

Projektdokumentation im Modul Semantic Web

Historische Personen in Videospielen

Sonja Mauersberger

31.07.2016

Recherchefragestellung: Welche historischen Persönlichkeiten wurden als Vorbild für Charaktere in japanischen Videospielen genommen?

1 Inhaltliche Interpretation der Fragestellung

Einige historische Ereignisse inspirieren die Entwickler von Videospielen zu neuen Spielen. Dabei werden oftmals auch reale Persönlichkeiten als Charaktere in die Spiele übernommen. Besonders dem jungen Spieler ist oftmals nicht bewusst, hinter welcher Figur sich ein historisches Vorbild verbirgt. Mit der Beantwortung dieser Problematik beschäftigt sich diese Arbeit.

Es gibt kaum Informationen im Web über die Verknüpfung von historischen Persönlichkeiten und Videospielen. Diese Informationen sind, insofern vorhanden, nur im Subtext zu den Spielen oder Spielfiguren zu finden. Da die automatisierte Kontextsuche zur Zeit jedoch nur in der Theorie existiert, bleibt nur der Namensvergleich von Spielcharakteren und historischen Personen. Um einen Stringvergleich der Namen ermöglichen zu können, werden daher nur Namen aus lateinischen Buchstaben akzeptiert, d.h. Daten aus beispielsweise kyrillischen Buchstaben oder japanischen Schriftzeichen werden ignoriert.

Für diesen Vergleich werden Namenslisten sowohl von Spielcharakteren, als auch von historischen Persönlichkeiten benötigt.

Hierzu müssen zuerst einige Einschränkungen und Festlegungen getroffen werden.

Im Rahmen dieser Arbeit sind per Festlegung jene Spiele als japanische Videospiele zu verstehen, die von japanischen Publishern veröffentlicht wurden. Da ein Spiel mehrere

Publisher haben kann, wird in diesem Fall nur der japanische Publisher genannt. Eine Person sei "historisch", wenn sie berühmt war und bereits verstorben ist. Des Weiteren sollen zur Vereinfachung nur Spiele für die Playstation 2 (Konsole) in Betracht gezogen werden.

2 Relevante Datenquellen

Zur Bearbeitung der Fragestellung steht zu Beginn die Datenbeschaffung. Die folgenden Datenquellen wurden dazu genutzt.

2.1 historische Datenbank geboren.am

Die Personendatenbank für historische und berühmte Persönlichkeiten ist als Linked Data Endpunkt im Web realisiert. Sie umfasst wichtige Persönlichkeiten vergangener Tage und Prominente von heute. Auf geboren.am sind über 4.000 bekannte Personen mit ihren Biografien versammelt. [Kor16]

Link	http://geboren.am/
Datenformat	HTML
Schnittstelle	Linked Data, HTTP
Lizenz	CC
Open Data	***

2.2 List of Every Game Ever von DataBaser

Auf der Internetseite pastebin.com kann jeder Nutzer nach erfolgter Anmeldung Text online speichern. [pas16] Der Nutzer mit dem Nickname DataBaser stellte hier seine mit Hilfe anderer User selbst erstellte und sehr umfangreiche Liste veröffentlichter Videospiele ein. Es handelt sich hierbei um eine (auch ohne Anmeldung bei Pastebin) herunterladbare Textdatei mit über 43.000 Datensätzen. Ein Datensatz besteht dabei aus dem Titel des Spiels, einer Info zur Version, dem Erscheinungsjahr, dem/den Publisher(n) und den mit dem Spiel kompatiblen Konsolen.

Link	https://pastebin.com/EuxZMbWT
Datenformat	Text (herunterladbar)
Schnittstelle	Texteditor
Lizenz	des Autors (keine angegeben)
Open Data	***

2.3 Videospielehub wikia.com

Wikia ist eine Datenbank mit Bereichen (den sogenannten „Hubs“) zu Themen wie Videospielen, Serien, Animes, Fernsehserien und Comics. Innerhalb dieser Bereiche befinden sich einzelne in sich verlinkte Wikis zu konkreten Themen. Hier kann nach erfolgter Anmeldung jeder eigene Beiträge erstellen und bearbeiten. Obwohl von Wikia einige Layouts und Stylevorschriften bereitgestellt werden, sind diese nicht bindend. Die so entstehenden Webseiten unterscheiden sich daher (teilweise stark) im Aufbau und können ohne Anmeldung angesehen werden. Zur Beantwortung der Fragestellung dieses Projekts ist nur der Videospielehub interessant. Hier findet man Informationen zu vielen Videospielen, sowie zu deren Charakteren. [wik16]

