

# **Festvortrag**

**von Prof. Dr. Dr. h.c. mult. Wolfgang Wahlster,  
Deutsches Forschungszentrum für Künstliche Intelligenz,  
Saarbrücken**

**anlässlich der Abschlussveranstaltung des  
Informatikjahres – Wissenschaftsjahr 2006**

**am 18. Dezember 2006 in Berlin**

**Es gilt das gesprochene Wort!**

**Sperrfrist: Beginn der Rede!**

Sehr geehrte Frau Ministerin Dr. Schavan, meine Damen und Herren,

Als Informatiker, der als zweites Hauptfach Sprachwissenschaft studiert hat und neben einem Lehrstuhl in der naturwissenschaftlichen-technischen Fakultät auch kooptierter Professor in der philosophischen Fakultät ist, möchte ich in diesem Festvortrag die Brücke zwischen dem jetzt ablaufenden, sehr erfolgreichen Informatikjahr und dem kommenden Jahr der Geisteswissenschaften schlagen, indem ich über die „Semantische Wende in der Informatik“ spreche.

„Semantisches Web, Semantische Technologien, Semantische Modellierung, Semantische Netze, Semantische Suche, Semantische Dienste, Semantische Navigation, Semantisches Grid, Semantische Integrität, Semantische Daten- und Geschäftsprozessintegration, Semantische Interoperabilität, Semantische Mediation“ all das sind aktuelle Themen der Informatik-Forschung.

„Semantik“ ist ein Begriff aus den Geisteswissenschaften und steht in der allgemeinen Zeichentheorie für die Bedeutungslehre. Es geht also um die Inhalte, den Sinn, die Bedeutung von Texten, aber auch von Bildern und Gesten bis hin zu künstlerischen Darbietungen. Zwischenmenschliche Kommunikation ist multimodal. Es werden meist verschiedene Sinnesorgane gleichzeitig angesprochen und unterschiedliche Zeichensysteme in synergistischer Kombination verwendet, wie das Beispiel des Sprechens dieses Festvortrages in Kombination mit Gestik, Mimik und Körpersprache verdeutlicht. Sprachen in der allgemeinen Bedeutung des Begriffs sind daher nicht nur auf die gesprochene Sprache und Schriftsysteme beschränkt, sondern umfassen eine weites Spektrum von Zeichensystemen von der Bildsprache über die Körpersprache bis hin zur Musiksprache. Mich hat es sehr gefreut, dass Sie Frau Ministerin Dr. Schavan das Thema „Sprachen des Menschen“ als inhaltliche Klammer des kommenden Jahres der Geisteswissenschaften mit den Akteuren vereinbart haben.

Die amerikanische Suchmaschine Google, die bekanntlich noch keine Semantik der Anfragen und der durchsuchten Dokumente berücksichtigt, sondern die Wörter nur als bedeutungslose Zeichenreihen betrachtet und statistische Beziehungen zwischen diesen auswertet, liefert aktuell fast zwei Millionen Dokumente zum Suchwort „Semantik“ zurück.

Aber wie kann nun dem Computer, der ja ursprünglich für die strukturelle Verarbeitung von Zeichenketten entwickelt wurde, eine Semantik vermittelt werden? Hier gibt es hauptsächlich zwei Ansätze: die intensionale Semantik und die Referenzsemantik. Die intensionale Semantik beschäftigt sich mit der inneren Strukturierung einer Begriffswelt, den Relationen zwischen den Zeichen und der Konstruktion komplexer Bedeutungsstrukturen mithilfe eines Inventars an Verknüpfungsregeln für Elementarbedeutungen. So formuliert der Informatiker in einer modernen Wissensrepräsentationssprache wie OWL: „Ein Düsenflugzeug ist eine Spezialisierung eines Flugzeuges. Ein Flugzeug kann ein Antriebsaggregat haben. Ein Düsenflugzeug ist ein Flugzeug, das mindestens ein Antriebsaggregat hat, das eine Jetdüse ist.“ Mit epistemologischen Relationen wie „Unterbegriff von“,

„Gegenteil von“ oder „Gleichbedeutend mit“ werden schrittweise komplexe Begriffsnetze, so genannte Ontologien aufgebaut.

