

Tagungsbericht

Eva Milgotin*

SEMANTiCS 2019 – oder die Antwort darauf, wie digital endlich real wird

<https://doi.org/10.1515/iwp-2019-2045>



Abbildung 1: Prof. Dr. Harald Sack (Bildmitte) spricht mit Prof. Dr. York Sure-Vetter (rechts). (Foto: Johannes Hüsich)

Der Veranstaltungsort der SEMANTiCS Konferenz 2019, Karlsruhe, mag irreführend klingen. Denn mit Ruhe haben weder das schöne Städtchen als Gastgeber noch die SEMANTiCS als Veranstaltung etwas zu tun. Spätestens seit 1984 hat Karlsruhe in der Tech-Community einen besonderen Klang, denn die erste E-Mail überhaupt wurde hier empfangen.¹ Dieser Ort ist also besonders prädestiniert, um die führenden Spezialistinnen und Spezialisten aus Forschung und Industrie zum gegenseitigen Wissenstransfer und Gedankenaustausch zu spannenden Themen wie die Erstellung von Wissensgraphen, Implementierung von Ontologien und Künstlicher Intelligenz in die Karlsruher Gartenhalle einzuladen.

¹ Siehe dazu bspw.: <https://de.wikipedia.org/wiki/E-Mail> [16.9.2019].

*Kontaktperson: Eva Milgotin, FH St. Pölten, Department Medien & Wirtschaft, Matthias Corvinus-Straße 15, 3100 St. Pölten, E-Mail: eva.milgotin@gmail.com

Seit 2005 öffnet die SEMANTiCS Tür und Tor für Interessierte aus aller Welt, um mit einem viertägigen Vortragsprogramm, bei dem Wirtschaft und Forschung gleichermaßen zu Wort kommen, aktuelle Problemstellungen aus praktischen Anwendungsbereichen mit dazu passenden innovativen Lösungsansätzen zu verbinden. So sollen Synergien entstehen, die den technologischen Fortschritt vorantreiben und einen tatsächlichen Mehrwert für alle Teilnehmenden darstellen. Auch 2019 konnte das SEMANTiCS-Team den 426 Besucherinnen und Besuchern ein aufregendes Programm bieten, das gespickt war mit Themen wie Computerlinguistik, Informationsmanagement, Legal Tech, Distributed Ledger Technology, Ontologie Management, Linked Data und Digital Humanities.

Michael J. Sullivan, Experte in Sachen Softwarearchitektur und Cloudcomputing bei Oracle, eröffnete die SEMANTiCS 2019 mit einer Demonstration, wie sich die Zugänge zur heutigen datengetriebenen Gesellschaft verändert haben. Seine Veranschaulichung machte deutlich, dass der Bedarf nach Semantik-basierten Problemlösungen in Sachen Datenanalyse und -verarbeitung schon seit den frühen 80er Jahren besteht und unverändert anhält. Getaggt, verlinkt, beschrieben, sortiert und interoperabel sollen Daten sein, um ihr volles Potenzial zu entfalten. Um gerade dieses Potenzial hat sich in den letzten zwei Jahrzehnten eine Vielzahl an Experten und Expertinnen aus diversen Bereichen wie etwa der Linguistik, der Geschichtswissenschaften, der Geschäftsentwicklung, der Medienindustrie oder der Software Entwicklung gebildet. Diese „Semantic Web Community“ hat es sich zur Aufgabe gemacht, ein Netz aus öffentlich zugänglichen Daten zu erstellen, das seither nicht nur an Größe, sondern auch an Bedeutung kontinuierlich zunimmt.



Abbildung 2: Der gut gefüllte Vortragssaal der SEMANTICS 2019. (Foto: Johannes Hüsich)



Abbildung 3: Michael J. Sullivan, Oracle. (Foto: Johannes Hüsich)

