

Eine Anwendung für die FIFA Fußball-Weltmeisterschaft mit "Semantic Web"-Technologien

Manuel Dudda, Robert Brylka und Frank Reichwein

Hochschule RheinMain Informatik Master of Science
Fachbereich Design Informatik Medien
Campus Unter den Eichen 5, 65195 Wiesbaden, Deutschland
{manuel.b.dudda,robert.c.brylka,frank.b.reichwein}@student.hs-rm.de
<https://github.com/manuel84/semanticwc>

Zusammenfassung. Große Sportereignisse haben schon immer große Menschenmassen begeistert. So auch die Events, die die Disziplin Fußball betreffen. Dieses Paper beschreibt ein Projekt, das sich als Ziel die Präsentation der Ereignisse der diesjährigen FIFA Fußball-Weltmeisterschaft gesetzt hat. Der Fokus dieser Anwendung liegt auf der Aktualität der präsentierten Daten sowie einer auf mobile Endgeräte angepassten Repräsentation. Für die Vielfalt der Daten sowie deren Richtigkeit sorgt eine Informationssynthese aus mehreren Quellen. Um die Informationen weitestgehend für eine automatische Weiterverarbeitung vorzubereiten, wurden alle Daten um semantische Bedeutung aus gängigen Sport-Ontologien ergänzt.

1 Einleitung

In den vielen Informationsseiten, welche die diesjährige FIFA Fußball-Weltmeisterschaft in Brasilien präsentieren, sucht man oft vergeblich nach gut strukturierten und leichtgewichtigen Applikationen, die schnell mit mobilen Endgeräten zu erreichen sind. Zudem werden viele Seiten nicht aktuell gehalten, was gerade bei einer so ereignisreichen Sportdisziplin wie Fußball unerlässlich ist. Bei unserem Projekt haben diese Aspekte die höchste Priorität erhalten. Um einem aktuellen Trend in der Webentwicklung zu folgen, wurde unsere Applikation mit Konzepten des semantischen Web entwickelt. Dabei werden alle Informationen mit einer für die Maschine verständlichen Syntax ergänzt. Dazu wurden die Informationen der FIFA-Webseite extrahiert und mit der Ontologie von BBC-Sport und der Ontologie Soccer-Voc des Hasso-Plattner-Instituts versehen. Für die Darstellung dieser Daten wurde eine Ruby on Rails Web-Applikation, welche zur Präsentation das "jQuery Mobile"-Framework einsetzt, entwickelt. Des Weiteren werden weiterführende Informationen, durch die Einbeziehung von DBpedia-Einträgen, eingeblendet.

2 Related Work

Bei der Recherche in dem Bereich der Fußballereignisse mit semantischen Hintergrund hat sich unsere Aufmerksamkeit auf drei Quellen gerichtet, die Sport-Ontologie¹ der BBC, die Soccer-Voc-Ontologie² des Hasso-Plattner-Institut sowie der "World Cup JSON"³-API.

Die "Future Media and Technology"-Abteilung der BBC setzt auf semantische Technologien, um die intern relevanten Inhalte automatisch verwalten zu können. Der erste große Einsatz der semantischen Technologien war die Gestaltung der Internetseite[1] mit Informationen zu der FIFA Fußball-Weltmeisterschaft 2010. Die dabei eingesetzte Ontologie ist seither frei zugänglich.

Die Arbeitsgruppe "Semantic Technologies"⁴ am Hasso-Plattner-Institut im Potsdam veröffentlichte 2013 ein Paper zu Linked Soccer Data[2]. Im Rahmen dieses Papers wurde ein Datensatz mit Fußballinformationen vorgestellt, der mit der LOD Cloud⁵ verknüpft wurde. Die dabei verwendete Ontologie kann frei verwendet werden und stellt eine gute Ergänzung der Sport-Ontologie der BBC dar.

Bei der World Cup API handelt sich um ein von Eric Stiens⁶ betreutes Projekt. Dessen Ziel ist es, die aktuellen Fußballereignisse für jeden frei verfügbar zu machen. Neben den frei verfügbaren Informationen, die als JSON Dateien vorliegen, ist auch die API⁷ für Weiterentwicklungen auf Github freigegeben.

