

Erkennung und Reparatur von Inkonsistenzen in einer medizinischen Ontologie

Darya Martyniuk

Bachelorarbeit • Studiengang Informatik • Fachbereich Informatik und Medien • 18.09.2017

Aufgabenstellung

Ziel der Bachelorarbeit ist die Untersuchung einer medizinischen Ontologie (die WNC-Ontologie) im Hinblick auf Inkonsistenzen sowie die Entwicklung eines Algorithmus zur Reparatur von widersprüchlichen Konzepten, die keine Ausnahmen darstellen und deren Modellierung einem der beiden in der Abb. 1 dargestellten Mustern entspricht.

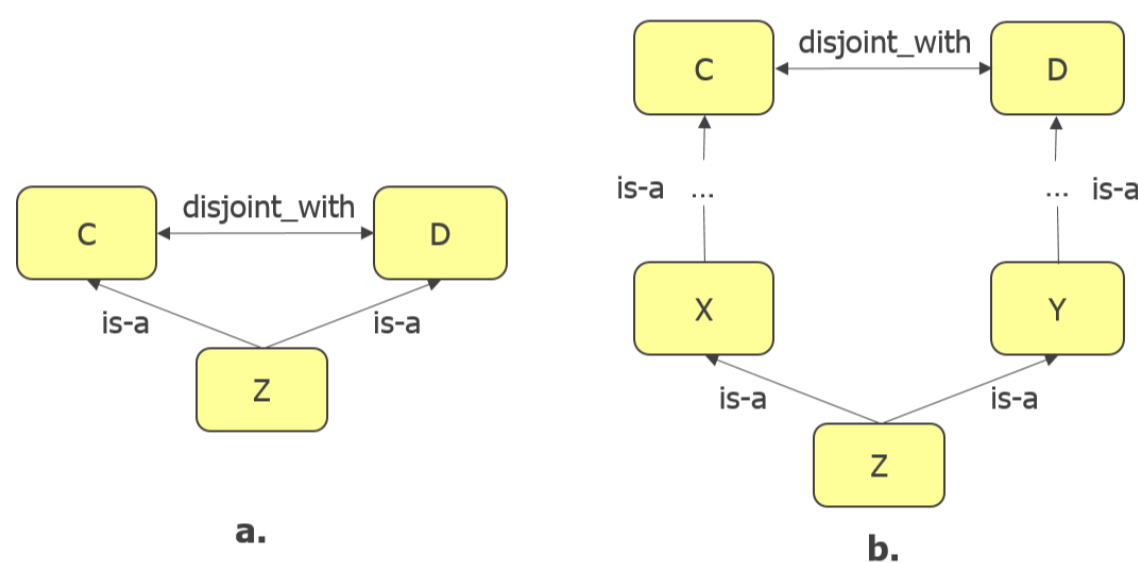


Abb.1: Modellierungsmuster von WNC-Konzepten, die mit dem Algorithmus repariert werden müssen.

Die WNC-Ontologie

Die WNC-Ontologie wird von dem Berliner Unternehmen ID GmbH & Co. KGaA entwickelt. Sie bildet die in der Wingert-Terminologie dargestellten medizinischen Begriffe als Konzepte und die Zusammenhänge zwischen diesen Begriffen als Relationen ab. Die Wissensbasis der WNC-Ontologie entstand als Ergebnis der Datenmigration aus der Darstellungsform als Semantisches Netz ID MACS® in Form einer Ontologie. Im Juli 2017 beinhaltete die Ontologie über 70 000 Konzepte und mehr als 370 000 Relationen, verteilt auf 350 Relationstypen.

Untersuchung der WNC-Ontologie

Im Rahmen der Arbeit erfolgte eine Untersuchung der WNC-Ontologie (Version vom 21. Juli 2017) auf logische, semantische und strukturelle Inkonsistenzen. Es wurden sowohl automatische als auch manuelle Methoden zur Diagnose von Inkonsistenzen eingesetzt.

Konzeption des Algorithmus

Der Algorithmus soll die Reparatur von WNC-Konzepten leisten, die durch zwei disjunkte Elternkonzepte subsumiert und aus diesem Grund unerfüllbar sind.

