

Entwicklung eines Frameworks zur Erstellung und Durchführung von Ontologietests im Kontext der Health Intelligence Platform

Jessica Truber

Masterarbeit • Studiengang Informatik • Fachbereich Informatik und Medien • 19.05.2017

Aufgabenstellung

Die Arbeit umfasst die Konzeption und Entwicklung eines Frameworks zur Erstellung und Ausführung von Ontologietests in der Health Intelligence Platform (HIP) [1]. Es werden Qualitätseigenschaften von Ontologien erarbeitet, die in einem zweiten Schritt der Ausarbeitung einer Testsystematik dienen. Die softwaretechnische Umsetzung hat zum Ziel, die schnelle und systematische Erstellung von Testscenarien während der HIP-Entwicklung aber auch zur Laufzeit während des Ontology Engineering Zyklus zu unterstützen.

Health Intelligence Platform (HIP)

Die ORTEC Gesellschaft für Organisations- und Technologieberatung mbH begann die Entwicklung der heutigen HIP als IT-Lösung zur Unterstützung von klinischen Studien, die Ein- und Ausschlusskriterien mit vorhandenen Patientendaten vergleicht und mögliche Probanden vorschlägt. Die aktuelle HIP bietet mittlerweile eine Sammlung von Anwendungen, welche auf der Verarbeitung und Verwaltung von strukturierten und unstrukturierten Patientendaten basieren. Die Kernkompetenz stellt die linguistische Analyse klinischer Dokumente und die Extraktion von Fakten dar. Diese Aufgabe übernimmt die Linguistische Pipeline in Zusammenarbeit mit der HION (Health Intelligence Ontology).

Wissensverwaltung in der HIP

Die sogenannte HION umfasst mehrere Ontologie-Dateien im Turtle-Format, die sich in medizinische Fachbereiche teilen und für jeden Kunden anwendungsorientiert zusammengestellt werden. Die aktuelle Repräsentation liegt in OWL vor, unter der Benutzung von RDFS, und umfasst vor allem Symptome, Diagnosen, Therapien, Attribute und Relationen.

Aufgrund der Wandelbarkeit medizinischen Wissens und der Anwendungsorientierung der HION spielt das Thema der Ontologieevolution eine große Rolle in der HIP. Zum einen werden Änderungen durch die Entwickler selbst durchgeführt, zum anderen sollen auch die Kunden in Zukunft dazu in der Lage sein. Um die Qualität der HION in jeder Phase der Wissensverwaltung kontrollieren und sicherstellen zu können, wird ein Framework benötigt, welches die Erstellung und Durchführung von Ontologietests erleichtert.

Konzeption

Zur Konzeption des Frameworks wurden Anforderungen aufgestellt, die in erster Linie die Überprüfung speziell auf die HION abgestimmter Qualitätskriterien beinhalten. Zusätzlich müssen die durch das Framework entstehenden Tests einfach bedienbar sein und zur Vereinfachung der späteren Integration in die HIP auf der Programmiersprache Java unter der Benutzung der Frameworks JUnit [2] und RDF4J [3] basieren.

Wie in Abbildung 1 zu erkennen ist, wurden in der Planung strukturelle und semantische Tests unterschieden, die über eine nicht zum Framework gehörende Zugriffsschicht angesprochen werden und auf einem Repository agieren, welches die zu testende HION enthält.