Die Referenzsemantik behandelt dagegen nicht die Relationen der Zeichen untereinander sondern die Relation zwischen den Zeichen und der Welt. Viele Computersysteme haben heute einen Zugang zur Realwelt durch Sensoren, etwa durch Digitalkameras, Temperaturfühler, Druckmesser oder RFID-Leser. So kann die extensionale Bedeutung eines Begriffs wie Düsenflugzeug für einen Computer dadurch definiert werden, dass sein Bildanalysesystem entlang der Tragfläche eines erkannten Flugzeuges nach geometrischen Formen sucht, die wie typische Jetdüsen aussehen, oder einen Propeller findet, der dann gegen ein Düsenflugzeug spricht.

Als Informatiker und damit als Ingenieur verfolge ich in meinen Forschungen zu den semantischen Informationstechnologien eine sehr humanistische Zielsetzung:

Wir wollen die Sprache der Computer an die Sprache der Menschen anpassen.

Der Mensch soll sich nicht länger an die Computer anpassen müssen. Umgekehrt muss sich die Informationstechnik an den Menschen anpassen – was in meinen Augen die größte Herausforderung für die Informatik in den nächsten Jahren ist und den Kern der semantischen Wende darstellt. Nicht die Interaktion über komplizierte Kunstsprachen mit Tastatur und Maus, sondern die Kooperation in der Alltagsbegrifflichkeit mit Sprache und Gestik sollen in Zukunft im Zentrum einer „Informatik für den Menschen“ stehen.

„Informatik für den Menschen!“ – das war das Thema der Jahrestagung der Gesellschaft für Informatik in Dresden und kein anderes Thema hätte besser zur dringend notwendigen Kooperation zwischen der Informatik und den Geisteswissenschaften gepasst.

Softwaresysteme müssen in der nächsten Dekade hauptsächlich aus zwei Gründen noch intelligenter werden: damit sie besser verstehen, was der Mensch von ihnen will, und damit sie sich umgekehrt dem Menschen einfacher verständlich machen. Unser Ziel ist es, dass jedermann, der heute zu Hause oder unterwegs ein Fernsehgerät, ein Radio oder ein Telefon nutzt, künftig eine Technik, die all diese Funktionen mit einer Vielzahl von anderen nützlichen Informatikdiensten vereint, mit der gleichen Selbstverständlichkeit einfach und effizient bedienen kann. Eine der wichtigsten Herausforderungen für die zukünftige Wissensgesellschaft ist daher die Schaffung intelligenter Technologien für die Mensch-Technik-Interaktion, die den natürlichen Kommunikationsstil von Technikern akzeptieren, einen direkten Dialog mit der Technik unterstützen und damit Hemmschwellen bei der Nutzung von Hochtechnologie abbauen. Da elektronische Interaktion ein integraler Bestandteil des täglichen Lebens, der Arbeit und der Erziehung sein wird, könnten sonst rasch erhebliche Nachteile für Bevölkerungskreise entstehen, die nicht in der Lage sind, diese Technologien für sich zu nutzen. Die nächste Generation von Bedienoberflächen für Computer wird daher nicht mehr von der Bürometapher mit elektronischem Schreibtisch, verschiedenen Ordnern und Papierkorb dominiert,

sondern sie wird wesentlich von der Vorstellung eines persönlichen digitalen Assistenten bestimmt sein, der ihn in möglichst vielen Situationen im Alltag und Berufsleben intelligent unterstützt.

Semantische Technologien überbrücken die Lücke zwischen der Fachsprache der Informatik und den Sprachen ihrer Anwender, weil sie es erlauben, verschiedene Begriffssysteme ohne Bedeutungsverlust ineinander zu übersetzen. Der Automobilingenieur, der Medizintechniker oder der Logistikexperte ist mit semantischen Technologien in der Lage, sein Wissen und seine Prozessmodelle digital in der eigenen Fachsprache zu formulieren, ohne die speziellen künstlichen Sprachen zur maschinellen Wissensrepräsentation erlernen zu müssen.

Nach einer Gartner-Studie entstehen heute noch 40 Prozent der IT-Kosten bei der notwendigen Integration unterschiedlicher Softwaresysteme. Eine Abhilfe können semantische Technologien in zweierlei Hinsicht schaffen: Zum einen bieten semantische Modelle einen höheren Abstraktionsgrad und sind in der Fachsprache der Geschäftsexperten ohne spezielle IT-Kenntnisse abgefasst. Damit schirmt man den Anwender von unnötigen technischen Details ab und erlaubt ihm stattdessen, in seiner gewohnten Terminologie zu arbeiten.

