnlichkeit zu den vom Benutzer negativ bewerteten Sendungen aufweisen, auch tatsächlich nicht von WatchMe vorgeschlagen wurden.

6 Zusammenfassung

Im Rahmen dieser Diplomarbeit wurde das Empfehlungssystem WatchMe für den Kabelnetzbetreiber LIWEST konzipiert und prototypisch implementiert.

Besondere Herausforderungen dabei waren, die passenden Merkmale für die Empfehlungsgenerierung zu identifizieren und für die jeweiligen Merkmale geeignete Ähnlichkeitsmaße zu finden. Zu diesem Zweck wurden mehrere Ansätze für Filmempfehlungssysteme und die darin verwendeten Merkmale und Ähnlichkeitsmaße analysiert.

Das Empfehlungssystem WatchMe wurde als 3-Schichten-Softwaresystem konzipiert, wobei die Benutzerschnittstelle als Android App und die Geschäftslogik als Web-Service realisiert wurde. Die Datenhaltung erfolgt in einer Oracle Datenbank.

Im Folgenden werden einige Ansatzpunkte für mögliche Weiterentwicklungen des Systems diskutiert:

Erstellung von Gruppenprofilen: Durch die Erstellung von Gruppenprofilen soll die Qualität von Empfehlungen in Situationen, in denen mehrere Personen gemeinsam fernsehen, verbessert werden.

Gewichtung der Merkmale: Da unterschiedlichen Benutzern die einzelnen Merkmale, aufgrund derer Empfehlungen generiert werden, unterschiedlich wichtig sind, soll die Gewichtung der einzelnen Merkmale zukünftig benutzerabhängig sein. Dazu kann der Benutzer entweder die gewünschte Gewichtung der einzelnen Merkmale über die WatchMe App einstellen oder die Gewichtung wird automatisch aufgrund der bisher bewerteten Sendungen ermittelt.

Erweiterung des Empfehlungsumfangs: Um durch WatchMe auch die Verkäufe von Pay-TV zu fördern, könnten zukünftig in einem eigenen Bereich Sendungen vorgeschlagen werden, die der Kunde nicht empfangen kann, da er den Sender nicht gebucht hat. Dabei könnte dem Kunden die Möglichkeit geboten werden, den entsprechenden Sender spontan zu bestellen, um sich die Sendung ansehen zu können.

WatchMe ist bereits jetzt imstande, Benutzern Sendungen, die ihren Vorlieben tatsächlich entsprechen, zu empfehlen.

Aus diesem Grund hat LIWEST im September 2015 die Absicht bekundet, das im Rahmen dieser Diplomarbeit prototypisch implementierte Empfehlungssystem WatchMe zur Produktreife zu bringen. Dafür sind folgende Schritte notwendig:

- Fertigstellung der GUI und Anpassung an das Corporate Design
- Durchführung von Maßnahmen zur Performanceoptimierung
- Durchführung umfangreicher Tests
- Überprüfung rechtlicher Rahmenbedingungen
- Beantragung der Freigabe der WatchMe-App durch Google und Facebook

