

Triologo: De Ontologia

Triologo: De Ontologia

Abstract

This article is about ontologies and their inspection as well as their discussion inside a philosophical dialogue from three different perspectives: a philosophical, a linguistic, and a quite central computer science perspective. Starting from the main question, what ontologies mean in each of the mentioned disciplines, we show in which way ontologies are so meaningful for the current and future development of the web. Besides this we discuss – in the same dialogue – the aspects of practicability and the more formal aspect of denotation using a concrete example. The discussion ends with a rather general contemplation on the interpretation of meaning and being from an ontological perspective.

Zusammenfassung

Im folgenden Artikel soll das Thema der Ontologien und seine Bedeutung im Kontext des Semantic Web in der Tradition eines philosophischen Lehrgesprächs aus drei unterschiedlichen Perspektiven betrachtet und diskutiert werden: aus einer philosophischen, einer linguistischen und ganz zentral einer informatischen Perspektive. Ausgehend von der Frage, was Ontologien in den einzelnen Disziplinen bedeuten, soll gezeigt werden, in welcher Form sie von so großer Bedeutung für die aktuelle Weiterentwicklung des Web sind. Dabei sollen auch Fragen der Praktikabilität und der formalen Beschreibung im Rahmen einer konkreten Umsetzung angesprochen werden. Das Gespräch endet mit einer allgemeinen Beurteilung dessen, wie Bedeutung und Sein im ontologischen Sinn zu interpretieren sind.

Schlüsselwörter: Ontologien, Semantic Web, RDF, Semantik, Wissensrepräsentation, Inferenz

Einleitung: Zur Genese des Trialogs

Bei meiner Teilnahme als Bibliothekar an den ersten drei SWIB-Kongressen habe ich mir immer wieder die Frage gestellt, welche von den ganzen Fachbegriffen, die so selbstverständlich und ohne weiteres benutzt werden, sind den meisten Zuhörern überhaupt bekannt? Wenn Worte wie Ontologie, Tripel, Graph, URI usw. benutzt werden, können sich die meisten etwas Präzises darunter vorstellen? Bei einigen Fragen an andere Kongressteilnehmer habe ich festgestellt, dass dies durchaus nicht der Fall war.

Unter anderem habe ich zufällig auch einen Informatiker und einen Linguisten gefragt, die gleich gutmütig angefangen haben, meine vielen Zweifel zu klären. Aus diesem zufälligen Treffen und mit der Absicht und Hoffnung, dass dies auch vielen anderen mit denselben Zweifeln willkommen sei, ist dem Linguisten die Idee eines mäeutischen Trialogs gekommen, dem sich auch der Philosoph und der Informatiker mit Enthusiasmus angeschlossen haben.

Der folgende Trialog ist das Ergebnis von mehreren, über Internet-Telefonie geführten Gesprächen mit einer zu meist wortgetreuen Wiedergabe dessen, was aus dem Gespräch selbst hervorgetreten ist.

Der Philosoph (Maurizio Grilli)

Triologo

Was sind Ontologien eigentlich und uneigentlich?

Der Linguist (René Schneider): Eigentlich ist die Ontologie ja etwas Philosophisches, oder sie gehörte früher zum Arbeitsfeld der Philosophen, denn diese haben zum ersten Mal darüber reflektiert und dieses Wort erfunden. Wie sehen die Philosophen die Ontologie?

Der Philosoph (Maurizio Grilli): Ontologie stammt etymologisch aus zwei Wörtern, ontos und logos und ontos ist das Partizip Präsens des Verbs εἶναι (Sein) und λόγος

Maurizio Grilli¹

Fabio Ricci²

René Schneider³

1 Claudiana,
Landesfachhochschule für
Gesundheitsberufe, Bozen,
Italien

2 semweb.ch, Küsnacht ZH,
Schweiz

3 Haute Ecole de Gestion,
Carouge GE, Schweiz

heisst das Wort, die Lehre, also geht es um die Lehre über das „Sein“.

Aber was ist eigentlich das Sein? Darüber muss man sich kurz Gedanken machen. Auf der Welt, in unserer Umgebung, kann „alles“ sein, d.h. das Sein ist das Grundsätzliche, die Grundlage von jedem Ding, auch von dem, was keinen Anfang hat und auch kein Ende. Auf dem Sein basiert alles, was es im Universum gibt. Wir bestehen teilweise aus Elementen, die mit dem Ende unseres Lebens nicht aufhören werden. Die gab es schon vorher, die wird es nachher geben. Diese Elemente leben nicht. Sie sind. Das sind die Atome, aus denen – wie sonst alles auch – die Lebewesen bestehen. Die Ontologie beschäftigt sich mit diesem Wesen aller Dinge. Ist das nachvollziehbar?

L: Ja sehr. Welches sind denn die Philosophen, die sich hauptsächlich mit Ontologien beschäftigt haben?

P: Angefangen hat es historisch gesehen mit Parmenides. Parmenides hat vor Sokrates im 6. Jahrhundert vor Chr. gelebt. Er hat sich systematisch mit der Ontologie beschäftigt, mit dem Grundsätzlichen, mit demjenigen, was alles Weitere verallgemeinert. Der uns Philosophen chronologisch am nächsten stehende Vorgänger, der sich auf grossartige Art und Weise damit beschäftigt hat, ist Martin Heidegger. Sein Hauptwerk „Sein und Zeit“ wurde in den 20er Jahren des letzten Jahrhunderts veröffentlicht.

Doch wie sieht es mit den Linguisten bzw. Computerlinguisten und den Informatikern aus? Haben sie dasselbe Verständnis von Ontologie? Denn die Semantic Web-Experten benutzen diesen Begriff ja ganz selbstverständlich, aber da verstehe ich kein Wort.