3 Bestimmung der Informationsquellen

Die FIFA Fußball-Weltmeisterschaft ist allgemein ein beliebtes Thema in der digitalen Welt. Um geeignete Quellen für die gesuchten Informationen zu finden, wurde eine tiefgründige Recherche durchgeführt. Die unten aufgeführte Tabelle zeigt einen Ausschnitt der Bewertung der Quellen. Wie man an der Zusammenstellung erkennen kann, verfügen nicht alle Quellen über die gleiche Informationsfülle. Dabei ist die "Open Football"-Datenbank aufgrund der nicht aktuellen Informationen für weitere Verwendung uninteressant. Die Football Data sowie die "World Cup"-Datenbank zeigen beide eine sehr gute Fülle an brauchbaren Informationen. Bei einer weiteren Recherche zu der Word Cup Datenbank wurde die offengelegte API gefunden, die einen direkten Zugriff auf die offizielle Seite der FIFA ermöglicht. Wie der letzten Spalte in der Zusammenstellung zu entnehmen ist, ist die offizielle Seite des internationalen Fußballverbandes die beste Informationsquelle. Nach der Bewertung der Quellen wurde beschlossen, dass mit einem Fork der Soccer for good API das beste Ergebnis zu erwarten ist. Dieser Fork wurde unseren Bedürfnissen angepasst. Durch die Anpassungen wurde

¹ <http://www.bbc.co.uk/ontologies/sport>

² <http://purl.org/hpi/soccer-voc/>

³ <http://worldcup.sfg.io/>

⁴ http://hpi.de/meinel/knowledge_tech/semantics.html

⁵ <http://linkeddata.org/>

⁶ <http://softwareforgood.com/soccer-good/>

⁷ https://github.com/estiens/world_cup_json/

die Schnittstelle um Informationen zu den Spielern, Trainern und Schiedsrichtern erweitert. Es wurden damit alle für die Anwendung relevanten Informationen integriert. Dieser Fork bietet entsprechende REST⁸-Endpunkte an. Die Rückgabe dieses Services ist eine JSON⁹-Datei mit allen relevanten Informationen.

Tabelle 1. Zusammenstellung der wichtigsten Informationsquellen

	Football Data ¹⁰	Open Football ¹¹	Football DB ¹²	World Cup ¹³	FIFA ¹⁴
Teams	+	+	+	+	+
Gruppen	-	+	-	+	+
Spiele	+	+	-	+	+
Stadien	+	+	-	+	+
Uhrzeiten	+	+	-	+	+
Spieler	+	+	-	-	+
Spielergebnisse	+	+	+	+	+
Torschützen/min	+ / +	+ / +	-	+	+ / +
Schiedsrichter	-	-	-	-	+
Trainer	-	-	-	-	+
Spieltag	+	+	+	+	+
Aktuell	-	-	+	+	+
Besuchersch	+	-	-	-	+

Eine wichtige Rolle bei der Wahl der API hat letztendlich auch der finanzielle Aspekt gespielt. Es gibt eine ganze Reihe an gut organisierten und stets aktuellen Datenbanken, die jedoch eine regelmäßige Gebühr mit sich tragen. Für dieses Projekt wurde jedoch die Vereinbarung getroffen, rein quelloffene Lösungen zu verwenden. Dadurch wurden finanziell abhängige Projekte direkt bei der Recherche ausgeschlossen.

4 Ontologien

Um die Daten semantisch korrekt darzustellen, wurden verschiedene Ontologien verwendet. Für die Darstellung der Weltmeisterschaft wurde die BBC-Sport-Ontologie verwendet. Mit dieser wurden die verschiedenen Gruppenphasen, Spieltypen, Spiele sowie die zugehörigen Stadien abgebildet. Für die fußballspezifischen Daten wie Spielevents (Tore, Ein- und Auswechselungen sowie