Literaturverzeichnis

- Android*. (2015). Abgerufen am 10. 08 2015 von Android: <https://www.android.com/>
- Android API 4.0*. (2015). Abgerufen am 02. 08 2015 von <http://developer.android.com/about/versions/android-4.0.html>
- Android API 4.1*. (2015). Abgerufen am 02. 08 2015 von Android API 4.1: <http://developer.android.com/about/versions/android-4.1.html>
- Android Platform Versions*. (2015). Abgerufen am 02. 08 2015 von <http://developer.android.com/about/dashboards/index.html>
- Android SDK*. (2015). Abgerufen am 02. 08 2015 von <http://developer.android.com/guide/topics/manifest/uses-sdk-element.html#ApiLevels>
- Burke, R. (11 2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction* (Volume 12 Issue 4), S. 331-370.
- Chok, N. S. (2010). *Pearson's versus Spearman's and Kendall's correlation*. University of Pittsburgh, Pittsburgh.
- Cleve, J., & Lämmel, U. (2014). *Data Mining*. Wiesmar: Oldenburger Wissenschaftsverlag.
- Contentwise*. (2015). Abgerufen am 25. 04 2015 von <http://www.contentwise.tv/cw/>
- Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.
- Deloitte. (2015). *Deloitte Media Consumer Survey*. Abgerufen am 15. 07 2015 von <http://www2.deloitte.com/de/de/pages/technology-media-and-telecommunications/articles/media-consumer-survey-2015.html>
- Doyle, D. (01. 08 2015). Von Ranks NL Webmaster Tools: <http://www.ranks.nl/stopwords/german> abgerufen
- EPG Data*. (2015). Abgerufen am 05. 07 2015 von <http://www.epgdata.com/>
- Filmaster*. (2015). Abgerufen am 30. 04 2015 von <http://filmaster.tv/>
- Good, N., Schafer, J., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., et al. (1999). *Combining Collaborative-Filtering with Personal Agents for Better*

Recommendations. University of Minnesota, Department of Computer Science and Engineering.

Gupta, A., Jain, R., & Song, S. (2008). *Movie Recommendations Using Social Networks*. Stanford University.

Holger Brandl API. (2015). Abgerufen am 10. 06 2015 von <https://github.com/holgerbrandl/themoviedbapi/>

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2011). *Recommender Systems – An Introduction*. Cambridge University Press, USA.

Jinni. (2015). Abgerufen am 17. 04 2015 von <http://www.jinni.com>

Klahold, A. (2009). *Empfehlungssysteme: Recommender Systems – Grundlagen, Konzepte und Lösungen* (1. Ausg.). Wiesbaden: Vieweg+Teubner.

MovieLens. (2015). Abgerufen am 01. 06 2015 von MovieLens: <https://movielens.org/>

Netflix Techblog. (2015). Abgerufen am 31. 07 2015 von <http://techblog.netflix.com/2012/04/netflix-recommendations-beyond-5-stars.html>

RedBee Media. (2015). Abgerufen am 20. 04 2015 von <http://www.ericsson.com/broadbandmedia/>

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (2011). *Recommender Systems Handbook*. Springer Science+Business Media.

Rotten Tomatoes. (2015). Abgerufen am 15. 04 2015 von <http://www.rottentomatoes.com>

Soares, M., & Viana, P. (01. 11 2014). TV Recommendation and Personalization Systems: Integrating Broadcast and Video On-demand Services. *Advances in Electrical and Computer Engineering* (Vol. 14).

Tastekid. (2015). Abgerufen am 17. 04 2015 von <https://www.tastekid.com>

The Filter. (2015). Abgerufen am 20. 04 2015 von <http://thefilter.com>

The Movie Database. (2015). Abgerufen am 15. 07 2015 von <https://www.themoviedb.org>

What to rent. (2015). Abgerufen am 10. 08 2015 von What to rent:

<http://www.whattorent.com/>

Xia, T., & Chai, Y. (03 2011). An Improvement to TF-IDF: Term Distribution. *Journal of Software* (6), S. 413-420.

Yu, Z., Zhou, X., Hao, Y., & Gu, J. (2006). *TV program recommendation for multiple viewers based on user profile merging.*