L: Eigentlich ist ja die Ontologie zum Exerzierfeld der Informatiker geworden! Wahrscheinlich beschäftigen sich derzeit mehr Informatiker mit Ontologien (oder mit dem, was sie darunter verstehen) als Philosophen! Von daher überlasse ich zuerst dem Informatiker das Wort!

Der Informatiker (Fabio Ricci): Ontologien habe ich erst im Zusammenhang mit KI (Künstlicher Intelligenz) gegen Ende der 80er Jahre des letzten Jahrhunderts kennen gelernt. In der KI hiess es immer: Bevor man irgendwas aussagt, programmiert, modelliert, sollte man immer die Objekte und die Beziehungen der Objekte zueinander aufschreiben (das wäre ja die ontologische Basis) und erst danach könnte man ein Verhalten hinzufügen – womit wir mit „Verhalten“ jetzt eher in Richtung KI gehen, und damit das eigentliche ontologische Feld verlassen. Das Objekt der Ontologie in der KI ist das schriftlich dargestellte Wissen (also nicht das Sein an sich) – wahrscheinlich wird der Linguist dazu noch mehr sagen – also zielt es darauf ab, eine Verbindung zwischen Mensch und Maschine herzustellen. D.h. was heute eine Ontologie enthält, sind Beschreibungen von Objekten und Beziehungen unter diesen beschriebenen Objekten.

P: Heißt „Objekte beschreiben“ auch „Objekte strukturieren“?

I: Ja, sicher! Wobei durch die Beschreibung sichergestellt ist, dass ein Merkmal an genau einer einzigen Stelle beschrieben ist, das nennt man dann orthogonal.

L: Ich habe einmal so definiert, was eine Ontologie ist: „Ontologien sind logisch einwandfrei definierte, maschinell interpretierbare Beschreibungen des Sprach- und Weltwissens.“

Für den Sprachwissenschaftler oder den Computerlinguisten ist die Ontologie von daher auch interessant in Bezug zur Künstlichen Intelligenz, nämlich inwiefern ich Maschinen das „Denken“ oder Maschinen die „Sprache“ beibringen kann. Für den Linguisten ist aber alles nur aus der Perspektive der Sprache relevant und die Ontologie ist dabei sehr wichtig, weil für den Linguisten eben die Welt über die Sprache dargestellt wird.

Es gibt für uns einerseits das Weltwissen – d.h. konkret die Aussagen, die wir über die Welt machen können – (und zwar mit den Mitteln der Sprache) und es gibt das Sprachwissen, d.h. das Wissen über das Instrument mit dem wir dies tun, nämlich die Sprache. Es gibt einerseits diese strenge Unterscheidung und andererseits die damit einhergehende Verbindung zwischen dem Weltwissen und dem sprachlichen Wissen. Und nur das ist von Relevanz. Letztlich kann man in Anlehnung an Wittgenstein sagen: „Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt“, also ontologisch ist für den Sprachwissenschaftler nur das, was auch sprachlich ausgedrückt werden kann.

Interessant aus der computerlinguistischen Perspektive sind die maschinelle Beschreibbarkeit oder die logisch einwandfreie Beschreibbarkeit und deren maschinelle Interpretierbarkeit.

I: Ist das auch für Philosophen verständlich?

P: Meint man hier mit „Weltwissen“ eine Art „objektives“ Wissen (was es ja natürlich nicht geben kann, denn die Menschen können Sachen nur subjektiv sehen) oder etwas anderes?

L: Im Prinzip kann man hier unter Weltwissen all das verstehen, was über die Welt gesagt werden kann – und das kann hier sowohl wahr als auch falsch sein, die Frage der Wahrheit ist letztlich hier untergeordnet; die Ontologie muss auch in der Lage sein, widersprüchliche Aussagen zu erfassen, weil Bedeutung ja auch etwas Dynamisches ist, und – wie wir jetzt wissen – die Welt sich auch ständig ändert. Also kommt es nicht darauf an, wie die Welt an sich ist, sondern wie die Welt durch den Menschen ausgedrückt werden kann.

I: Da jeder beschreibbare Teil der Realität an sich und für sich auch einfach „ist“, ist er Teil des Seins, und auch einer Domäne, für die es eine Ontologie geben kann.

L: Wenn man davon ausgeht, dass das philosophische Ontologiekonzept das am weitesten gefasste ist, dann ist das sprachliche bereits eine Teilmenge davon und das informatik-orientierte Konzept wiederum eine Teilmenge des sprachlichen. Kann man das so sagen?

P: Vielleicht?! Aber was ist eigentlich mit der Berechenbarkeit? Ist eine Ontologie berechenbar, oder was ist eigentlich berechenbar von diesem beschriebenen Wissen? Ist die Ontologie von Nutzen? Wenn ja, wie kann sie genutzt werden? Wird es heute schon nützlich gemacht?

I: Eine Ontologie ist ja eine logische Beschreibung, eine in sich abgeschlossene logische Beschreibung, die aller-

dings auch Widersprüche zulässt (also unterschiedliche Herleitungen) und die Frage ist: „Wozu ist diese Beschreibung nun wirklich nützlich?“ Um eine Beschreibung meines Weltwissens zu dokumentieren oder um auch Antworten zu Fragen zu berechnen?

Praktische Nutzung

P: Nützlich im praktischen Sinne ist die philosophische Ontologie bestimmt nicht. Aber diese eröffnet wiederum den Horizont, ohne den jedes menschliche Schaffen blind und deswegen auch sehr schädlich sein kann. Ich habe es so verstanden, dass eine Ontologie im Sinne des Informatikers so etwas wie eine „Vereinbarung“ ist, eine Abmachung unter Menschen in einem gewissen Tätigkeitsbereich, um die Realität oder gewisse Elemente der Realität aus gewissen Gesichtspunkten zu beschreiben.