Die Ursachen für die widersprüchliche Modellierung können sehr unterschiedlich sein: ein menschlicher Fehler, eine inkorrekte Verwaltung von Homonymen, ein falsches Modellierungsvorgehen bei der Abbildung von Prozessen und weitere. Die automatische Reparatur der Konzepte nach einem Muster würde die Ursachen der Entstehung von Widersprüchen und somit die Semantik der einzelnen widersprüchlichen Konzepte nicht berücksichtigen. Aus diesem Grund wurde in dieser Arbeit auf eine komplett automatische Reparaturmethode verzichtet.

Stattdessen wurde entschieden, die Reparatur stark auf die

Kommunikation mit den Domänenexperten und somit auf die Berücksichtigung der Semantik der Konzepte zu beziehen.

Zu diesem Zweck wurden Fragen in natürlicher Sprache (engl. Natural Language Questions, Abk.: NLQs) ausgearbeitet, die aus der Ontologie automatisch generiert und den Domänenexperten gestellt werden müssen. Die Beantwortung dieser Fragen bestimmt neue konsistente Modellierungsvorschläge für widersprüchliche Konzepte. Ein wesentlicher Vorteil dieser Vorgehensweise ist die Unterstützung der effektiven Kommunikation zwischen IT-Spezialisten und Domänenexperten [1].

Implementierung

Für die Arbeit mit der Ontologie wurde die OWL API [2] als eine Java-Bibliothek in dem Eclipse Editor eingebunden. Der Input für den Algorithmus sind Konzepte, deren Modellierung einem der beiden in der Abb. 1 dargestellten Mustern entspricht und die keine Ausnahmen aus den üblichen Zuständen darstellen. In der Vorbereitungsphase sollen solche Konzepte und ihre disjunkten Elternkonzepte herausgefunden werden. Dies erfolgte mithilfe des ELK Reasoners [3] (für die Abfragen der impliziten Konzepthierarchie) und des Structural Reasoners [4] (für die Abfragen der expliziten Konzepthierarchie). Bei dem automatischen Erzeugen von NLQs aus der Ontologie und der Generierung der neun Modellierungsvorschläge wurde zwischen Fall a und Fall b in der Abb. 1 differenziert. Das primäre Testen hat gezeigt, dass die Fragen und die Modellierungsvorschläge korrekt und verständlich generiert wurden. Vorgeschlagene Änderungen ermöglichen, die Konsistenz von einigen widersprüchlichen Konzepten herzustellen. Der Algorithmus lässt sich jedoch optimieren und erweitern.

Fazit

Diese Arbeit beschreibt den Prozess und die Ergebnisse der Untersuchung der WNC-Ontologie auf Inkonsistenzen. Sie stellt außerdem einen Algorithmus zur Reparatur von widersprüchlichen Konzepten der Ontologie vor. Der Algorithmus basiert auf der Überprüfung der Semantik dieser Konzepte mithilfe der aus der Ontologie generierten Fragen in natürlicher Sprache und Beantwortung der Fragen von Domänenexperten. Der entwickelte Algorithmus kann zukünftig durch die Automatisierung bestimmter Schritte und die Erweiterung um weitere Reparaturmöglichkeiten optimiert werden.

Quellen

- [1] Abacha et al.: Medical Ontology Validation through Question Answering. In Proceeding: 14th Conference on Artificial Intelligence in Medicine, S. 196 - 205, Springer, 2013.
- [2] OWL API: http://semanticweb.org/wiki/OWL_API.html, letzter Zugriff am 09.09.2017.
- [3] ELK Reasoner: <https://www.cs.ox.ac.uk/isg/tools/ELK/>, letzter Zugriff am 09.09.2017.
- [4] Structural Reasoner: <http://soft.vub.ac.be/svn-pub/PlatformKit/platformkit-kb-owlapi3-doc/doc/owlapi3/javadoc/org/semanticweb/owlapi/reasoner/structural/StructuralReasoner.html>, letzter Zugriff am 09.09.2017.