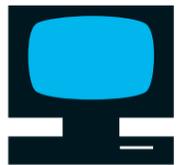


# ICH VERSTEHEN DICH ...

Andere sind noch beim Web 2.0, da arbeiten visionäre Entwickler schon an der nächsten Internet-Generation. Das **WEB 3.0** wird intelligent – ein Google mit Grips.



Von Manfred Buchner

**ALS FRANÇOIS BRY 1974** erstmals eine Computermaus sah, war der frisch gebackene Abiturient tief beeindruckt: „Viele hielten das für eine Spinnerei. Für mich aber war das kleine Ding eine visionäre Entwicklung.“ Heute beschäftigt sich Bry, mittlerweile Informatikprofessor an der Universität München, mit einem Thema, das in einigen Jahren ähnlich alltäglich sein könnte, wie es heute die Maus ist: dem semantischen Web. Während die Internet-Gemeinde gegenwärtig noch um das Goldene Kalb namens Web 2.0 tanzt und bei Wikipedia, in Blogs und Videotauschportalen wie YouTube den

Spaß an der eigenen Medienproduktion entdeckt, sehen Visionäre bereits ein völlig neues Internet heranreifen – das Web 3.0.

**DAS SEMANTISCHE WEB** – frei übersetzt „verstehendes Web“ – ist eine Idee von Tim Berners-Lee, dem Erfinder des World Wide Web. Noch vor dem großen Internet-Goldrausch hatte der von der britischen Queen zum Sir geadelte Informatiker die Gefahren der anrollenden Datenflut und das drohende Informationschaos vorausgesehen. Sein Rettungsplan: Er will Computer befähigen, Daten nicht nur zu lesen, sondern auch zu verstehen. Als Hel-

fer mit Grips sollen sie die Infoschätze im Internet wesentlich besser als bislang nutzen und damit Berners-Lees Traum erfüllen. Demnach ist „das Web ein Ort, an dem die emotionalen Fähigkeiten eines Menschen und das logische Denken eines Computers in einer idealen, leistungsfähigen Mischung nebeneinander existieren“. Der Weg dorthin ist allerdings weit. Das machen Eingaben bei Suchmaschinen deutlich, die endlose Listen vermeintlicher Hits produzieren. Wer die Worte „Wann fliegt die nächste Maschine von Berlin nach München?“ bei Google eingibt (ohne Anführungszeichen), bekommt 95 600 Seiten genannt – aber keine Antwort. Nach einer empirischen Untersuchung der Universität Düsseldorf sind mehr als zwei Drittel der bei Google angezeigten Ergebnisse wertlos, weil es ihnen an Relevanz mangelt. Hinzu kommt: Viele Infoquellen bleiben ungenutzt, weil Google und Co. Dokumente in Datenbanken oder Daten in Tabellen meist gar nicht erfassen. Sie sind Teil des „Invisible Web“, dessen Umfang deutlich größer ist als der sichtbare Teil des globalen Wissenslabyrinths.

**DIE OFT VERGEBLICHE** Sucherei kostet Zeit und Geld. So beklagte Bill Gates auf dem CEO Summit 2006 in Redmond vor Firmenchefs aus aller Welt die dadurch entstehenden Produktivitätseinbußen. Nach Schätzungen des Unternehmens verbringen Angestellte rund ein Drittel ihrer Arbeitszeit mit der Suche nach Informationen, was pro Mitarbeiter jährlich im Schnitt 18 000 Dollar kostete. Mit dem semantischen Web ist Abhilfe in Sicht. Als Direktor des World Wide Web Consortium (W3C) hat Tim Berners-Lee technische Standards und Verfahren für die nötige Erweiterung des Webs definiert. Dabei geht es um zwei konkrete Ziele: 1. den Austausch von Daten erleichtern und alle Datenbestände maschinell erschließbar machen, 2. den Bezug von Daten auf Personen und den Kontext hin organisieren. „Google mit Köpfchen“ nennt der Münchner Professor François Bry die Fähigkeiten

des neuen Internets, mit denen Surfer vereinzelt schon heute Bekanntschaft machen können. Etwa beim Online-Versandhändler Amazon. Sämtliche Seitenaufrufe von Besuchern werden dort registriert. Daraus lassen sich persönliche Empfehlungen generieren, zum Beispiel Buchtitel. Oder bei Google Maps und Windows Live Local. Wer hier eine Adresse eintippt, bekommt automatisch einen Stadtplan mit Straßenausschnitt angezeigt – auf Wunsch sogar dreidimensional. Möglich macht's die Verknüpfung von Text- und Grafikdateien.