I: Eine Ontologie ist ja meistens eine domänenbezogene Wissensbeschreibung, das sind Ausdrücke, die innerhalb einer Domäne eine eindeutige Bedeutung haben und in einem Wabensystem ingenieurmäßig zusammengeknüpft, eingesetzt bzw. kombiniert werden und das ist heutzutage eigentlich noch eine offene Frage an die Forschung:

Wie richtet man zwei oder mehrere Ontologien über dieselbe Domäne untereinander aus? Das heisst, wie werden Begriffe aus zwei oder mehreren Ontologien so gruppiert, dass sie z.B. als Synonyme, Meronyme (also die „ist-Teilvon-Relation“, bspw. „Das Fenster ist ein Teil des Hauses“) und ähnliches untereinander erkannt werden?

L: Ich glaube, das ist ein sehr wichtiger Punkt – der Begriff der Domäne. Die Ontologien der Informatik sind ja domänenorientiert.

P: Dem gegenüber gehen die philosophischen Ontologien viel weiter. Wie kann ich, wenn ich unterschiedliche Domänen miteinander verbinde, diese sozusagen „matchen“, d.h. in Einklang zueinander bringen?

L: Da kommt meiner Meinung nach die Sprache wieder ins Spiel. Wie ist dies hier, wie ist jenes dort ausgedrückt, und wie ist die semantische Relation zwischen diesen Begriffen herstellbar? Handelt es sich hier um Synonyme, dort um Hyperonyme, also: wie sind die Bedeutungsunterschiede bzw. Bedeutungsidentitäten zwischen diesen einzelnen Domänen? Und ganz eindeutig muss für den Informatiker der Begriff der Wahrheit sein: Jedes (beschriebene) Faktum muss per se wahr sein.

I: Die Domänenorientierung ist lediglich eine von Informatikern zweckmässigerweise herangezogene Strukturierung in kleinere Ontologie(-dateien), um die sonst enorme Beschreibungscomplexität zu bändigen.

Mit der Domänenorientierung kommen natürlich auch noch alle Instrumente zum „Zusammennähen“ von Ontologien (sofern sinnvoll) zum Vorschein. Wahrheit geht ja mit Herleitbarkeit einher, also mit einer Semantik, die zur Auswertung als Grundlage herangezogen wird. Der Informatiker muss sich immer fragen, wie berechenbar Fragestellungen mit Hilfe einer Ontologie sind, wollen wir, dass eine Ontologie innerhalb eines terminierenden Programms genutzt wird; gibt es Anhaltspunkte, wann ein Programm terminiert, aber die allgemeine Frage der Ter-

minierung, d.h. wann überhaupt eine Auswertung zu einem vorgegebenen Stopp kommt, ist ein unentscheidbares Problem. Entscheidbar heisst, nach endlich vielen Schritten hält jede Berechnung innerhalb dieser Ontologie an, unentscheidbar heisst, dass die Anzahl der Berechnungsschritte nicht vorhersagbar ist.

Die Ontologiewelt für sich genommen ist im Moment grundsätzlich als Modellierungsmedium zu sehen, statt als allgemeine Berechnungsgrundlage. Es gibt Spezialfälle (von Ontologien), die sehr domänenbezogen sind, also innerhalb einer Domäne, in denen die Herleitung von Fakten, Begriffen sowie die Herleitung weiterer Fakten und Begriffe mittels Inferenz möglich sind.

Inferenz ist hier als Methode zu verstehen, wie innerhalb eines formalen Systems aus Aussagen – die von Termen beschrieben sind, weitere Aussagen hergeleitet werden können.

L: Diese Spezialfälle sind übrigens gerade aufgrund ihrer Domänenspezifität die einzigen, die auch wirklich nützlich bzw. nutzbar sind.

Mächtigkeiten und Inferenzen

L: Von daher ist es in diesem Zusammenhang vielleicht auch wichtig zu wissen, dass es im Semantic Web, das ja nichts anderes als der Ontologiebereich des Web ist, drei unterschiedliche Mächtigkeiten von Ontologien gibt: Das sind OWL-LITE, OWL-DL (Description Logics) und OWL-Full. OWL= Web Ontology Language

I: Genau, das ist eine pragmatische, taxonomische Unterteilung, die im Semantic Web sehr nützlich ist: OWL ist ein gutes Beispiel von drei Modellstärken, die je eine andere Auswertungsmächtigkeit voraussetzen.

P: Was wird hier eigentlich mit „Mächtigkeit“ gemeint?

I: Der Begriff der Mächtigkeit ist direkt verbunden mit dem jeweiligen formalen System. Eine Ontologie ist ja ein formales System. Ein solches formales System besteht aus Elementen (Individuen), aus Operationen auf diesen Elementen und aus einem Herleitungsverfahren, wann eine Aussage aus Fakten (Begriffen) oder anderen Aussagen in diesem formalen System wahr (das heisst, herleitbar) ist.

Individuen werden mittels Termen beschrieben. Terme können einfach (wie z.B. einzelne Wörter) oder strukturiert sein (wie Funktionen, $f(h(x), k)$). Je strukturierter Terme werden, desto komplexer eine Berechnung/Auswertung/Herleitung innerhalb eines formalen Systems.

P: Ein Beispiel?