Abbildungsverzeichnis

Abbildung 1: Berechnung Schnellempfehlung	28
Abbildung 2: Ähnlichkeitsermittlung Genre	29
Abbildung 3: Berechnung prozentueller Erreichung der maximal erreichbaren Werte	31
Abbildung 4: Prozess Authentifizierung durchführen	35
Abbildung 5: Prozess Sendungsbewertung durchführen	36
Abbildung 6: Prozess Benutzerverwaltung durchführen	37
Abbildung 7: Prozess Empfehlungsgenerierung durchführen	38
Abbildung 8: Prozess Sendungsdaten-Import durchführen	40
Abbildung 9: Prozess Benutzerdaten-Import durchführen	42
Abbildung 10: Datenmodell	44
Abbildung 11: Datenmodell Film/Serie	45
Abbildung 12: Architekturübersicht	46
Abbildung 13: Codebeispiel Cosinus-Ähnlichkeit	51
Abbildung 14: XML-Beispiel Fernsehprogramm	53
Abbildung 15: Beispielcode Abfrage bewertete Sendungen	55
Abbildung 16: Facebook Sendungsbewertung	56
Abbildung 17: WatchMe Startbildschirm	58
Abbildung 18: WatchMe Facebook-Login	58
Abbildung 19: WatchMe Datumsauswahl	59
Abbildung 20: WatchMe Uhrzeitauswahl	59
Abbildung 21: WatchMe Empfehlungsgenerierung	59
Abbildung 22: Empfehlungsauswahl	59
Abbildung 23: Tablespace erstellen	69
Abbildung 24: DB-Benutzer anlegen	69
Abbildung 25: Sendungsdaten-Import Properties Datei	70
Abbildung 26: Web-Services config.properties	70

Tabellenverzeichnis

Tabelle 1: Übersicht Empfehlungssysteme	12
Tabelle 2: Übersicht Empfehlungsmethoden	19
Tabelle 3: Merkmalsübersicht Sendungsdaten	20
Tabelle 4: Merkmalsübersicht Benutzerpräferenzen	21
Tabelle 5: Merkmalsübersicht Benutzerinformationen	21
Tabelle 6: Ähnlichkeitsberechnung	22
Tabelle 7: Vergleich beliebteste Schauspieler	27
Tabelle 8: Normalisierte Ähnlichkeitswerte vom Vergleich der Genres	30
Tabelle 9: Gewichtete Werte aller Merkmale	31
Tabelle 10: Übersicht Ähnlichkeiten von Sendungen	32
Tabelle 11: WatchMe Funktionalitäten	34
Tabelle 12: Elementbeschreibung Fernsehprogramm	53
Tabelle 13: Benutzerdaten-Import Parameter	55
Tabelle 14: Beschreibung Elemente Sendungsbewertung	56
Tabelle 15: Individuelle Beliebtheit der Genres	57
Tabelle 16: Sendungsbewertungen	58
Tabelle 17: Vergleich Empfehlungswerte	60

Anhang

A. Installationsanleitung

Im folgenden Abschnitt werden die Installationsschritte für den WatchMe Prototypen beschrieben. Dabei wird die Installation für die Datenbank, den Sendungsdaten-Import, die Web-Services und die Android App beschrieben. Alle für die Installation benötigten Installationsdateien und -skripte sowie der gesamte Quellcode liegen der Diplomarbeit auf CD bei.

Installation Datenbank

Für das Einrichten der Datenbank wird eine installierte Oracle Datenbank in der Version 12c vorausgesetzt.

Zuerst muss der Tablespace für die Datenfiles angelegt werden:

```
create tablespace watchme datafile 'watchme.dbf' size
5G autoextend on next 500m maxsize 20g;
```

Abbildung 23: Tablespace erstellen

Danach wird der Benutzer mit den entsprechenden Berechtigungen angelegt:

```
create user c##WATCHME_USER identified by xxxxxxxxxxxx
default tablespace watchme;
ALTER USER c##WATCHME_USER QUOTA UNLIMITED ON
watchme;
grant connect, resource to c##WATCHME_USER;
```

Abbildung 24: DB-Benutzer anlegen

Zum Anlegen der Tabellen muss die Datei create_schema.sql, welche sich auf der Installations-CD im Unterordner DB befindet ausgeführt werden.

Installation Sendungsdaten-Import

Um den Sendungsdaten-Import zu starten, muss auf dem Computer JRE 1.7 installiert sein. Zusätzlich muss eine Verbindung zum Datenbankserver hergestellt werden können.