⁸ **R**epresentational **S**tate **T**ransfer

⁹ **J**avaScript **O**bject **N**otation

¹⁰ <https://github.com/footballdata/fifadata>

¹¹ <https://github.com/openfootball/world-cup>

¹² <http://footballdb.herokuapp.com/api/v1/>

¹³ <http://www.worldcup.sfg.io/>

¹⁴ <http://www.fifa.com/>

Karten) wurde die Soccer-Voc-Ontologie des Hasso-Platner-Instituts verwendet. Diese erweitert die "Friend Of A Friend"-Ontologie. Außerdem wurde, für die zeitliche Einordnung der Spielevents, die Event-Ontologie verwendet.

```

@prefix semwc: <http://cs.hs-rm.de/~mdudd001/semanticwc/> .
@prefix bbcsport: <http://www.bbc.co.uk/ontologies/sport/> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix soccervoc: <http://purl.org/hpi/soccer-voc/> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .

semwc:GER_POR a bbcsport:Match.

semwc:GER_POR bbcsport:homeCompetitor dbpedia:Germany_national_football_team .

semwc:Kramer a soccervoc:SoccerPlayer;
             soccervoc:playsFor dbpedia:Germany_national_football_team;
             foaf:name "Kramer" .

```

Abb. 1. Exemplarische Ausschnitt aus der RDF.

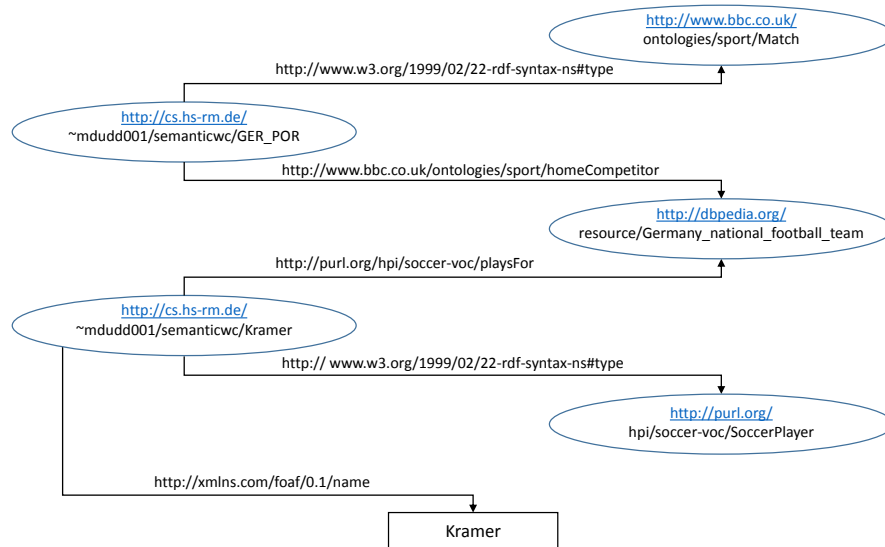


Abb. 2. Graph über den exemplarischen Ausschnitt aus der obigen RDF.

Die Abbildung 1 zeigt einen Ausschnitt aus den mit den Ontologien versehenen Daten. Der dazugehörige Graph ist in der Abbildung 2 zu sehen. An diesem Beispiel ist die Verflechtung der Ontologien sehr gut sichtbar. Aus den Tripeln ist ersichtlich, dass für jedes Spiel ein eigener Knoten mit den Länderkürzeln der

Spielteilnehmer erstellt wurde *GER_POR*. Diese Knoten sind vom Typ *bbsport:Match* und haben als *bbsport:homeCompetitor*-Property die Knoten der jeweiligen Nationalmannschaft *Germany_national_football_team*. Weiterhin gibt es einen Knoten für alle Spieler, welche vom Typ *soccer-voc:SoccerPlayer* sind. Diese sind mit dem Prädikat *soccer-voc:playsFor* mit den Mannschaften verbunden. Das zuletzt genannte Prädikat und der zuletzt genannte Typ stammen von der Soccer-Voc-Ontologie. Da *SoccerPlayer* eine Unterklasse von *foaf:Agent* ist, kann ein *SoccerPlayer* mit einem Prädikat der FOAF-Ontologie für einen Namen versehen werden.