Die Vision vom digitalen Krankenhaus der Zukunft wird wesentlich durch semantische IT-Technologien bestimmt sein, welche die Interoperabilität aller vernetzten medizintechnischen Systeme auf der semantischen Ebene gewährleisten. Die Mustererkennung eines Tumors im Anfangstadium muss mit den Daten aus der digitalen Patientenakte verglichen werden, in ein medizinisches Expertensystem eingespeist werden, das den Arzt mit aktuellem Fachwissen bei der Diagnose unterstützt, und der Therapievorschlag muss dann von einer Software zur Arzneimittelinteraktion überprüft werden. Heute sterben in Deutschland mehr Menschen durch falsche Arzneimittelverordnungen und unerwünschte Interaktion von Medikamenten als im Straßenverkehr, weil der automatisierte semantische Abgleich zwischen Befund, Diagnose, Therapie, und Arzneimittelwirkung und -interaktion noch fehlt.

Semantische Technologien ermöglichen aber auch einen höheren Grad an Automatisierung und Wiederverwendung von Informatikdiensten. Durch eine ontologisch angereicherte Beschreibung von Softwarediensten können auf Anfrage mithilfe von automatischen Planungsverfahren Basisdienste zu höherwertigen Diensten kombiniert werden. In dem derzeit größten deutschen Leitinnovationsprojekt zu semantischen Technologien, das von Ihrem Hause, Frau Ministerin Dr. Schavan, als BMBF-Projekt unter dem Namen SmartWeb gefördert wird, kann beispielsweise die Anfrage eines Autofahrers „Wo gibt es hier das billigste Superbenzin?“ dadurch beantwortet werden, dass nach der Bestimmung des aktuellen Aufenthaltsortes der Fahrzeuges in der Nähe nach Tankstellen gesucht wird und aus den Internetseiten der gefundenen Tankketten die aktuellen Preise für Superbenzin bestimmt werden, um dann in einem weiteren Dienst den Ort des günstigsten Anbieters in das Navigationssystem des Fahrzeuges einzuspeisen. Eine gezielte Suche und Komposition von Webdiensten könnte ohne eine für den Rechner

verständliche semantische Beschreibung der relevanten Softwarekomponenten nicht realisiert werden.

Semantische Technologien ermöglichen auch erst den Austausch von Benutzermodellen, so dass meine Computer mich als Benutzer immer besser kennen lernen. Die meisten Menschen kommen heute täglich mit Dutzenden von Informatiksystemen in Kontakt - nicht nur am Arbeitsplatz sondern auch im Auto, am Telefon, beim Einkauf und im Wohnzimmer. Aus der Interaktion mit einer Vielzahl eingebetteter Systeme können zahlreiche Eigenschaften, Verhaltensmuster und Vorlieben eines Individuums abgeleitet werden und über semantische Technologien zu einem umfassenden Benutzermodell integriert werden, das zur situationsabhängigen Personalisierung von Informatiksystemen genutzt werden kann. Durch ubiquitäre Benutzermodellierung werden adaptive und proaktive Assistenzsysteme realisierbar, ohne dass der Benutzer diese umständlich immer wieder selbst an seine Bedürfnisse anpassen muss. Das Gebiet der Benutzermodellierung, das in diesem Jahr seinen zwanzigsten Geburtstag feiert, hat übrigens wie so viele grundlegende Informatikkonzepte seinen Ursprung in Deutschland, wo wir 1986 die erste internationale Konferenz zur Benutzermodellierung im Kloster Maria Laach veranstalteten.

Meine Damen und Herren, das klassische Web 1.0 hat den weltweiten Zugang zu digital gespeicherter Information drastisch verbessert. Aber dort sind die Inhalte nur maschinenlesbar, ohne maschinell verstehbar zu sein. Das semantische Web basiert auf der inhaltlichen Beschreibung digitaler Dokumente mit standardisierten Vokabularien, die eine maschinell verstehbare Semantik haben. Die semantische Wende führt uns vom syntaktischen Web mit seiner Informationsüberflutung durch die Reduktion von Textdokumenten auf die sinnfreie Kombination von Buchstaben zu einem hochpräzisen Antwortverhalten in einem Web, das Sinnzusammenhänge in den Mittelpunkt stellt.

