

„Semantische Techniken stehen kurz vor dem Durchbruch“

Skepsis gegenüber dem Semantischen Web ist weit verbreitet. Wir fragten Professor Frank van Harmelen, einen der führenden Köpfe auf diesem Gebiet, nach den Chancen und dem Praxisnutzen der semantischen Techniken.

c't: Forschung und Industrie haben sich mit Lust auf das Web 2.0 gestürzt, das Mitmach-Web droht, dem Semantischen Web das Wasser abzugraben – nicht zuletzt aufgrund vieler großer Versprechungen, die Letzteres nicht einhalten konnte. Wie sehen Sie die Koexistenz dieser beiden Welten?

Frank van Harmelen: Die Frage fußt auf der falschen Annahme, dass das Semantische Web in Ungnade fiel und seine Versprechungen nicht einhalten konnte. Schauen wir uns stattdessen lieber einige Fakten an: Die SemTech, eine auf die Belange der Industrie ausgerichtete Konferenz in San Jose, konnte seit 2005 ihre Besucherzahlen mehr als verdoppeln, auf über 700 Teilnehmer. Die ESTC in Wien, das europäische Gegenstück, kam auf mehr als 200 Besucher, drei Viertel da-

von Firmen. Also ist entweder die Frage falsch oder viele hundert Geschäftsleute irren sich – Ihre Entscheidung.

Ganz im Gegenteil: Semantische Techniken stehen kurz vor einem Durchbruch im industriellen Umfeld. Das sieht auch ein Gartner-Report vom Mai dieses Jahres so. Der Branchendienst, der nicht als Freund kurzlebiger Hypes bekannt ist, spricht von „außerordentlichen Fortschritten in der Sichtbarkeit und Verwertung von Informationen“ in den kommenden zehn Jahren dank semantischer Strukturen in Webdokumenten. Glücklicherweise erklärt die Analystenfirma nicht vorschnell das Scheitern: „Die große Vision des Semantischen Webs wird sich in vielen evolutionären Schritten manifestieren und kleine Initiativen sind oft die besten Startpunkte.“

Zurück zum Kern der Frage: Die Wissenschaft ist sich weitgehend einig, dass Web 2.0 und Semantisches Web (oder Web 3.0) komplementär sind. Das Web 2.0 weist eine niedrige Einstiegsschwelle auf, man kommt aber auch schnell ans Ende der Fahnenstange – viel lässt sich mit Folksonomies nun mal nicht anfangen. Das Web 3.0

setzt anfangs eine höhere Hürde, reicht aber auch wesentlich höher hinaus.

Der zitierte Gartner-Report empfiehlt die Verbindung beider Welten und sagt ein langsames Wachstum voraus, das vom derzeitigen Web über leichte, einfach nutzbare Semantik-Techniken zum kostenintensiveren, aber auch ertragreicheren Web 3.0 führt.

c't: Der Semantic-Web-Initiative wird oftmals vorgeworfen, dass sie dieselben Problemen aufgreift, für welche die künstliche Intelligenz (KI) bereits vor 30 Jahren keine adäquaten Antworten parat hielt. Wissensrepräsentation und induktive Logik gelten als Eckpfeiler dieser Bestrebungen und brachten Systeme wie KL-ONE, Cyc oder die konzeptuellen Graphen von John Sowa hervor. Der Erfolg blieb weitgehend aus. Was unterscheidet das Semantische Web von diesen fruchtlosen Unterfangen?

van Harmelen: In der Tat gibt es das weit verbreitete Missverständnis, dass das Semantische Web alter Wein in neuen Schläuchen sei. Obwohl es mit der KI einige Werkzeuge gemein hat – eben die Ontologien und das logische Schließen –,

sind die Zielsetzungen völlig unterschiedlich. Das Semantische Web gibt sich bescheiden. Seine Bestimmung besteht *nicht* darin, die große, allumfassende und Internet-basierende Intelligenz zu kreieren. Vielmehr geht es um Interoperabilität zwischen Datenbeständen, die über das Web zugänglich sind.

Tim Berners-Lee widmete im Juli letzten Jahres diesem Thema einen eigenen Vortrag [9], worin er festhält: Das Semantische Web ist nicht KI und KI ist nicht Semantisches Web; die KI ist ein Forschungsfeld, das Semantische Web ein Projekt; das Semantische Web steht in der Schuld der KI, da es einige ihrer Werkzeuge benutzt. Daher sollte das Semantische Web eine exzellente Spielwiese für die KI sein.

Berners-Lees Vortrag räumt zudem mit einigen weiteren Mythen auf, etwa dass das Semantische Web lediglich mit von Hand annotierten Textdokumenten arbeite oder dass es eine einzelne umfassende, von allen genutzte Ontologie erfordere.

c't: Wie steht es um Werkzeuge zum automatisierten Lernen von Ontologien, von Beziehungen



zwischen Entitäten und so weiter – also Natural Language Processing (NLP), Text Mining und statistisch gestützte Extraktion von Wissen? Was halten Sie von diesen Techniken im Hinblick auf den emsigen Bau am Semantischen Web?

van Harmelen: Mein Standpunkt zu automatisch erstellten Ontologien und Klassifizierungen ist: Schön, wenn es funktioniert. Rein manuelles Modellieren von Ontologien ist natürlich sehr aufwendig und kostet Zeit. Daher gehe ich davon aus, dass die von Ihnen genannten Techniken in Zukunft eine größere Rolle spielen werden. Ich bin recht zuversichtlich, dass diese Techniken einen zunehmend

Frank van Harmelen, Professor für Informatik in Amsterdam, begann vor zwei Jahrzehnten mit der Arbeit an KI-Projekten. Er wirkte seit den Anfängen des Semantischen Webs daran mit und ist einer der Koautoren der Web Ontology Language (OWL).

wertvollen Beitrag leisten werden.

c't: Alle Meilensteine der Technik zeichnen sich durch einen großen Wurf aus, eine „Killer-Anwendung“. Was könnte die Killer-Anwendung des Semantischen Webs sein – und wird es überhaupt eine geben?

van Harmelen: Ich finde die ewige Frage nach einer „Killer-Anwendung“ ein wenig naiv. Nehmen wir XML: Wenn wir uns einigen können, dass das weit verbreitete XML eine wichtige technische Innovation war – was war dann seine Killer-Anwendung? Gab es eine einzelne? Nein, das allgegenwärtige XML arbeitet sozusagen unter der Haube.

Semantic-Web-Techniken sind hauptsächlich Infrastruktur und somit den Augen des Anwenders entzogen. Er wird lediglich bemerken, dass Websites personalisierter werden (weil sich dank semantischer Techniken persönliche Profildaten mit Websites austauschen lassen), dass Suchmaschinen die Ergebnisse besser gruppieren (weil Ontologien bei der Klassifikation helfen) oder dass die Desktop-Suche plötzlich die Namen von Autoren lokaler Dokumente mit E-Mail-Adressen im Adressbuch verknüpft (weil Semantik diese Datenformate interoperabel macht). Doch keine dieser Anwendungen wird das Schildchen „Semantisches Web“ tragen.

Das Semantische Web ist wie die Nikasil-Beschichtung im Zylinder eines Automotors: Sehr wenige Fahrer wissen davon, aber alle bemerken den reduzierten Benzinverbrauch, die gesteigerte Höchstgeschwindigkeit und die längere Lebensdauer des Motors. Das Semantische Web ist das Nikasil der nächsten Generation von menschenfreundlichen Computer-Anwendungen – Anwendungen, die bereits jetzt entwickelt werden.