Link	<code>http://de.videospielehub.wikia.com/wiki/Videospiele_Hub</code>
Datenformat	HTML
Schnittstelle	Linked Data, HTTP
Lizenz	CC-SA
Open Data	★★★★★

3 Extraktion relevanter Daten

Nach Lokalisation geeigneter Quellen, müssen die relevanten Daten extrahiert werden.

3.1 Extraktion historische Personen

In geboren.am gibt es diverse Kategorien, nach denen die Persönlichkeiten sortiert sind. Für eine möglichst ausführliche Extraktion der Namen empfiehlt sich die Kategorie “A-Z”, in der die Personen alphabetisch aufgelistet sind. Mittels Mouseover-Aktion über einen Namen, werden dort zudem das Geburtsjahr und ggf. das Sterbejahr angezeigt. Die Namen fungieren als Links zu den jeweiligen Seiten der Personen mit zusätzlichen Informationen und einem kurzen Infotext.

Die Informationen werden mittels Parsen der Webseiten aus dem Internet bezogen. Hierzu wurde mit dem Java HTML Parser “Jsoup” gearbeitet. [Hed16]

Listing 1: Beispiel für Jsoup-Extraktion von Personendaten

```
Document doc = Jsoup.connect("http://geboren.am/namen/a").get();
Elements charaktere = doc.select("body div li a.bub-b-top");
```

In Listing 1 sieht man die Funktionsweise des HTML-Parsers am Beispiels einer Persönlichkeit aus geboren.am. Zuerst wird die Verbindung mit der Webseite hergestellt (hier: "http://geboren.am/namen/a" für Liste aller Personennamen beginnend mit "A"). Dort werden dann alle Zeilen ausgewählt, deren Code-Hierarchie einem bestimmten Schema (hier: "body div li a.bub-b-top") entspricht.

Listing 2: Beispiel für Zerlegung von Jsoup extrahierten Daten

```
Elements caractere2 = caractere.eq(0);
String subjectlabel2 = caractere2.attr("Title");
String subject = ('http://geboren.am/' + caractere2.attr("Href"));
String[] subjectlabel = subjectlabel2.split( Pattern.quote( "(" ) );
// Speicherung der Daten historischer Personen in RDF-Format
```

Wie in Listing 2 zu sehen, wähle man eine der durch Listing 1 geparsten Zeilen und zerlege diese in ihren Titel ("Title" als String für Personen-Namen und Lebensdaten) und ihren Referenz-Link ("Href" als String für genauere personenbezogene Informationen). Danach wird der Titel weiter unterteilt in den Namen (hier: subjectlabel[0]) und die Lebensdaten (hier: subjectlabel[1]). Da die Lebensdaten in Klammern angegeben sind, ist die öffnende Klammer ein geeignetes Trennzeichen. Anhand des Vorhandenseins eines Sterbejahres kann nun festgestellt werden, welche der Personen als "historisch" zu betrachten ist und in die Datenbank mittels RDF-Format übergeben wird. Dieses Vorgehen wird anschließend für die anderen Zeilen wiederholt bis alle relevanten Personendaten von der Webseite geparkt wurden.

3.2 Extraktion Videospiele

Zuerst muss die Textdatei "List of Every Game Ever" [Bas16] von pastebin.com heruntergeladen werden. Anschließend kann durch den Import in eine Exceltabelle mit den entsprechenden Spaltenanpassungen, eine Vorsortierung getroffen werden. Für dieses Projekt werden nur die Spiele für die Playstation 2 - Konsole von japanischen Publishern verwendet.

Bei welchem der Publisher es sich um einen Japanischen handelt, kann mittels Vergleich mit der Wikipedia-Liste der Videospiel-Publisher [pub16] herausgefunden werden. In dieser Liste findet man die Namen der Publisher mit ihrem jeweiligem Hauptsitz und weiteren hier irrelevanten Daten.