I: Angenommen unser formales System enthalte (der Einfachheit halber) folgende Aussagen:

A: Sokrates ist ein Grieche

R: Alle Griechen sind sterblich

A ist eine Grundaussage, während R eine Regel in der Prädikatenlogik darstellt: Für alle x, wenn x Grieche ist, dann x ist-sterblich

Welche Aussagen kann man aus A, bzw. R noch herleiten?

1. Aus A und R kann mit *modus ponens* „Sokrates ist sterblich“ hergeleitet werden.

2. Umgekehrt, aus einer Aussage „x ist nicht Sterblich“, und R kann mit *modus tollens* die Aussage „wenn jemand nicht sterblich ist, dann ist es auch nicht Sokrates“ hergeleitet werden.

Modus ponens ist dabei eine Inferenzregel, bei der aus einer bewiesenen Aussage A und einer passenden Regel ($A \rightarrow B$) die Folgerung B hergeleitet wird.

Modus tollens ist eine Inferenzregel, bei der aus einer logisch negierten Folgerung (nicht B) und einer Regel ($A \rightarrow B$) die (logisch) negierte Prämisse (nicht A) hergeleitet wird.

Die Herleitung ist also ein formaler Prozess, der innerhalb eines logischen Kalküls weitere Aussagen berechnet. Wie sieht es im Semantic Web aus? Folgendes Beispiel lässt sich in OWL-DL darstellen:

Die Regel R:

```
<owl:Class rdf:ID="Griechen">
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:hasValue>
        <Sterblichkeit rdf:ID="Sterblich"/>
      </owl:hasValue>
      <owl:onProperty>
        <owl:ObjectProperty rdf:ID="hat_Lebensspanne"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

Und die Grundaussage A:

```
<Griechen rdf:ID="Sokrates"/>
```

L: Ja gut, aber wer macht die Inferenz?

I: Im Semantic Web, oder besser: innerhalb von Ontologien werden Inferenzen von sogenannten „Reasoners“, letztlich also Softwareprogrammen durchgeführt. Diese berechnen aus RDFS-Aussagen weitere Aussagen. RDFS (RDF Schema) ist eine Empfehlung vom W3C (<http://www.w3.org/>), wie man RDF-Aussagen interpretieren soll.

L: Und wozu sind Inferenzen sinnvoll?

I: Oft wird das Wissen in Ontologien intensional repräsentiert, d.h. mit Hilfe von Abstraktionen, Regeln. Ein Gegenbeispiel bilden Grundaussagen oder Fakten, d.h. die extensionalen Wissensbeschreibungen. Intensionale Wissensrepräsentationen enthalten also nicht alle Fakten, sondern Grundaussagen und Regeln (das sind auch Aussagen), mit Hilfe derer ein Reasoner in einem Kalkül weitere Aussagen berechnet.

In der Prädikatenlogik erster Stufe sieht die Sache schon etwas anders aus: Die Variablen können Terme (Ausdrücke) enthalten, die beliebige Tiefen bzw. Längen besitzen dürfen und ein komplexes Herleitungsverfahren voraussetzen. Auswertungen in der Prädikatenlogik zweiter Stufe) sind für den allgemeinen Fall noch weniger lösbar.

L: Dies ist letztlich ein pragmatischer Aspekt. Im Semantic Web wird nicht ohne Grund unterschieden zwischen LITE, DL und FULL und deren Mächtigkeiten, wobei LITE das wichtigste ist, denn es erlaubt, bestimmte ontologische

Zusammenhänge domänenspezifisch zu beschreiben. DL scheint sich als gängige Ontologieform zu etablieren. P: Moment, das mit den OWL/Lite/DL/RDF verstehe ich nicht. Könnt ihr das mir erklären?

I: Gerne – also man beginnt die Konstruktion einer Ontologie meistens mit OWL-Lite, das ist eine Art von Ontologie im Semantic Web, die entscheidbar und exponentiell ist (exponentiell bezieht sich hierbei auf die Komplexität der Berechnung einer Problemlösung mit dieser Ontologie) gefolgt von OWL-DL, einer etwas ausdrucksstärkeren Ontologieart, basierend auf Description Logics, die ebenso entscheidbar ist (d.h. es gibt Softwarewerkzeuge, die die Berechnung in endlich vielen Schritten durchführen) aber komplexer zu berechnen (mehr als exponentiell). OWL-Full enthält die maximale Ausdruckskraft, dafür ist sie unentscheidbar. Der Begriff Unentscheidbarkeit drückt dabei aus, dass eine Berechnung (z.B. eine Herleitung) u.U. nicht terminieren kann, d.h. nicht nach endlich vielen Schritten zu einem vorgegebenem Halt kommen kann.

L: Also ist doch OWL-Full eine Ontologie im philosophischen Sinn?

I: Das würde ich nicht glauben, aber es geht in diese Richtung. Man müsste sich hier auf einen Beweis oder Nachweis berufen. Wenn eine Ontologieform nicht maschinell berechenbar ist, wie ja bekanntlich OWL-Full, muss dies nicht bedeuten, dass man mit dieser Ontologieform philosophische Ontologien darstellen kann. Die nächste Einschränkung kommt aus der Endlichkeit einer Ontologiebeschreibung (man muss endlich viele Zeichen einsetzen) und aus der Unendlichkeit der Wahrheiten (z.B. der Menge aller aussagenlogischen Formeln bzw. Ausdrücke). OWL-Full bietet zwar eine höhere Modellierungsmächtigkeit an, unterliegt aber der Endlichkeit seiner Beschreibung.

RDF-Tripel und URI

P: Setzt man sich näher mit dem Semantic Web auseinander, fallen oft die Begriffe URI und RDF. Wie hat man sich das bzw. die Zusammenhänge zwischen diesen beiden Konzepten vorzustellen?

I: Ein URI, also ein Universal Resource Identifier wird benutzt, um eine Ressource über eine Adresse (URL=Uniform Resource Location) im Web darzustellen. Mit einem URI und einem Browser oder ähnlichen Dienst kann man die dargestellte oder identifizierte Ressource abrufen. Ein RDF ist ein Format (letztlich ein Dialekt von XML) mit dem Entitäten beschrieben werden.

L: Es geht auch darum, dass jede Information, die kommuniziert wird, quasi einen festen Ort hat und abrufbar ist. Und dieses eindeutig Abrufbare, das bewerkstelligt der „URI“, so wie seit jeher Webseiten ihren (mehr oder weniger) festen Platz hatten, so bekommen jetzt eben auch die einzelnen Entitäten, also alles, was in RDF-Tripeln erfasst werden kann, ihren festen Platz, ihre feste Adresse. Diese Entitäten werden dann in RDF-Tripeln miteinander verknüpft.

P: RDF ist also der Grundbestandteil einer Ontologie im Semantic Web, oder?



Abbildung 1: Beispiel einer Kante im Semantic Web Graph (ein Tripel); Quelle: Dittrich KR, Semantic Web. Seminar Datenbanktechnologie fürs Web. Institut für Informatik der Universität Zürich. Sommersemester 2001. [Internet], c2001. Erreichbar unter: <http://www.inf.uni-konstanz.de/dbis/teaching/ss01/data-on-the-web/local/semanticweb.pdf>

I: RDF ist ein Format in dem sich ganz allgemein Informationen beschreiben lassen. Es definiert die Art und Weise, wie eine Entität als Ressource im Webuniversum dargestellt werden kann. Eine Ontologie an sich wird als komplexes Netz oder baumartiges Objekt (Graph) mit einem Konzept an jedem Knoten dargestellt.

Und die Kanten dieses Graphs stellen Prädikate dar, die entweder ein Konzept mit einem anderen Konzept verbinden oder aber es werden über Assoziationen Konnotationen zu anderen Entitäten erzeugt, je nachdem welche Granularität man wählt. Also ist eine Ontologie hier ein komplexes Objekt, das in RDF aber auch in einem anderen Format vorliegen kann und von einem Computer verarbeitet wird.

Folgendes Bild erklärt wie ein einfacher Satz des Semantic Web, ein Tripel, eine Kante im Graph darstellt (vgl. Abbildung 1):

Die Aussage hinter der Kante kann wie folgt interpretiert werden:

«Das Konzept 10034334 ist eine Verallgemeinerung des Konzepts 10034335».

Eine Ressource, das Prädikat (Eigenschaft) und der beschreibende Inhalt (Wert) dieser Eigenschaft, bilden im RDF-Kontext als Tripel eine Einheit, bestehend aus **Subjekt, Prädikat** und **Objekt** – sie bilden eine RDF-Aussage. Eine RDF-Aussage kann als gerichteter Graph dargestellt werden: Subjekte und Objekte sind Knoten und jedem Prädikat (Eigenschaft) entspricht eine Kante, die vom Subjekt zum Objekt weist. Die Gesamtheit aller Tripel dieser Art bildet einen Semantic Web Graph.

L: Ich möchte noch dazu sagen, dass das Besondere bei RDF ist, dass es über das normale XML hinausgeht, in dem Sinne, dass es eine semantische Ebene enthält, also eine semantische Information kommuniziert. Hingegen wird über XML letztlich nur rein syntaktische Information kommuniziert.

I: Einverstanden! In RDF sind viel komplexere Beschreibungen enthalten, durch die – auf einfache Weise – beliebig komplexe Beschreibungen anhand von Vokabularen (kontrollierte Vokabulare, unterschiedliche Thesauri) formalisiert sind.

So kommen jene ulkigen Ausdrücke zustande, die einen Doppelpunkt enthalten: Links vor dem Doppelpunkt, quasi als Präfix, wird der Worttyp aus dem jeweiligen Vo-

kabular angekündigt, und rechts nach dem Doppelpunkt stehen der Name (auch Literal genannt) und das Prädikat. Auch das Prädikat stammt aus demselben Vokabular wie der Name. Das, was vor dem Doppelpunkt steht, das Präfix also, bildet dann einen „Namespace“, einen Namensraum. An dieser Stelle wird also immer ein Name innerhalb seines Vokabulars – alias Namespace – oder der Name eines Graphen oder weitere formale Spezifikation erwähnt, und dieser Name wird irgendwo über seinen entsprechenden URI dargestellt bzw. spezifiziert.

Beispiel:

Das bekannte Tripel (siehe Abbildung 1) aus dem TheSoz Graph sieht in einem menschlich lesbaren Format so aus: thesoz:concept/10034335 skos:broaden thesoz:concept/10034334

Im Tripel werden zwei Namensräume (Kontexte) verwendet:

thesoz: steht hier als Name des Graphs <<http://lod.gesis.org/thesoz/>>

skos:Verweis zur SKOS Spezifikation <<http://www.w3.org/2009/08/skos-reference/skos.html>>

Skos steht dabei für Simple Knowledge Organisation System, einen Standard zur Gestaltung von Thesauri im Semantic Web.

Obiges Tripel sagt also formal noch aus, dass das TheSoz-Konzept Nr. 10034334 allgemeiner ist als das TheSoz-Konzept Nr. 10034335 nach dem Skos-Prädikat „broaden“.

In RDF würde ein solches Tripel wie folgt aussehen:

