

Nur wenige Forschungsprojekte erreichen diese Größe: 5,5 Millionen Euro auf vier Jahre verteilt macht die Europäische Union und das Schweizer Bundesamt für Forschung und Bildung innerhalb ihres sechsten Rahmenprogramms für REWERSE locker. Das steht für **RE**asoning on the **WE**b with **R**ules and **SE**mantics. Das Projekt wird von LMU-Informatikern koordiniert und beschäftigt rund 100 Wissenschaftler aus 27 europäischen Einrichtungen, die mit ihrer Arbeit das Internet der nächsten Generation entstehen lassen wollen: das semantische Web.

ULRIKE OSTLER

## REWERSE MACHT COMPUTER „INTELLIGENT“

**S**eit gut einem Jahr dreht sich in dem internationalen Projekt REWERSE alles um Regeln, Beschreibungssprachen und Klassifikationssysteme. REWERSE beschäftigt sich mit einer Vision für das Web der Zukunft, in dem es auch Computern möglich sein soll, mit Informationen „intelligent“ umzugehen. Handelt es sich also bei dem semantischen Web um die Entwicklung neuer Killerapplikationen für das Internet? Nein, sagt Tim Berners-Lee, der als Erfinder des World Wide Web (WWW) gilt, in einem Artikel aus dem Jahr 2001. Vielmehr sei das semantische Web selbst der Killer des jetzigen Netzes. 1989 schlug er seinem damaligen Arbeitgeber, dem Europäischen Labor für Teilchenphysik, Cern, ein Projekt vor, das den weltweiten Austausch und die Aktualisierung von Informationen zwischen Wissenschaftlern vereinfachen sollte. Als grundlegendes Prinzip sollte Hypertext dienen, Links von Zeichenketten oder Bildern auf andere Zeichenketten oder Bilder. Obwohl diese Richtung, wie sich Professor François Bry von der Lehr- und Forschungseinheit Programmier- und Modellierungssprachen an der LMU erinnert, damals für so manchen Experten als Spinnerei galt, verwirklichte Tim Berners-Lee das Projekt und entwickelte den ersten Web-Server. Die Idee des Webs war so einfach wie genial, sagt François Bry. Allerdings funktioniert sie nur dank menschlicher Interaktion. Das Projekt REWERSE, das von François Bry gemeinsam mit seinem LMU-Kollegen Professor Hans Jürgen Ohlbach sowie Professor Jan Małuszynski von der schwedischen Universität Linköping koordiniert wird, will nun den nächsten Schritt machen und Prozesse automatisieren. Dazu ein Beispiel: Die Suchmaschine *Google* kennt zwei „François Bry“. Das Abfrageergebnis listet neben dem LMU-Professor auch einen Musiker aus Toulouse auf. Die Unterscheidung ist für Menschen kein Problem, aber Suchroboter können nicht erkennen, nach welcher Person im Web gefahndet wird. So kommt es, dass nach jüngsten Untersuchungen der Abteilung Informationswissenschaft an der Düsseldorfer Hochschule Treffer von Suchmaschinen im deutschen Web, das aus rund vier Millionen Domains mit 320 Millionen Seiten besteht, in mindestens vier von zehn Fällen keine Relevanz haben. Da hilft auch keine Einschränkung nach Datum. Nur vier bis sechs Dokumente werden zeitlich richtig zugeordnet. Es fehlen eindeutige, für Rechner lesbare Bedeutungen und vor allem Methoden für Schlussfolgerungen und Möglichkeiten, das Folgern mit Anfragen zu koppeln.

Das wichtigste Gremium zur Entwicklung von Formalismen zur Daten-Modellierung und von Methoden zur Daten-Verarbeitung im Web ist das World Wide Web Consortium (W3C), dem Tim Berners-Lee vorsteht. Es ist ein offenes Forum für Unternehmen und Organisationen, das patentfreie Standards rund um das Web verabschiedet. Es arbeitet an Strukturen, mit denen sich Informationen kategorisieren lassen. Dabei entsteht neben einer Typisierung auch

► Mit der Sprache CaTTS können Termine und Zeiträume definiert werden, d.h. in Bezug auf kulturelle Kalendare, „Arbeitskalender“ und Zeitzonen. So können Menschen an verschiedenen Orten Termine vereinbaren. Dabei kann jeder die Zeitangaben von allen daran Beteiligten nach seinem eigenen Kalender und seiner Zeitzone erfahren. Mit der Sprache MPPL können ferner Wege gesucht werden, um ein Treffen zu vereinbaren. Dabei werden nicht nur topologische Aspekte (wie etwa Entfernung oder Koordinaten) berücksichtigt, sondern auch Kontexte (wie etwa „zu Fuß“, „im Auto“ usw.).