**DOCH DAS IST ERST** der Anfang. Durch Kombination von bekannten Methoden wie Business-Intelligence, Expertensystemen oder Data-Mining mit neuen semantischen Verfahren wird das digitale Verhalten von Nutzern inklusive des Umfelds künftig systematisch durchleuchtet. Dazu wertet man sämtliche Datenein- und -ausgaben aus, die verwendeten Dokumente, angeklickten Websites und E-Mails. Aus den gewonnenen Benutzerprofilen produziert der PC automatisch personalisierte,

## EU-PROJEKT

Das EU-Forschungsprojekt **REVERSE (REasoning on the WEb with Rules and SEmantics)** soll Intelligenz ins Internet bringen. Einige Beispiele aus der Themenliste:

- **Rule Markup Languages: einheitliche Auszeichnungssprache mit Werkzeugen für die Zuordnung von Begriffen**
- **Composition and Typing: Methoden und Regeln für die Softwareinteroperabilität im Internet**
- **Reasoning-aware Querying: logisches Schließen bei Web-Anfragen**
- **Entscheidungshilfen für Ereignis-, Zeit- und geografische Daten**
- **Personalisierte Systeme für benutzeradaptive Info- und Lehrsysteme**

auf Zeit und Situation bezogene Infos und Aktionsvorschläge. Dafür sorgt vor allem Resource Description Framework (RDF), die Schlüsseltechnologie des Semantic Web. Diese Programmiersprache funktioniert ähnlich wie das Satzbaumuster „Subjekt – Prädikat – Objekt“ im Grammatikunterricht in der Schule. Mithilfe sogenannter Metadaten werden Informationskategorien gebildet, mit denen Computer in einem Dokument zum Beispiel den Autor, das Erstellungsdatum, den Titel oder andere Bestandteile identifizieren können. Durch Zuordnung von realen Begriffen zu Metadaten entstehen maschinenlesbare Datenlexika – das Gedächtnis des Rechners.

**VIELES STECKT ALLERDINGS** noch im Forschungsstadium. Das zeigt ein Blick in die Sprachwelt des Semantic Web, in der es von fremdartigen Begriffen wie Annotation, Ontologie, Reasoning und Taxonomien nur so wimmelt. Thema des EU-Projekts Reverse (siehe Kasten) ist etwa die Entwicklung von semantischen Methoden auf Basis der von den W3C-Gremien gesetzten Standards. Rund 100 Wissenschaftler aus 27 europäischen Forschungseinrichtungen sind an dem 5,5 Millionen Euro teuren Vorhaben beteiligt, das Professor François Bry zusammen mit Kollegen der schwedischen Uni Linköping koordiniert. Bis das intelligente Web in großem Stil verfügbar ist, werden nach Schätzung der Analysten von Gartner noch fünf bis zehn Jahre vergehen. Doch das Abwarten lohnt, denn die Vorteile sind enorm. So können laut einer aktuellen Studie des Fraunhofer-Instituts für Rechnerarchitektur und Softwaretechnik Unternehmen bei Geschäftsprozessen mit Effizienzgewinnen von bis zu 50 Prozent rechnen. Dass es nicht schneller vorangeht, hat nach Meinung von Web-Visionär Tim Berners-Lee eine einfache Ursache – den Konservatismus der Anwender: „Sie sind nur schwer von der neuen Technik zu überzeugen. Es war für die Leute ja auch schwierig, das World Wide Web zu verstehen, bevor es existierte.“ ||