```
<rdf:Description rdf:about="concept/10034335">
  <skos:broaden rdf:resource="concept/10034334"/>
</rdf:Description>
```

L: Man sollte, wenn man diese verschachtelten formalen Ausdrücke liest, nicht vergessen, dass es beim Semantic Web und damit zusammenhängend RDF und Ontologien letztlich darum geht, Maschinen Wissen verständlich zu machen, also eine Wissensrepräsentation für Maschinen zu erstellen. Das muss vorweg gestellt werden. Es geht nicht nur darum, Wissen digital in Maschinen darzustellen, sondern dass Maschinen selbst konstruktiv mit diesem Wissen umgehen.

P: Also mit einem URI können Maschinen auch was anfangen.

L: Ja, URI ist quasi eine Adresse. So wie eine Signatur beim Buch, sie ordnet dem Wissen nur seinen Platz zu. Also eine Beschreibung des Orts, wo ich das Buch abholen kann.

P: Enthält ein URI nicht besondere Informationen über eine Entität selbst oder ist ein URI nur ein Wegweiser?

I: Ein URI muss nicht unbedingt ein Objekt direkt darstellen, aber es ist eine gute Praxis z.B. Bereiche in einer URL zu kodieren. Normalerweise besteht ein URI aus unterschiedlichen Segmenten und diese könnten sinnvollerweise bereits als Wegweiser innerhalb einer Taxonomie dienen.

Einige Beispiele dazu:

1. Die URI <http://lod.gesis.org/pubby/page/thesoz/ext/CompoundEquivalence> sagt aus, dass <http://lod.gesis.org/pubby/page/thesoz/term/10034303-en> des Bereichs (Semantic Web Graph) <http://lod.gesis.org/thesoz/> die (englische Definition der) CompoundEquivalence extern angelegt ist.
2. <http://lod.gesis.org/pubby/page/thesoz/concept/10034304> sagt aus, dass 10034304 ein Konzept innerhalb des Graphs <http://lod.gesis.org/thesoz/> ist.
3. <http://lod.gesis.org/pubby/page/thesoz/term/10034303-en> beschreibt den Sachverhalt, dass 10034303-en ein englischer Term im Graph <http://lod.gesis.org/thesoz/> ist.

P: Wie kann dann eine Maschine an Hand eines URI eine Entität lesen? Sie muss doch Tripel abrufen bzw. lesen, um die semantischen Informationen abzurufen ...

L: Richtig! Die Tripel sind entscheidend, weil durch die Tripel eine Relation ausgedrückt wird. Das einfachste ist, ein Tripel als einen einfachen Satz (Subjekt, Prädikat, Objekt) aufzufassen. Man kann sich ja auch vorstellen, dass selbst dieses Gespräch aus lauter Tripeln (Subjekt, Prädikat, Objekt) bestehen würde. Die syntaktisch-semantische Repräsentation der Ontologien im Web geht über diese Strukturen nicht hinaus, d.h. wir können unendlich ...

I: ... vielleicht besser „beliebig“ ...

L: einverstanden, also beliebig viele Sätze produzieren, die diese Struktur haben und das entscheidende ist, dass die Zusammenhänge über Verknüpfungen der Tripel stattfinden.

P: Also könnte man sagen, dass diese URI „Tore“ sind, die Maschinen erlauben, Zugang zu den Tripeln zu haben?

I: Die sind Tor und Tür zu semantischen Informationen.

L: Da fällt mir wieder Wittgenstein ein und sein Vergleich der Sprache mit einer Stadt: „Ein Gewinkel von Gässchen und Plätzen, alten und neuen Häusern mit Zubauten aus verschiedenen Zeiten: und dies umgeben von einer Menge Vororte mit geraden und regelmäßigen Straßen und mit einförmigen Häusern.“ Aus der Perspektive der Wittgensteinianer gesehen ist es einfach so, dass je nachdem, welche Ebene man aufschliesst, hinter einem URI(-Tor) eine ganze Stadt oder nur Strassenzüge stehen, am Ende aber nur die Tore und Türen zu den Häusern. Das wären dann die Tripel.

I: Ich bin nicht sicher, ob das gilt, dass hinter jeder URI genau ein Tripel steht...

L: Nein, hinter jeder URI steht ja ein Element eines Tripels oder?

I: Auch nicht. Es ist mir nicht bekannt, ob das eine Forderung ist. Hinter einem URI könnte auch eine ganze Ontologie sein. Ein URI ist grundsätzlich eine universelle Adresse.

L: Ok, dann ist es – wie Du gesagt hast – Tor und Tür. Wenn wir dann noch einmal die Wittgenstein-Metapher bemühen, dann ist es wirklich ein Tor zu einer Stadt, in dieser Stadt gibt es Viertel, darin Strassen und in allen Häusern ... irgendwo hört es auf, ich habe eine atomare Ebene, aber ich kann diese noch einmal verschachteln oder ineinander packen.

I: Ja, das stimmt und ich möchte dazu ergänzen, wenn ich ein anderes Bild nehmen darf, ein Tripel kann auch Adresse sein zu einem sehr komplexen System, das in sich selbst referenziell sein kann, aber das brauchen wir hier nicht weiter zu besprechen.

Wir haben also mehrere Repräsentationsebenen und wir können durch eine URI wie gesagt sowohl ein einzelnes Grundobjekt als Tripel darstellen oder auch eine Ansammlung von solchen Objekten, wie etwa eine Enzyklopädie, bekanntlich Ontologie.

P: Also lasst mich noch ein Beispiel machen, um sicher zu sein, dass ich alles verstanden habe. Nehmen wir an, wir haben einen Datensatz, der ein Buch beschreibt, dann kann sich also ein URI sowohl auf den ganzen Datensatz als auch auf die einzelnen Elemente des Datensatzes, als auch auf die Teile der Elemente beziehen. Es ist nicht vorgeschrieben, auf welche Repräsentationsebene sich die URIs beziehen sollen.