eine Ontologie, die vergleichbar ist mit der Systematik eines Bibliothekskatalogs, erweitert um Regeln wie: „Autoren sind Menschen“ und „Bücher sind keine Menschen“. Den REWERSE-Forschern dienen die W3C-Standards als Grundlage, beispielsweise das *Resource Description Framework* (RDF). RDF gilt als eine Ausprägung des Datenaustauschformats *Extended Markup Language* (XML). Diese bereits weithin eingesetzte Beschreibungssprache liefert die Syntax. RDF als Auszeichnungssprache für Metadaten ist dreigliedrig, unterscheidet Subjekt (*subject*), Prädikat (*predicate*) oder auch Eigenschaft (*property*) und Objekt (*object*). Im Beispiel „Öttingerstraße 67 ist Büroadresse von François Bry“ wäre „Öttingerstraße 67“ das Subjekt, „ist Büroadresse“ das Prädikat und „François Bry“ das Objekt. Eine Website-Adresse, *Unicore Resource Locator* (URL), kann ebenfalls als Subjekt dienen, die Eigenschaft „Autor“ als Prädikat und der Name des Autors als Objekt. Somit lassen sich mit RDF Aussagen über so genannte Ressourcen treffen.

#### PRINZIPIEN DES SEMANTISCHEN WEB

*Uniform Resource Identification* (URI) ergänzen die URLs. Sie verleihen Eindeutigkeit, ermöglichen eine Identifizierung und schließen Verwechslungen aus. URLs, die Ressourcen lokalisieren, können als Sonderfall von URIs interpretiert werden. Mit Hilfe der *Web Ontology Language* (OWL) lassen sich sodann Regeln, beziehungsweise Eigenschaften vollständig darstellen und überprüfen: „Kein Dozent ist ein Erstsemestler“ oder „Ein Seminarteilnehmer ist (entweder) Studierender oder Dozent“. RDF und OWL erlauben es, WWW-Links aber auch Web-Seiten Annotationen hinzuzufügen, die ihre Bedeutung klären. So könnte aus der Beschreibung der Website „François Bry“ hervorgehen, dass es sich um einen Informatik-Professor handelt, der zum Thema Reasoning im Web doziert, publiziert und Mitarbeiter sucht. Letztendlich kann es sogar sein, dass eine Suchmaschine die Frage „Welche Veranstaltung hat François Bry 2005 in Malta moderiert?“ direkt richtig beantwortet. Auf dem Weg dorthin hat der LMU-Fachbereich Programmier- und Modellierungssprachen bereits zwei neue Sprachprototypen entwickelt: Xcerpt, eine deduktive Abfragesprache sowie XChange, eine reaktive Sprache. Denn RDF und OWL können zwar beschreiben, aber eine Bearbeitung der Daten fehlt weitgehend. Während sich mit Xcerpt Daten und Metadaten befragen lassen, kann XChange für die Mitteilung von Änderungen und Änderungswünschen verwendet werden.

Bislang fehlen zur Entwicklung des semantischen Web vor allem Werkzeuge, also Sprachen und Methoden. Dabei können auch scheinbar kleine Aufgaben sehr schnell komplex werden. Das zeigt Forschungsblock A1, in dem es um Zeit-, Orts- und Ereignisangaben geht. Diese sind keineswegs eindimensional, zum Beispiel wenn „die nächstgelegene Apotheke“ gesucht wird. Die Entfernung allein reicht nicht zur Klassifizierung; denn die nächste Apotheke könnte schon geschlossen haben oder nur mühsam erreichbar sein, wenn etwa ein Fluss oder eine Autobahn zwischen Suchendem und der Adresse liegt. Zudem weisen die meisten Web-Dokumente einen zeitlichen Bezug auf, der unterschiedlich formuliert sein kann. Von präzisen Angaben wie „12:45 Uhr mitteleuropäischer Zeit“ über „zweistündige Dauer“ und „Ende um etwa drei Uhr nachmittags“ bis „Schluss“ reicht das Spektrum, ohne dass unterschiedliche kulturelle und persönliche Kalender und Bezugspunkte wie „Ostern“ und „Pfingsten“, Geburtstage, wöchentliche Meetings, Zeitzonen, Schreibweisen und Ereignisse wie „zur Fußballweltmeisterschaft 2006“ bereits berücksichtigt wären. So beschäftigt sich das REWERSE-Team A1 mit dem so genannten *geotemporal reasoning*, dem *geospatial reasoning* und *topical information*. Letzteres ist der Versuch, relevante Informationen hierarchisch nach Themenfeldern zu ordnen, beziehungsweise entsprechende Taxonomien und Klassifikations-schemata zu entwerfen. Raumbezogene Daten, spatiale Daten, basieren auf dem bekannten Koordinatensystem, mit