3.3 Extraktion Videospiel-Charaktere

Da die Wikis in Wikia nicht einheitlich angelegt sind, ist es erforderlich die Auswahl der benötigten Seiten und Kontrolle der Informationen manuell durchzuführen. Aufgrund fehlender Einträge zu gewissen Spielen bzw. deren Charaktere dezimiert sich die Anzahl der verwendbaren Spiele von etwa 400 auf 24. Dabei fallen primär Spiele raus, die außerhalb Japans nie veröffentlicht wurden und zu denen es daher auch keine Übersetzungen im deutsch- oder englisch-sprachrige Wikia gibt.

Das Parsen der Charaktere von den noch übrigen 24 Spiele-Seiten geschieht wie bereits in Kapitel 2.1 beschrieben mittels Jsoup.

4 Verlinkung von Ressourcen und Import in einen Triplestore

Die Verlinkungen der Daten werden bereits bei der Extraktion der Daten erfasst. So werden die Spielcharaktere direkt von der Webseite des Spiels bezogen und diese Verbindung in Tripeln dargestellt. Die wichtigsten Beziehungen sind dabei:

- * PUBLISHER veroeffentlich SPIEL
- * SPIEL hascharakter CHARAKTER
- * PERSON gehoertzu KATEGORIE

Wobei KATEGORIE dabei entweder die Ausprägung “historische Person” oder “lebendige Person” annehmen kann, abhängig davon, ob die Person zum Zeitpunkt der Abfrage noch am Leben ist (bzw. dies in geboren.am vermerkt ist).

Listing 3: Beispiel für RDF-Schreibweise eines Tripels

```
<rdf:Spiel rdf:about="http://shadowhearts.wikia.com/...">
<rdfs:label xml:lang="de">Shadow Hearts:Covenant
</rdfs:label>
<dbp:hascharakter>
  <rdf:Charakter rdf:about="'http://shadowhearts.wikia.com/wiki/Rasputin">
    <dbp:name>Rasputin
    </dbp:name>
  </rdf:Charakter>
</dbp:hascharakter>
</rdf:Spiel>
```

Eine Zwischenspeicherung der Tripel erfolgt anschließend im XML/RDF-Format [ED06] (wie in Listing 3 dargestellt), damit im Falle eines Serverausfalls der Datenquellen immer noch auf diese Daten zurückgegriffen werden kann. Abschließend werden die Tripel

einem Triple-Store hinzugefügt. Für dieses Projekt wurde der Triplestore “Stardog” verwendet.

5 Anfrage an die Forschungswissensbasis

Zur Beantwortung der Fragestellung muss eine Anfrage an die Datenbank des Triplestores gestellt werden um relevante Daten herauszufiltern und sinnvoll in Beziehung zueinander zu setzen.

5.1 SPARQL-Anfrage

Bei der SPARQL-Anfrage werden die Strings von den Charakter-Namen der Videospielfiguren mit denen der historischen Persönlichkeiten verglichen.

Listing 4: SPARQL-Anfrage

```
PREFIX dbp: <http://dbpedia.org/property/>

SELECT *
WHERE { ?Namen rdf:type rdf:Person.
        ?Namen rdfs:label ?string1.
        ?Charakter rdf:type rdf:Charakter.
        ?Charakter dbp:name ?string2.
        Filter(contains(?string1,?string2))
}
```

5.2 Ergebnis der Anfrage

Die SPARQL-Anfrage aus Kapitel 5.1 (Listing 4) liefert als Ergebnis eine vierspaltige Tabelle mit 790 Einträgen. Ein Ausschnitt dieser Ergebnismenge ist in Tabelle 1 dargestellt. Hierbei treten jedoch diverse Probleme auf. Namen haben keine einheitliche Schreibweise, dies liegt zu meist in unterschiedlicher Übersetzung aus anderen Sprachen begründet. Des Weiteren beziehen sich selbst Namen, die gleich geschrieben werden nicht zwingend auf ein und dieselbe Person. Als zusätzliche Schwierigkeit kommt hinzu, dass viele Videospiel-Charaktere nur einen Namensteil besitzen. Dabei handelt es sich zu meist, aber nicht ausschließlich um den Vornamen.