5 Implementierung

Bei der Wahl des Framework für diese Anwendung fiel die Entscheidung zu Gunsten von "Ruby on Rails"¹⁵ aus. Durch die verfügbaren Erweiterungen bietet es eine sehr gute Unterstützung bei der Entwicklung von "Semantik Web"-Applikationen. Sei es bei dem Erzeugen von RDF-Daten oder der Erstellung von Abfragen mittels SPARQL. Für die Gestaltung der Seiten, sowie als Unterstützung für die Mobilien-Endgeräte, wurde die "jQuery Mobile"¹⁶-Bibliothek eingesetzt. Durch diese Bibliothek werden auf Mobilien-Endgeräten z. B. Swipe-Gesten unterstützt.

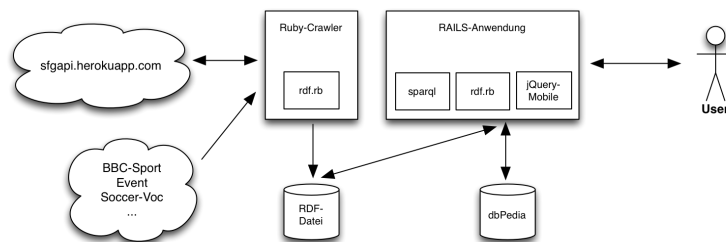


Abb. 3. Technik-Stack

Die Abbildung 3 zeigt eine Übersicht über den Technik-Stack. Die Applikation wurde in drei Bereiche unterteilt. Ein REST-Endpunkt, ein Crawler und die "Ruby on Rails"-Applikation, welche für die Präsentation der Daten verantwortlich ist.

Die "Ruby on Rails"-Applikation, welche die Informationen der FIFA-Seite crawlt und diese in einem REST-Endpunkt zur Verfügung stellt, ist ein Fork des SSoccer for good¹⁷-Projektes. Der entwickelte REST-Endpunkt ist auch separat über <http://sfgapi.herokuapp.com/matches> erreichbar. Der Service

¹⁵ <http://www.rubyonrails.org/>

¹⁶ <http://www.jquerymobile.com/>

¹⁷ https://github.com/estiens/world_cup_json/

wurde erweitert, indem die Aufstellung der Mannschaften, Trainer sowie die Schiedsrichter für jedes Spiel angezeigt werden.

Als zweiter Teil der Anwendung wurde ein Crawler implementiert. Der Crawler benutzt die Informationen der REST-API, versteht diese mit den passenden Ontologien und erstellt einen Graphen. Für die Erstellung des Graphen wird das Ruby Gem `rdf.rb`¹⁸ benutzt. Da die Anwendung sich zunächst auf ein sportliches Ereignis beschränkt, in diesem Fall die Fußball-Weltmeisterschaft, wurde für die Speicherung keine zusätzliche Datenbank verwendet. Stattdessen werden alle Datensätze als RDF-Tripel in einer Datei (*worldcup.ttl*) gespeichert. Für die Serialisierung der Daten wurde hierbei die Turtle-Syntax verwendet. Diese Datei enthält über 10.000 Tripel und ist ca. 900 KB groß. Dadurch kann der komplette Datenbestand, ohne Probleme, im RAM gehalten werden.