Das semantische Web oder auch die Linked Open Data Cloud offeriert jedoch nicht nur eine gigantische Menge an miteinander verbundenen Daten, die zur eigenen Auslese verwendet werden können. Viel mehr noch beinhaltet sie Datensätze, die durch das Versehen mit Metadaten zueinander in Kontext gebracht werden können. Das heißt, kurz gesagt, dass Maschinen ihre Suche und Analyse nicht länger auf reiner Stichwortbasis durchführen, sondern auf Informationen zurückgreifen können, die mit-

einander in Relation gebracht werden und mit Sinn versehen sind. Mittels URI's, so genannte Uniform Resource Identifier, werden zum Beispiel Doppeldeutigkeiten und multiple Verwendungszwecke von Worten für Maschinen erkennbar. Siri ist so in der Lage, den Unterschied zwischen einer Sitzbank oder einer Bank im Sinne eines Geldinstitutes zu erkennen, je nachdem, in welchem Kontext das Wort „Bank“ zur Anwendung kommt. Darüber hinaus können auf diese Weise vermeintlich voneinander unabhängige Datenmengen aus diversen Datenquellen miteinander verknüpft und interoperable genutzt werden.

Verknüpfte Daten bringen Mehrwert

Diese Form der Datenverknüpfung steht jedoch nicht nur der Öffentlichkeit im World Wide Web zur Verfügung, sondern kann in jedem Unternehmen oder Institut implementiert werden. „Durchschnittlich werden weniger als die Hälfte der strukturierten Daten eines Unternehmens zur Entscheidungsfindung herangezogen und weniger als ein Prozent der unstrukturierten Daten wird analysiert oder verwendet“, zitiert Sullivan aus einer Harvard Studie aus dem Jahre 2017. Damit verdeutlicht er das enorme Ausmaß an Potenzial, dass auf Grund mangelnder oder unausgereifter technologischer Unterstützung für Unternehmen ungenutzt bleibt. Besonders in einer Zeit, die an den Goldrausch des 19. Jahrhunderts erinnert, in der Daten wahrhaftig geschürft sowie in rohen Mengen kostenintensiv

gelagert werden, kann dieses ungenutzte Potenzial Wettbewerbsvorteile kosten.

Sogar einen Schritt weiter gehen Andy Boyd und Brendan Nielsen, Mitarbeiter des Knowledge Management Teams der Royal Dutch Shell. Sie behaupten in ihrer Keynote zum Thema „High-grading Business Decisions through Semantic Technology“: *„Semantik kann helfen, Leben zu retten!“*. Das fragmentierte Businessmodell von Shell bedingt auch die Diversität der Unternehmensdaten. Daher setzt Shell auf ein ausgereiftes Wissensmanagement, nicht nur um die Sicherheit der Kundinnen und Kunden sowie Mitarbeiterinnen und Mitarbeiter zu gewährleisten, sondern auch, um einen zielgerichteten, effizienten und automatisierten Informationsfluss innerhalb des Unternehmens sicherzustellen.

Und Wissen ist Macht

Aber nicht nur Unternehmen profitieren von semantischen Technologien, erklärt Michel Dumontier in seinen Ausführungen zum Thema „Accelerating Biomedical Discovery with an Internet of FAIR Data and Services“. In seiner Rolle als Professor an der Universität Maastrich zählt Dumontier zu den führenden Spezialisten in Sachen Data Science. Er verweist auf die Tragweite semantischer Technologien und vor allem des Open Linked Data Web in Sachen Forschung. Der Weg zur Innovation ist für ihn zukünftig mit FAIR Data gepflastert. Findable, Accessible, Interoperable und Reusable müssen Daten aufbereitet sein, um valide Forschungsergebnisse zu Tage zu bringen und diverse Forschungsarbeiten vielfältiger Untersuchungsgegenstände miteinander zu verbinden. Darüber hinaus vereinfacht die faire Aufbereitung von Datensätzen das Nachvollziehen von wissenschaftlichen Analysen und schließt die Lücke der individuellen Interpretation. Die Vorteile des semantischen Webs sind für ihn in fünf Aspekten begründet:

1. *„Es ist einfacher, denselben Datensatz für unterschiedliche Zwecke wiederzuverwenden.“*
2. *Es ist einfacher für andere, meine Daten zu finden, zu verwerten, zu begutachten und zu verstehen, was ich im Gegenzug dafür erwarte.*
3. *Es ist einfacher für andere, meine Arbeit nachzuvollziehen und zu überprüfen.*
4. *Es stellt sicher, dass meine Daten auch in Zukunft für andere verfügbar sind, ohne, dass ich die Verantwortung der Veröffentlichung tragen muss.*
5. *Die Aufbereitung meiner Daten erfüllt die Anforderungen der Datenmanagement Systeme von Institutionen, Investoren, Fachzeitschriften und der Community.“*