dessen Hilfe sich die geografische Lage von Städten und Flüssen beschreiben lässt. Das REWERSE-Projekt befasst sich insbesondere mit ungenauen räumlichen Begriffen wie „in der Richtung“, „im Süden von“ oder „in der Nachbarschaft von“. Geotemporale Informationen schließlich finden sich in persönlichen oder in Firmenkalendern und Semesterplänen: „vergangenen Sonntag“, „zu Beginn der Sommerferien“, „die nächsten Bundestagswahlen“. Die Arbeitsgruppe A2 erforscht speziell die Bio-Informatik. DNA- und Protein-Forscher suchen etwa nach geeigneten Regelsprachen, mit denen sich die Modellierung von Sequenzeigenschaften leichter darstellen lassen als mit Programmiersprachen. Außerdem stehen im Fokus der neun REWERSE-Gruppen, die sich mit Bioinformatik beschäftigen, Text-Mining, Gendaten-Analyse, Stoffwechselsystematiken und der Informationsaustausch zwischen Proteinen.

#### PERSONALISIERTE SUCHE DURCH SOFTWARE-AGENTEN

Zu den Gebieten, auf denen REWERSE-Forscher arbeiten, gehören darüber hinaus personalisierte Informationssysteme. Die Gruppe A3 untersucht entsprechende Methoden und wendet sie prototypisch auf die Bereiche Reise und Touristik, E-Learning und Gesundheitswesen sowie die Suche im Web an. Wenn zum Beispiel jemand nach „Friedrich Schiller“ sucht, könnte sofort ein Software-Agent nach sämtlichen Werken suchen oder auch eine weitere Auswahl anbieten: Werke, Interpretationen, Rezensionen, Theateraufführungen, Sonderveranstaltungen zum Schillerjahr und Biografien beispielsweise. Ausgehend von einem aktuellen Ereignis wie einer Online-Suche lassen sich aber auch ganz andere Szenarien ersinnen. Die letzte Maschine kann zum Beispiel am Freitagabend in Heathrow nur mit großer Verspätung nach München starten. Es wäre doch schön, wenn dem genervten Geschäftsmann, der seinen Anschluss verpasst, nun auf seinem Taschencomputer Alternativen und To-Do-Listen seiner Möglichkeiten aufgezeigt würden: Ein Hotelzimmer in Flughafen-Nähe buchen, Ehefrau anrufen, den nächstmöglichen Abflugtermin angeben etc. „Da lassen sich raffinierte Dienste ersinnen“, sagt François Bry. Heute bucht der Reisewillige entweder über ein Reisebüro, weil es dort einen guten Service gibt, oder online, weil es billiger ist. Hierbei entfällt allerdings auch jegliche Beratung und die Suche im Netz kann sich recht langwierig gestalten. Zudem gibt es keine Garantie, das beste Angebot herausgefischt zu haben. Ein in einem Software-Agenten hinterlegtes Profil könnte das Problem lösen. Zum Beispiel sollte ein Flug in den Urlaub günstig sein, um mehr Geld am Urlaubsort zur Verfügung zu haben. Ein Flug zum Geschäftstermin sollte möglichst entspannend sein, ohne Umsteigen, der Preis spielt eine untergeordnete Rolle, weil schwierige Verhandlungen angesagt sind, zu denen der Reisende möglichst ausgeruht,