Den Kern der semantischen Webtechnologie bilden Markierungssprachen, die eine formale Semantik haben und in Form einer Ontologie eine standardisierte Begrifflichkeit zur Beschreibung digitaler Inhalte bereitstellen. Bei der Suche nach Information können Schlussfolgerungen den Suchprozess beschleunigen oder gesuchte Information ableiten. Methoden des maschinellen Lernens, des Data Mining und des Text, Speech und Video Mining reduzieren beim Aufbau und bei der Wartung von Ontologien den manuellen Aufwand.

Das Web 2.0 ist noch ein primär syntaktisches Web, das aber durch die private Annotation von multimedialen Dokumenten und deren Bereitstellen für die Öffentlichkeit eine große soziale Dynamik entfaltet. Allerdings fehlt den ad-hoc Markierungen im Web 2.0 die Systematik und Präzision der Ontologien im semantischen Web, auf der letztlich auch ein maschinelles Verstehen basieren muss. Wenn es gelingt, die Ontologien des semantischen Webs in die Massenbewegung des Web 2.0 einzubringen, so dass die privaten Endnutzer in einfacher und effizienter Weise bei der Erstellung von Wikis, Blogs und Mashups auf wohlfundierte Ontologiewerkzeuge zurückgreifen können, dann kann in der nächsten

Dekade ein Web 3.0 entstehen, das die Vorteile des semantischen Web und des Web 2.0 verknüpft und deren Nachteile überwindet.

Das vom BMBF geförderte und vom DFKI geleitete SmartWeb-Konsortium demonstriert den zum Semantischen Web durch multimodalen Dialog über ein mobiles Endgerät. Der Benutzer kann mit SmartWeb seine Fragen frei in Umgangssprache formuliert in das mobile Gerät einsprechen und mit Zeigegesten auf dem Bildschirm unterstützen. Die Faktenfrage „Wie hoch ist der Mount Fuji?“ wird nicht durch eine Vielzahl von Dokumenten beantwortet, in denen der Benutzer die Höhe selbst herausuchen muss, sondern prägnant mit der gewünschten Angabe „3776 m“. Die Weiterentwicklung der situationsadaptiven Dialogkomponente zum Zugang zur digitalen Wissensinfrastruktur ist auch eine Basistechnologie für das geplante Theseus-Vorhaben, das derzeit vom BMWi als Komponente der Hightech-Strategie vorbereitet wird. Durch die Konvergenz der Informatik-Trends „Semantisches Web“ und „Web 2.0“ ergibt sich in Theseus die Chance, als Web 3.0 eine innovative Wissensinfrastruktur zu schaffen, die völlig neuartige Formen des Wissenszugriffs, des Wissensmanagements und der webbasierten Dienste, Anwendungen und Geschäftsmodelle ermöglicht.

Im Web 3.0 werden Gegenstände des alltäglichen Lebens online vernetzt zu einem „Internet der Dinge“ – vom Mobiltelefon zum Fotoapparat, vom Auto bis zum Einkaufswagen. Die Menschen werden das digitale Netz um sie herum nicht mehr wahrnehmen – es wird als Umgebungsintelligenz einfach da sein. Die eingebetteten Systeme in unseren Fahrzeugen, die mittlerweile mit über 70 Mikrocomputern zu fahrenden Computernetzwerken geworden sind, müssen untereinander von Fahrzeug zu Fahrzeug kommunizieren, um eine noch höhere mobile Sicherheit zu realisieren und der Vision des hundertprozentig sicheren Autos näher zu kommen. Auch hier geht es nicht ohne semantische Technologien. Denn wenn ein französischer Peugeot über seinen Wassersensor feststellt, dass auf der Fahrbahn durch eine große Wasserlache die Gefahr von Aquaplaning vorliegt, soll er dies über eine ad-hoc Internet-Verbindung nicht auf französisch sondern in einer universell verstehbaren Semantik an das lokale Gefahrenwarnsystem eines folgenden BMW-Motorrads melden. Dort bekommt dann der Motorradfahrer die für ihn vielleicht lebensrettende Warnung: „Achtung in 200m Aquaplaning!“. Unsere deutsche Automobilindustrie ist bei dieser Form der Car-to-X Kommunikation zwischen Fahrzeugen und der Verkehrsinfrastruktur weltweit führend. Wir wollen im nächsten Jahr einen großen Feldversuch starten, um danach diese wichtige Innovation in den Markt zu bringen.