I: Korrekt.

P: Dann bin ich ein Stückchen weiter gekommen.

L: Man kann also sagen: die grundsätzlichen Elemente sind die Tripel, die aus Subjekt/Prädikat/Objekt bestehen ...

I: Ja, man kann besser sagen: die atomaren Elemente.

L: Das heisst aber nicht, dass man die atomaren Elemente nicht weiter teilen kann. Nehmen wir dafür ein Beispiel: „Wittgenstein ist der Autor der Philosophischen Untersuchungen“, dann gilt als Subjekt Wittgenstein, „ist der Autor der“ als Prädikat und das Objekt sind die „Philosophischen Untersuchungen“ und wenn ich jetzt genauer hinschauen könnte ich „Philosophische Untersuchungen“ noch genauer in zwei Teile aufsplitten und auch das Prädikat in mehrere Teile, letztlich die Lexeme, also die einzelnen Wörter. Letzteres ist aber für die Tripel an sich nicht von Interesse.

I: Dieser Beispielsatz mit den dazugehörigen NameSpaces sieht dann wie folgt aus:

person:Wittgenstein

verben:ist-Autor-von

Werke:Philosophische Untersuchungen

Der erste Teil des Paares ist ein Namespace, ein Begriff, der den Kontext darstellt und der zweite Teil ist das Substantiv und das ist das Wichtigste. Das gilt für jede Einheit dann so.

P: Ja, aber Namespace, was heißt das eigentlich?

L: In diesem Beispiel wenn ich richtig verstehe, sind die Namespaces: person, verben, werke.

I: Korrekt.

L: Das heisst, jedes Atom besteht zunächst aus zwei Hälften: einem positiven und einem negativen Teil. Im ersten Fall wäre quasi der negative Teil Person, der positive Teil Wittgenstein. Das ist die Grundstruktur oder?

I: Ja es geht um ein Tripel von Paaren im Grunde genommen.

Aber um noch auf die eben gestellte Frage zurückzugreifen: wie der Name sagt, ist „namespace“ ein Raum, in dem man Bedeutungen festlegt. Im Grunde ist es ein Vokabular, dieses Vokabular hat einen Namen und hier habe ich beispielsweise Person genommen. Im Semantic Web ist FOAF (d.h. friend of a friend) ein bekanntes Vokabular für Personen, also im Semantic Web würde ich in diesem Fall FOAF:Wittgenstein schreiben. Ist es für Dich so verständlich?

P: Ja, ich denke schon. Könnte man dann z.B. sagen, dass im PubMed die Feldernamen Autor, Titel, Abstract usw. Namespaces sind?

I: Sie sind eher Elemente in einem bibliographischen System, letzteres kann ein Namespace sein – so wie es in der Dublin Core Metadata Initiative vorgesehen ist. „Armut“ könnte in einem Kontext Person sein in einem anderen Verb, je nachdem was als Präfix davor steht.

P: Ok, alles klar.

Das Problem der Semantik

I: Ganz schön komplex... und eigentlich sehe ich hier die Semantik noch nicht. Ich sehe hier den Versuch, den Maschinen möglichst genaue Beschreibungen (formalisiert) zur Verfügung zu stellen oder vorzubereiten. Es geht also um eine Kontextschaffung syntaktischer Art.

P: Semantik heißt, wie man einen Satz interpretiert. Das hängt im Grunde genommen vom Betrachter ab.

L: Klar. Es hängt auch natürlich mit der Referenz zusammen, mit der Relation zwischen der Beschreibung und dem Objekt oder dem Konzept das dahinter steht. Also quasi dem Link, d.h. der Verbindung zur Welt an sich.

I: Die Verbindung zur schriftlich fest erfassbaren Welt.

L: Repräsentierbaren Welt.

I: Richtig, zur schriftlich repräsentierbaren Welt, die im Grunde ganz anders und viel reichhaltiger ist als unsere eigene. Auch wenn es banal klingt, aber ich finde es immer spannend, diese beiden Welten auseinander zu halten. Wir sprechen dabei letztlich von einer Maschinensemantik oder von darstellbarer Semantik.

L: Semantik, alle reden über Semantik, aber: wie entsteht eigentlich Semantik? Diese Frage ist meiner Meinung nach noch gar nicht gelöst und vielleicht auch in letzter Instanz unlösbar.

Tim Berners Lee hat das World Wide Web erfunden und das war natürlich eine große Sache. Das erste was er in der Folge als besonders interessant angesehen hat, war das Semantic Web. Für ihn war die Frage der Semantik also sehr wichtig. Gehen wir ein paar Jahre zurück, als

Lotfi Zadeh die Unscharfe Logik, die Fuzzy Logic erfunden hat. Was forderte er danach: Fuzzy Semantics. Kürzlich hörte ich einen informationswissenschaftlichen Vortrag von Alan N. Shapiro und wofür plädierte er wohl: für Semantic Information Science. Semantik ist wohl ein Thema, das für Leute, die sich viel ausdenken und gute Ideen haben, oft gleich an zweiter Stelle kommt, ohne dass diese Leute aber genau wissen, wie komplex das Problem der Semantik überhaupt ist.

P: Ich würde Euch aber gerne, bevor sich unser Gespräch dem Ende zuneigt, noch auf einen philosophischen Ausflug mitnehmen. Unser Ziel wird das Verhältnis zwischen der Realität und der Semantik sein. Schopenhauer hätte von Wille und Vorstellung geredet.