Wie in Tabelle 1 ersichtlich, sind Ergebnisse, die nur nach dem Namensstring verglichen wurden, nicht eindeutig. Um die Fragestellung aussagekräftig beantworten zu können, ist

Namen	string1	Charakter	string2
...
geboren.am/person/...	Edgar Rice Burroughs	shadowhearts.wikia.com/...	Edgar
geboren.am/person/...	Ernest Rutherford	shadowhearts.wikia.com/...	Ernest
geboren.am/person/...	John Rawls	shadowhearts.wikia.com/...	John
geboren.am/person/...	John D. Rockefeller	shadowhearts.wikia.com/...	John
geboren.am/person/...	John August Roebling	shadowhearts.wikia.com/...	John
geboren.am/person/...	Grigori Rasputin	shadowhearts.wikia.com/...	Rasputin
...

Tabelle 1: Ausschnitt aus der Ergebnis-Tabelle

noch eine Verifikation mittels des Kontextes nötig. Dazu müsste die Charakterbeschreibung der Spielfigur mit der Beschreibung der historischen Person verglichen werden. Unter der zugehörigen URL findet man einen solchen kontextbezogenen Infotext.

Für das in Tabelle 1 dargestellte Beispiel stimmt nur ein Kontext überein. So ist Rasputin aus dem Spiel “Shadow Heards: Covenant” dem historischen Grigori Rasputin nachempfunden.

5.3 Verbesserungsvorschläge

Die Nutzung einer separaten Datenbank für alternative Schreibweisen würde das Problem der unterschiedlichen Schreibweise/Übersetzung von Namen beheben. So bietet beispielsweise DBpedia für einige Namen alternative Schreibweisen (unter `rdfs:label`) [dp16], jedoch nicht für alle Personen und es besteht das Problem, dass die Persönlichkeiten von geboren.am erst einmal mit den entsprechenden DBpedia-Einträgen verknüpft werden müssten.

Ein weiterer Ansatz wäre eine Abänderung der SPARQL-Anfrage. So könnte man beispielsweise nur Spiel-Charaktere mit 2 Namensteilen (Wörtern) im Stringvergleich berücksichtigen. Dies führt zu einer kleineren Ergebnismenge von höherer Qualität. So wird die Figur “John” nicht mehr mit jedem John aus der historischen Datenbank verbunden. Allerdings werden dann Charaktere wie “Rasputin” (Nennung per Nachname), deren Nennung im Spiel nur über einen Namensteil erfolgt, auch nicht mehr gefunden (obwohl der Kontextvergleich ergibt, dass er dem historischen russischen Leibarzt der Zarenfamilie nachempfunden ist [ras16]).

Eine Trennung des Namensstrings der Spiele-Charaktere in Vor- und Nachname (insofern vorhanden) und SPAQL-Abfrage nach beiden Namensteilen getrennt, löst das Problem mit der Reihenfolge (bei japanischen Namen). Dabei ist jedoch zu bedenken, dass eine automatisierte Trennung nur schwierig ist und einer Überwachung bedarf. Dies liegt

darin begründet, dass das Vorhandensein von Titeln (z.B. Prinzessin) oder zusätzlichen Vornamen (zweiter Vorname), ebenso wie die Reihenfolge dieser nicht einheitlich sind. In den Wikia-Einträgen zu Charakteren kann es sogar vorkommen, dass Zwillinge oder andere stark verbundene Charaktergruppen (wie “99 Dalmatiener” im Spiel “Kingdom Hearts”) zu einem Eintrag zusammengefasst werden.

Der vielversprechendste Ansatz wäre jedoch die Videospieler-Gemeinde für die Kontextsuche einzuspannen. Mittels Gamifikation könnte man dem manuellen Kontext-Vergleich einen Spieleanreiz verleihen. Durch Entwicklung einer Anwendung, die gleichzeitig sowohl den Kontext des Charakters als auch die Kontexte aller in Frage kommender historischer Persönlichkeiten anzeigt, kann eine Zuweisung getroffen werden. Dabei ist jedoch auch eine Zurückweisungsoption notwendig, falls keine der Persönlichkeiten einen Treffen darstellt. Freiwillige Spieler könnten nach diesem Prinzip Punkte für jeden Kontextvergleich bekommen und sich so mit Mitstreitern spielerisch zu messen. Eine endgültige Entscheidung, welche Figuren zu welcher Person zugeordnet oder vollständig aussortiert werden, wird erst durch mehrmalige Abgabe der gleichen Antwort getroffen um Fehler einzelner Spieler zu ignorieren.