Meine Damen und Herren, ich bin mir sicher, dass wir in wenigen Jahren mit unseren Autos häufiger in die Werkstatt müssen für einen semantischen Software-Update als für einen Ölwechsel!

Deutschland spielt bei der Innovation mit semantischen Technologien schon sehr lange in der ersten Liga: Historisch hat Gottfried Wilhelm Leibniz nicht nur die erste mechanische Rechenmaschine auf der Basis des binären Zahlensystems entwickelt, sondern auch mit dem nach ihm benannten Leibniz-Gesetz die Grundlage für die

semantische Informationsverarbeitung gelegt. Und der Deutsche Konrad Zuse hat nicht nur den ersten frei programmierbaren Computer Z1 entwickelt, sondern mit seinem Plankalkül auch die erste universelle höhere Programmiersprache mit einer anspruchsvollen Semantik. Die Grundlagen für die heutigen Wissensrepräsentationssprachen wurden in Mitte der achtziger Jahre von deutschen Informatikern gelegt und der aktuelle Standard OWL wurde maßgeblich von Deutschen mitgeprägt.

Meine Damen und Herren, Innovationen entstehen heute meist an den Schnittstellen der Disziplinen. Die Computerlinguistik, also das Studium der maschinellen Verarbeitung natürlicher Sprachen, bildet eine der Brücken zwischen der Informatik und der Geisteswissenschaft. Es ist nach meinem Wissen das einzige Fach, das sich aktiv und mit gleichem Recht am Jahr der Informatik beteiligt hat und am Jahr der Geisteswissenschaft beteiligen wird. Übrigens hat dieses Brückenfach bereits das geschafft, was die Informatik unbedingt anstreben muss: Die Zahl der weiblichen Absolventen liegt bei fast 50 Prozent. Einer der Gründer der Computerlinguistik, Professor Martin Kay empfiehlt seinen Studenten an der Eliteuniversität Stanford inzwischen, ihr Studium in Deutschland fortzusetzen, weil wir hier zum Beispiel in Saarbrücken eines der weltweit größten und fachlich einzigartigen Zentren für Computerlinguistik aufgebaut haben. Ein besseres Kompliment für die deutsche Forschung gibt es wohl kaum.

Ich möchte zum Schluss meines Festvortrages den Bogen zum Anfang spannen und den Kollegen Kargerzmann zitieren, der in seiner Rede zum Auftakt des Informatikjahres im Januar zu Recht forderte „Das Jahr der Informatik darf kein Strohfeuer sein, das nur für zwölf Monate geschürt wird.“

Das Jahr 2006 als Wissenschaftsjahr für die Informatik war überfällig. Es gab in den letzten Jahren kaum eine Innovation, die nicht durch die Querschnittstechnologie der Informatik geprägt oder zumindest unterstützt wurde – und das wird noch viele Jahre so bleiben: Die Informatik ist einfach der Innovationsmotor Nummer Eins. Die Physik konnte neben ihrem eigentlichen Wissenschaftsjahr 2000 mit dem Einsteinjahr in 2005 bereits in einem zweiten Jahr die Aufmerksamkeit der Öffentlichkeit auf sich lenken. Die Informatik als wichtigster Wachstumsbeschleuniger für unsere Wirtschaft sollte auch in den nächsten Jahren ein weiteres auf eine überragende Forscherpersönlichkeit ausgerichtetes Jahr anstreben: Ich setze mich daher für eine Leibnizjahr oder ein Zusejahr ein. Durch die Personifizierung von Wissenschaft durch diese deutschen Pioniere der Informatik und über die Faszinationskraft dieser großen Erfinder und facettenreichen Persönlichkeiten können wir die Jugend für Informatik begeistern – auch über das diesjährige Informatikjahr hinaus.

Ich danke Ihnen für Ihre Aufmerksamkeit.