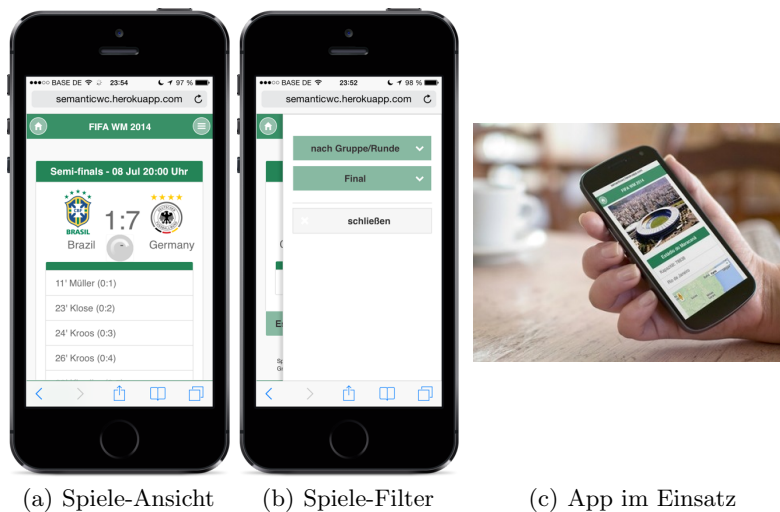


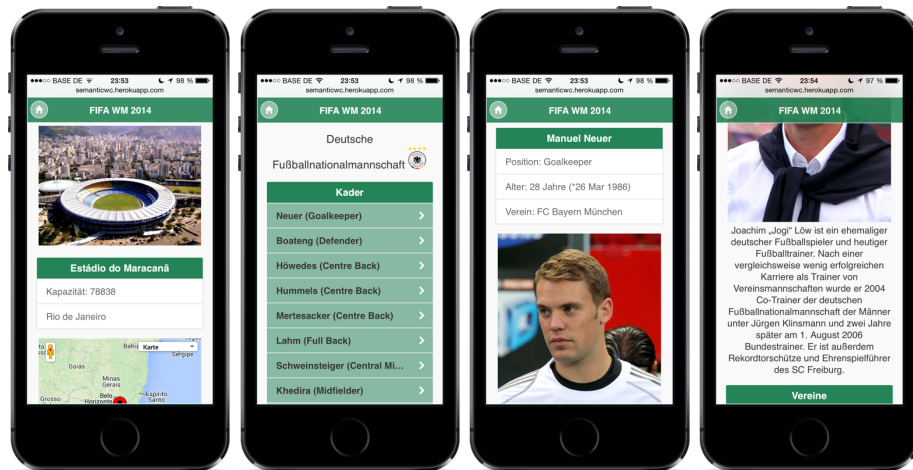
Abb. 4. Multi-Views

Für die Präsentation der Ergebnisse ist eine "Ruby on Rails"-Applikation entwickelt worden, die zur Darstellung "jQuery Mobile" verwendet. Für die Darstellung wurden verschiedene SPARQL-Abfragen erstellt, um die entsprechenden Informationen aus der Datei zu extrahieren. Da die Informationen in der Datei mit DBpedia verknüpft sind, mussten außerdem SPARQL-Abfragen für den SPARQL-Endpunkt von DBpedia entwickelt werden. Die Anwendung stellt alle Spiele mit Ergebnissen, Toren und Torschützen da. Außerdem gibt es für jede Mannschaft eine Ansicht, die den Kader darstellt. Die Informationen zu den jeweiligen Spielern kommen von DBpedia. Des Weiteren gibt es eine Ansicht der Stadien, welche die Position des Stadions mittels Google-Maps darstellt. Die Geo-Position der Stadien wird ebenfalls bei DBpedia abgefragt. Die Applikation

¹⁸ <https://github.com/ruby-rdf/rdf>

ist unter <http://semanticwc.herokuapp.com/>, der zugehörige Quellcode unter <https://github.com/manuel84/semanticwc> zu erreichen.

In der Abbildung 4 sind mehrere Screenshots aus unterschiedlichen Ansichten der Anwendung zu sehen. In Abb. 4(a) ist die Startseite zu sehen, auf der immer der aktuelle Spieltag angezeigt wird. Abb. 5(a) ist ein Ausschnitt aus der Ansicht, die Informationen zu Stadien präsentiert. Abb. 5(d) zeigt wiederum eine Ansicht, die die Informationen zu einem Trainer präsentiert. Da die Anwendung vorwiegend auf mobile Endgeräte ausgelegt ist, wurde bei der Gestaltung der einzelnen Ansichten großer Wert auf die Übersichtlichkeit gelegt. Für eine schnelle Navigation wurden zudem zwei Buttons eingepflegt. Der "Home"-Button, der in jeder Ansicht in der oberen linken Ecke zu sehen ist, führt immer zu der Seite mit dem aktuellen Spieltag. Mit dem "Bar"-Button, in der oberen rechten Ecke der Spiele-Ansicht, kann eine Seitenansicht mit mehreren Filteroptionen eingeblendet werden. Dabei kann der Inhalt nach dem Team, der Gruppe/Runde, dem austragenden Stadion oder dem Datum eingeschränkt werden.