6 Interpretation und Zusammenfassung

Da das Semantic Web auf einem Open-World-Szenario beruht und gewisse Daten nicht allgemein zugänglich oder auffindbar sind, kann die Lösung der in dieser Arbeit erörterten Fragestellung, keinen Anspruch auf Vollständigkeit erheben. Um dennoch eine umfangreichere Antwort zu bekommen, kann die Datenbasis beliebig erweitert werden.

Anregungen für solche Ergänzungen können sein:

- * auch Spiele nicht-japanischer Publisher verwenden
- * weitere Spiele für andere Konsolen (oder PC) inklusive ihrer Charaktere hinzufügen

Wie im EA Blog für digitale Spielkultur [Lor16] nachzulesen ist, findet man in aktuellen Titeln wie “Assassins Creed” und “Civilization” für PC viele berühmte historische Ereignisse und Persönlichkeiten.

- * weitere Quellen für historische Persönlichkeiten ausfindig machen

So liefert geboren.am kaum Einträge zu berühmten Angehörigen des Militärs, aber gerade diese werden gern in Videospielen referenziert. Die Liste der Persönlichkeiten kann nachträglich noch durch andere Quellen (z.B. Wikipedia) ergänzt werden.

Da es in dieser Arbeit um japanische Videospiele geht, sind besonders die Länder Japan und China interessant. Die unter den folgenden zwei Links aufgelisteten asiatischen

Persönlichkeiten bieten eine gute Ergänzung zu den Daten aus geboren.am:

* <https://de.wikipedia.org/wiki/Portal:China/Personen>

* <https://de.wikipedia.org/wiki/Portal:Japan/Personen>

Hier findet man beispielsweise auch den Feldherren Oda Nobunaga, der im Japan des 15. Jh. versuchte die einzelnen befeindeten Gebiete unter sich zu vereinigen [nob16] und im Spiel Samurai Warriors zu neuem Leben erwacht.

Das in dieser Arbeit vorgestellte Verfahren ist auf Grund der Schwierigkeiten bzgl. der unterschiedlichen Schreibweisen von Namen noch verbesserungswürdig. Die Antwort zu dieser Fragestellung konnte durch das hier vorgestellte Verfahren zwar eingegrenzt werden, macht jedoch eine nachträgliche Kontextuntersuchung notwendig. Die unter 5.3 vorgestellten Verbesserungsvorschläge könnten nun eine endgültige Antwort liefern.

Literatur

- [Bas16] DATA Baser. List of Every Video Game Ever. pastebin, besucht am 09.05.2016. <http://pastebin.com/EuxZMbWT>.
- [dbp16] Grigori Jefimowitsch Rasputin. DBpedia, besucht am 25.07.2016. http://dbpedia.org/page/Grigori_Rasputin.
- [ED06] et al. Elena Demidova, Shirish Kucheria. Einführung in RDF - RDF in wissenschaftlichen Bibliotheken. 2006. <http://www.iwi-iuk.org/seminarNotes/2/rdf-reader-v2.pdf>.
- [Hed16] Jonathan Hedley. Jsoup cookbook. herausgegeben von Jonathan Hedley, besucht am 25.07.2016. <http://jsoup.org/cookbook/>.
- [Kor16] Daniel Korieth. geboren.am. herausgegeben von Daniel Korieth, besucht am 21.06.2016. <http://geboren.am/>.
- [Lor16] Martin Lorber. Geschichte in Computer- und Videospielen. EA-Blog, besucht am 30.07.2016. <http://spielkultur.ea.de/themen/forschung-und-wissenschaft/geschichte-in-computer-und-videospielen/>.
- [nob16] Oda Nobunaga. Wikipedia, besucht am 31.07.2016. https://de.wikipedia.org/wiki/Oda_Nobunaga.
- [pas16] pastebin. Wikipedia, besucht am 21.07.2016. <http://de.wikipedia.org/wiki/Pastebin/>.
- [pub16] List of Video Game Publishers. Wikipedia, besucht am 21.06.2016. https://en.wikipedia.org/wiki/List_of_video_game_publishers.

- [ras16] Rasputin. Wikia, besucht am 31.07.2016. <http://shadowhearts.wikia.com/wiki/Rasputin>.
- [wik16] Videospielehub Wikia. Wikia, besucht am 10.05.2016. <http://de.wikia.com/Lizenz>.