Ein steiniger Weg

Auch wenn die Community rund um semantische Technologie in den letzten Jahren sprichwörtlich Berge versetzt hat, kann von Rasten keine Rede sein. Um das volle Potenzial von Open Linked Data nutzen zu können, muss noch einiges getan werden. Über gängige Probleme, wie zum Beispiel die Vereinbarung von multiplen Datensätzen, die ausschließlich in unterschiedlichen Datenformaten verfügbar sind oder das regelmäßige Aufkommen von Datensatzduplikaten aus verschiedenen Datenquellen bis hin zu der Notwendigkeit einer Vereinheitlichung der für die Katalogisierung von Datensätzen notwendigen Metadaten wird auf der SEMANTICS rege diskutiert.

Für Michel Dumontier sind noch einige Voraussetzungen zu erfüllen, bis das semantische Web im Sinne der FAIR Statuten operabel ist. Zum einen müssen eindeutige Identifikatoren und hochwertige Metadaten zur Anwendung kommen. Es muss sich auf ein geteiltes Vokabular und Gemeinschaftsstandards geeinigt werden. Auch die Speichersysteme und Nutzungsbedingungen müssen einen möglichst simplen gemeinsamen Nenner finden. Zusätzlich appelliert er an die sozialen und technologischen Verpflichtungen der Open Linked Data User in der Gestaltung des semantischen Webs.



Abbildung 4: Valentina Presutti, National Research Council (NRC). (Foto: Johannes Hüsich)

Valentina Presutti, Forscherin am National Research Council (NRC) in Rom, gibt sich unterdessen im seman-

tischen Web auf die Suche nach dem „Common Sense“, zu Deutsch dem „gesunden Menschenverstand“. *„Der gesunde Menschenverstand ist diese Art von Wissen, die wir für selbstverständlich erachten, wenn wir mit anderen Menschen kommunizieren, unabhängig vom Kommunikationsmittel (z.B. Text, mündlich). Wir gehen davon aus, dass (eine Gruppe von) Menschen es teilen. Der gesunde Menschenverstand beinhaltet auch die Art der Argumentation, die wir mit diesem Wissen führen“*, definiert sie und unterstreicht, dass Maschinen, egal wie intelligent sie sind, dieses Wissen nicht teilen, solange sie nicht damit versorgt werden. Da die Open Linked Data Cloud Basis für Künstliche Intelligenz in Form von zum Beispiel intelligenten Assistenzsystemen wie Siri und Google Now darstellt, müsse das Web mit so genanntem „Common Sense Knowledge“ versehen werden.

Trotz der überragenden Füllmenge an Faktenwissen ist dies zu großen Teilen bereits der Fall. Denn durch die Verlinkungen bzw. den beschreibenden Charakter des semantischen Webs wurden Datenpunkte mit Metadaten versehen, die auf gängigen Konventionen basieren. Nichts desto trotz sind diese „Common Sense“ Informationen oft *„[...] unvollständig, meist informell, weisen Heterogenität auf, sind nicht kontextualisiert und kaum verknüpft“*.

Eva Milgotin

FH St. Pölten

Department Medien & Wirtschaft

Matthias Corvinus-Straße 15

3100 St. Pölten

eva.milgotin@gmail.com

Doch auch Rom wurde nicht an einem Tag erbaut...

...weiß Persutti, während sie an eine Erweiterung des bisher auf einer modelltheoretischen Semantik aufbauenden Data Webs appelliert. Trotz der schnellen Entwicklung des semantischen Datennetzes und der dazugehörigen Technologien ist also noch einiges zu tun, bis wir alle endlich dieselbe Sprache sprechen. Da sind sich die Fachleute der Community einig.

Es wird also wie versprochen nicht langweilig rund um die SEMANTICS, egal wo sie stattfinden. Der spannende Diskurs über das Data Web und die Zukunft von Wissensgraphen sowie Künstlicher Intelligenz wird 2020 zum 16. Mal weitergeführt: zum ersten Mal mit einer US SEMANTICS vom 21.–23. April in Austin, Texas. In alter Manier findet man dann wieder in Europa zusammen – diesmal in Amsterdam, Niederlande von 7. bis zum 11. September 2020.

Deskriptoren: Tagung, Semantik, Ontologie, Künstliche Intelligenz, Wissensgraph, Linked Data, Open Data