Der Sendungsdaten-Import liegt als Executable-Jar vor und beinhaltet alle benötigten Libraries. Die vom Sendungsdaten-Import benötigten Werte für die Verbindungen zur Datenbank, den EPG-Daten und zu TMDb können über die Datei importer.properties gesetzt werden. Die Datei muss im gleichen Ordner wie die Executeable-Jar Datei liegen. In der Datei importer.properties werden folgende Werte gesetzt (siehe auch Abbildung 25):

- dbUrl: URL zum Datenbankserver, Beispiel: jdbc:oracle:thin:@localhost:1521:orcl
- dbUser: Datenbankbenutzer, Beispiel: c##WATCHME_USER

- dbPW: Passwort des Datenbankbenutzers
- tmdbKey: Kennung zur Abfrage von TMDB – wird von TMDB vergeben
- epGKey: Kennung zur Abfrage der EPG Daten – wird von epGdata.com vergeben

```
# Properties Datei mit Einstellungen für den WatchMe Sendungsdaten-Import
dbUrl:jdbc:oracle:thin:@localhost:1521:orcl
dbUser:c##WATCHME_USER
dbPW:xxxxxxxxx
tmdbKey:071812ff0ecae110994640919c0a7bc2
epgKey:Gv100v4x150aKG58PUI4uom2OPc8rYyd2iztajfO
```

Abbildung 25: Sendungsdaten-Import Properties Datei

Um den Sendungsdaten-Import regelmäßig auszuführen, kann ein beliebiges Werkzeug zur zeitbasierten Ausführung von Prozessen verwendet werden (z. B. Cronjobs in unixartigen Betriebssystemen und geplante Tasks in Windows).

Die JAR-Datei und die Datei importer.properties liegen auf der Installations-CD im Unterordner Sendungsdaten-Import.

Installation Web-Services

Für die Installation der Web-Services wird ein installierter Web-Server und eine JRE (Version 6 oder höher) vorausgesetzt. Die Web-Services werden als WAR-Archive zur Verfügung gestellt. Das WAR-Archive wurde für Tomcat 7 optimiert, kann jedoch auch auf anderen Web-Servern installiert werden. Das WAR-Archive befindet sich auf der Installations-CD im Unterordner Web_Services.

Da das Web-Service mit der Oracle Datenbank kommuniziert, muss sichergestellt sein, dass die Kommunikation zwischen Web-Server und Datenbank funktioniert. Die Datenbankverbindung kann in der Datei config.properties, welche im WEB-INF/classes/config/ abgelegt wird, angepasst werden. Dazu müssen die gleichen Parameter für die Datenbankverbindung wie bei der Installation des Sendungsdaten-Imports eingetragen werden: dbUrl, dbUser, dbPW.

```
# Properties Datei mit den Web-Service Einstellungen
dbUrl:jdbc:oracle:thin:@localhost:1521:orcl
dbUser:c##WATCHME_USER
dbPW:xxxxxxxxx
```

Abbildung 26: Web-Services config.properties

Installation Android App

Für die Installation der Android App wird ein Android Smartphone oder Tablet mit einer Android Version 4.1 (Jelly Bean) oder höher vorausgesetzt.

Um die Android App zu installieren, muss die watchMe.apk Datei per USB oder mittels SD-Karte auf das Android Handy kopiert werden. Die Datei befindet sich auf der Installations-CD im Ordner App. Danach muss auf dem Smartphone unter Einstellungen/Sicherheit die Einstellung „Unbekannte Quellen“ angehakt werden. Anschließend kann die Apk-Datei geöffnet und dadurch die Installation gestartet werden.

B. Glossar

API

Application Programming Interface; Eine API ist eine Programmierschnittstelle

EPG

Electronic Program Guide; EPG ist ein elektronisches Fernsehprogramm

Live-TV-Sendungen

Alle Fernsehsendungen die zum Zeitpunkt ihrer Ausstrahlung konsumiert werden

IPTV

Internet Protocol Television; Übertragung von Fernsehsendungen über das Internet Protokoll

VoD

Video-on-Demand; Angebot von Filmen oder sonstige Sendungen bei dem der Benutzer den Zeitpunkt der Nutzung frei bestimmen kann.