(a) Stadion-Ansicht (b) Team-Ansicht (c) Spieler-Ansicht (d) Trainer-Ansicht

Abb. 5. Single-Views

Aus der Hauptansicht, in der immer der aktuelle Spieltag angezeigt wird, kann mit einer Wischgeste nach links oder rechts zu den bereits stattgefundenen, respektive in der Zukunft liegenden Spielen gewechselt werden. Die gleiche Funktionalität liegt auch bei der Ansicht, die durch Anwenden der Filteroperation *nach Team*, *nach Stadion*, *nach Gruppe/Runde* sowie *nach Tag* gewählt wird. In den Single-Views (Abb. 5) kann mit einer Wischgeste von links nach rechts zurück navigiert werden.

6 Fazit und Ausblick

Im Rahmen dieses Projekts wurde eine Anwendung entwickelt, die alle wichtigen Informationen, zu der 2014 in Brasilien stattfindenden FIFA Fußball-Weltmeisterschaft, bereithält. Als Informationsquelle dient hierbei die offizielle FIFA-Webseite, die mit Daten aus der DBpedia ergänzt wurde. Die semantische Aufbereitung wurde mit Hilfe der BBC-Sport sowie der Soccer-Voc-Ontologie gestaltet.

Unsere Anwendung bietet eine solide Grundlage für Weiterentwicklungen, die nicht nur den Bereich Fußball betreffen. Bezogen auf die Fußballereignisse kann eine leichte Portierung auf die nächste Weltmeisterschaft oder Europameisterschaft stattfinden. Denkbar ist aber auch eine Abwandlung zur Informationspräsentation der Spiele der nationalen Meisterschaften, beispielsweise der Fußball-Bundesliga.

Die Struktur- und Spieldaten der Weltmeisterschaft umfassen über 10.000 RDF-Tripel. Diese sind weitestgehend mit DBpedia-URIs verknüpft. Für die Datenabfragen wurden über 200 Zeilen SPARQL-Code implementiert. Leider ist eine Zuverlässigkeit von DBpedia nicht garantiert. Während der Implementierungsphase mussten wir Änderungen am Sparql-Endpoint feststellen, die zum Ausfall der Applikation führten. Des Weiteren hat man immer wieder mit angekündigten Wartungsarbeiten sowie Performanceproblemen seitens DBpedia zu kämpfen. Unserer Einschätzung nach steht das DBpedia-Projekt noch in den Kinderschuhen - die Daten sind veraltet, Bildreferenzen teilweise nicht mehr existent, Attribute werden inkonsistent mit Prädikaten versehen, Sprachattribute sind oftmals falsch. So entstanden große, unübersichtliche SPARQL-Abfragen, die auf viele optionale Prädikate der gleichen Semantik abzielen. Dieser Umstand ist insbesondere bei Namen zu beobachten, die bis zu 10 verschiedene Prädikate aus 5 Ontologien berücksichtigen. Die Qualität und Verfügbarkeit der Daten ist inkonsistent und führt nur durch heuristische Methoden zu akzeptablen Ergebnissen.

Literatur

1. BBC ONTOLOGIES, A simple ontology for representing competitive sports events.
 Letzter Zugriff: 26. Juni 2014
http://www.bbc.co.uk/blogs/legacy/bbcinternet/2010/07/the_world_cup_and_a_call_to_ac.html
2. Linked Soccer Data, Tanja Bergmann, Stefan Bunk, Johannes Eschrig, Christian Hentschel, Magnus Knuth, Harald Sack, and Ricarda Schüller
 Hasso Plattner Institute for Software Systems Engineering, Potsdam, Germany
 Letzter Zugriff: 22. Juni 2014
<http://ceur-ws.org/Vol-1026/paper6.pdf>
3. Semantic Media Mining,
 Letzter Zugriff: 22. Juli 2014
<https://smm2013.blogspot.de/2012/11/soccer-voc.html>
4. DBpedia, Letzter Zugriff: 22. Juli 2014
<http://http://dbpedia.org>