Also zunächst die Realität. Nehmen wir irgendetwas. Ein Berg, ein Baum, das Meer, es spielt keine Rolle. Also nehmen wir den Berg als Beispiel. Einen Berg gibt es von uns Menschen abgesehen. Wenn wir Menschen aber klein wie Elektronen oder groß wie der Abstand zwischen Erde und Sonne wären, würde der Berg je nachdem ganz anders aussehen. Also der Berg, wie wir Menschen ihn kennen, ist nur eine seiner Dimensionen. Aber alle Dimensionen sind real. Der Berg – egal in welcher Dimension – gibt es, ist nicht Ergebnis einer Idee, wie manche es gerne hätten, um die volle Macht über die Dinge zu haben. Den Berg gibt es von uns abgesehen. Den gab es schon lange, bevor es Leben auf der Erde gab und wird es höchst wahrscheinlich auch viel länger geben als das Leben. Das war die kurze Klärung zur Realität.

Nun, ist die Semantik dran. Die Semantik ist eine Darstellung der Realität und ist im Gegensatz zur Realität eine menschliche Sache. Ohne Mensch gäbe es sie nicht. Der Mensch neigt dazu, darzustellen. Die Darstellung erfolgt durch Zeichen, seien es Farben, Striche, Klänge, Worte oder sonst was. Mit der Darstellung äußert der Mensch etwas. Und wenn auf der anderen Seite ein anderer Mensch die Darstellung rezipiert, entsteht Kommunikation.

Ferdinand de Saussure, der Vater der strukturalistischen Semantik hat gemeint, Worte würden aus einer Form und aus einem Inhalt bestehen. Also Worte würden in sich eine Bedeutung tragen, sie seien quasi ein Vehikel von Kenntnis unter den Menschen.

Das gilt aber meiner Meinung nach nicht für jede Kenntnis. Je komplexer die Kenntnisse sind, desto untauglicher wird die Semantik, desto weniger sind die Zeichen Vehikel von Kenntnis, sondern eher Wegweiser zur Kenntnis. Wenn das Objekt der Kenntnis komplexer wird, spielt die Erfahrung eine größere Rolle und Menschen verstehen sich nur, wenn sie eine ähnliche Erfahrung gemacht haben. Je komplexer die Kenntnis ist, die man vermitteln will, desto weniger geht es um das, was man sagt und desto mehr geht es stattdessen um das, was man meint. Je komplexer die Kenntnis ist, desto weniger geht es um die Bedeutung der Worte, sondern eher um das, was sie in dem Kontext meinen, was durch Ontologien nicht ausdrückbar ist.

Vom Anfang und vom Ende

P: Plato, der Vater der Philosophie, hat seine eigentliche Lehre gar nicht beschrieben. Seine Schriften sind eine Darstellung seiner einfachsten Gedanken. Seine eigentliche Lehre war nur Objekt seiner Reden. Seine eigentliche Lehre wurde nur den Schülern erschlossen, die mit der Erfahrung schon weit genug waren, um zu verstehen, nicht das, was er sagte, sondern das, was er meinte. Also ist das Semantic Web wie jede andere Form der menschlichen Darstellung Ausdruck von Kenntnis nur bis zu einem gewissen Komplexitätsgrad. Natürlich will ich damit den Wert des Semantic Web nicht vermindern. Dagegen denke ich, dass das heute die mächtigste Kommunikationsform ist, weil die Informationen meistens im Internet präsent sind und verbreitet werden. Ich möchte nur auf die Grenzen des Semantic Web aufmerksam machen, denn die Menschen neigen oft zur Allmächtigkeit. Es gibt nichts Schlimmeres als Menschen, die die Grenzen ihrer Dimension nicht erkennen.

L: Ich finde, dass die „Dimension“ hier ein sehr interessanter Begriff ist! Ich glaube, dass die unterschiedliche Sichtweise dessen, was eine Ontologie ist, sehr viel mit diesen Dimensionen zu tun hat, und das, was der Darstellung vorausgeht, für die philosophische Ontologie grundsätzlich ist. Dann kommt erst der Aspekt der Darstellung, das ist der Aspekt, der in der Sprachwissenschaft und ihrer Königsdisziplin, der Semantik, sehr zentral ist.

In der Informatik hat die Ontologie sehr viel mit Operabilität dessen zu tun und vor allen Dingen mit der Kommu-

nikation mit und zwischen den Maschinen. Wir dürfen nicht vergessen, dass es beim Semantic Web um maschinelle Operabilität geht, mit dem Ziel, die Kommunikation zwischen Maschinen zu erleichtern.

Zu allerletzt sollten wir dabei nicht vergessen, dass wir mit dem Semantic Web auch das Zeitalter der Maschinen vorbereiten und noch nicht wissen, ob wir damit nicht auch das Ende des Zeitalters der Menschen einläuten.

Korrespondenzadresse:

Maurizio Grilli
Claudiana, Landesfachhochschule für Gesundheitsberufe,
Lorenz-Böhler-Strasse 13, 39100 Bozen, Italien
Maurizio.Grilli@claudiana.bz.it

Bitte zitieren als

Grilli M, Ricci F, Schneider R. Trialogo: De Ontologia. *GMS Med Bibl Inf*. 2013;13(1-2):Doc15.

DOI: 10.3205/mbi000279, URN: urn:nbn:de:0183-mbi0002798

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/mbi/2013-13/mbi000279.shtml>

Veröffentlicht: 15.11.2013

Copyright

©2013 Grilli et al. Dieser Artikel ist ein Open Access-Artikel und steht unter den Creative Commons Lizenzbedingungen (<http://creativecommons.org/licenses/by-nc-nd/3.0/deed.de>). Er darf vervielfältigt, verbreitet und öffentlich zugänglich gemacht werden, vorausgesetzt dass Autor und Quelle genannt werden.