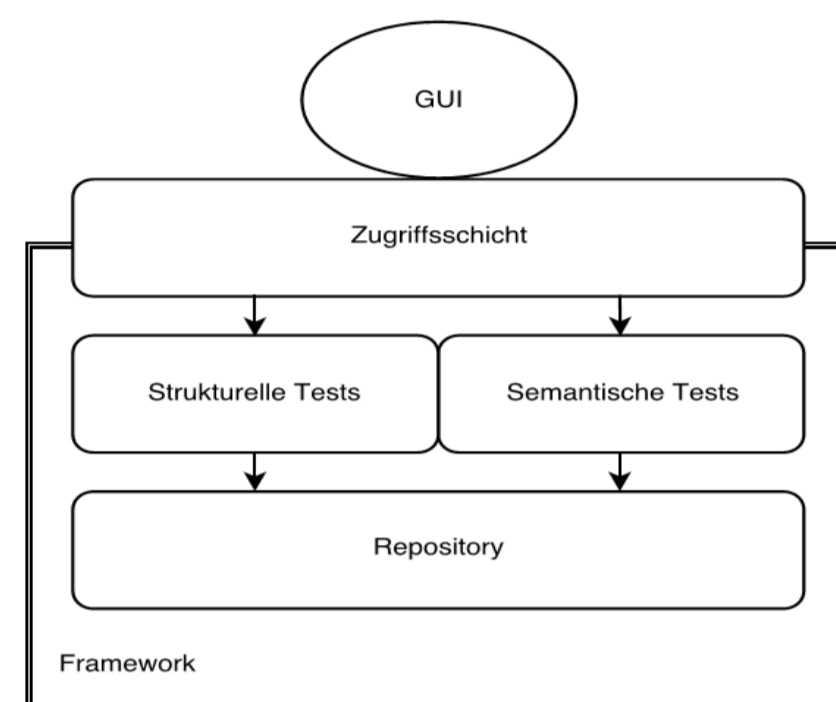


Abb. 1: Komponenten des Test-Frameworks

Strukturelle Tests umfassen dabei die Überprüfung der strukturellen Qualitätskriterien und der von der HION erwarteten Struktur, während semantische Tests dafür sorgen, dass die Ontologie in ihrer Anwendung wie erwartet funktioniert. In diesem Fall muss sie in Zusammenarbeit mit der Linguistischen Pipeline aus einem Dokument die erwarteten Fakten generieren.

Umsetzung

Die HIP-unabhängige Implementierung des Test-Frameworks wurde in der Entwicklungsumgebung IntelliJ IDEA [4] durchgeführt. Die strukturellen Tests entstanden prototypisch nach dem Vorbild der sogenannten *Test-Driven Linked-Data Quality Methodology* [5], während die semantischen Tests in Anlehnung an schon vorhandene *OntologyPipelineTests* der HIP gemockt wurden. Für jeden Anwendungsfall des Frameworks muss eine neue Test Suite implementiert werden, welche die einzelnen Tests ansteuert und ausführt. Das Framework stellt abstrakte Klassen zur Verfügung, um dem Nutzer die Erstellung der konkreten Tests zu vereinfachen.

Zum Testen des Frameworks wurden einige Testfälle beispielhaft umgesetzt.

Fazit

Im Rahmen dieser Arbeit entstand ein Ontologietest-Framework, welches in Form einer JAR-Datei in jedes beliebige Maven-Projekt integriert und dort genutzt werden kann. Vor der eigentlichen Verwendung in der HIP müssen allerdings noch einige Verbesserungen, wie die Integration in den Entwicklungsprozess, die endgültige Einbindung der *OntologyPipelineTests* oder die Implementierung der Zugriffsschicht, vorgenommen werden.

Quellen

- [1] HIP: <http://www.ortec.org>, letzter Zugriff am 18.05.2017
- [2] JUnit: <http://junit.org/junit4/>, letzter Zugriff am 18.05.2017
- [3] Eclipse RDF4J: <http://rdf4j.org/>, letzter Zugriff am 18.05.2017
- [4] JetBrains IntelliJ IDEA: Version 2017.1.2, <https://www.jetbrains.com/idea/>, letzter Zugriff am 18.05.2017
- [5] D. Kontokostas, P. Westphal, S. Auer, S. Hellmann, J. Lehmann, R. Cornelissen: Test-Driven Evaluation of Linked Data Quality. WWW '14 Proceedings of the 23rd international conference on World wide web, S. 747-758. 2014