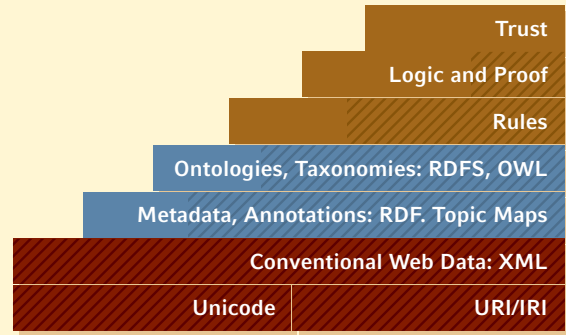
frisch und gutgelaunt erscheinen sollte. Die Suche nach adäquaten Angeboten im Netz führt ein Software-Agent durch, halbautomatisch, so dass der Suchende das Profil gegebenenfalls modifizieren oder gänzlich ändern kann. Solche hinterlegten Profile könnten auch helfen, sich die Tageszeitung individuell nach Interessenslage zusammenzustellen – nicht von einem Anbieter, sondern aus vielen Nachrichtenquellen oder Fachpublikationen. Es wäre aber auch möglich, an fremden Orten schneller Freundschaften zu schließen. Jemand, der sich an einer Universität einschreibt, braucht ein, zwei Semester, um sich in dem Sozialgefüge zurechtzufinden. Ein Profil mit Angabe der Hobbys, der Studienfächer oder auch der Stärken und Schwächen im jeweiligen Fach könnte bei der Anbahnung von Freundschaften helfen. Über das Computer-Device eines musizierenden Studenten suchen Software-Agenten nach anderen Musikern, die eine Band gründen wollen oder die Lehrkraft weiß, dass Defizite in Algebra bestehen, und kann den ultimativen Buch- oder E-Learning-Tipp geben. Eine Einordnung in ein Beziehungsgeflecht wie „ist Student an der LMU“, „lebt in Haidhausen“, „ist Sohn von Anke und Herbert“, „kommt vom Gymnasium Schloss Neuhaus“, „ist Bruder von Lars“ ermöglicht darüber hinaus, über das Umfeld Kontakte zu knüpfen. Vielleicht entdeckt ein Neuankömmling zwar kein bekanntes Gesicht aber eine Cello-Spielerin aus der Musikschule des Nachbarorts, die mit der Freundin des Bruders in den Urlaub gefahren ist.


#### DER SCHLÜSSEL ZUR PERSON

Die vernetzten Beziehungen lassen sich aber auch für Analysen nutzen. Zum Beispiel wenn es sich um wissenschaftliche Gemeinschaften handelt. Ballungen im Netz derer, die zusammenarbeiten, lassen erkennen, wer besonders aktiv ist und wer sich eher am Rand mit dem Thema beschäftigt oder auch wo ein neues Thema entsteht. Mit *Friend Of A Friend* (FOAF) existiert bereits ein weltweiter Prototyp solcher vernetzter Beziehungen. Die Kunst besteht jeweils darin, aus vorhandenen, rudimentären Informationen, neue zu folgern. Das Reasoning ist bis jetzt jedoch noch weitgehend den Menschen vorbehalten. REWERSE will das ändern. Die Arbeitsgruppe I1 beschäftigt sich mit der Modellierung von Regeln, die die Voraussetzung für das automatische Abarbeiten eines Prozesses darstellen. Dazu gehören die Formalien einer Regelsprache, ein Metamodell sowie eine abstrakte Syntax. Die Arbeitsgruppe I2 beschäftigt sich zwar auch mit Regeln, aber solchen die es ermöglichen, die Privatsphäre zu schützen und Verträge abzuschließen. *Composition and Typing* lautet das Aufgabengebiet der Arbeitsgruppe I3. Hier geht es im Wesentlichen um die Wiederverwendung von einmal erdachten Komponenten, sowohl was die Applikationen angeht als auch die der Ontologie und der Sprache. Sprachkonstrukte, die RDF und OWL ergänzen und damit das Automatisieren von Inferenzmethoden erst ermöglichen, sind das Spezialthema des REWERSE-Teams I4. Arbeitsgruppe I5 schließlich sieht das Web als lebenden Organismus, um dessen Evolution und Reaktionsfähigkeit zu analysieren. Hierbei geht es um die Nachvollziehbarkeit und Transparenz der Prozesse. Ein Beispiel dazu: Es stehen zwei Frühmaschinen nach London Heathrow zur Auswahl, von British Airways (BA) und Lufthansa. Der Lufthansa-Flug startet 20 Minuten später, der Reisende kann länger schlafen und bucht. Damit ist ein Platz mehr in dieser Maschine belegt, im BA-Flug bleibt ein Sitzplatz verfügbar. Eine Entscheidung ändert somit die Bedingungen für nachgeordnete Prozesse. Damit so etwas automatisiert ablaufen kann, bedarf es deklarativer regelbasierter Sprachen wie XChange sowie Methoden und Tools, mit denen sich die Änderungen im Web beschreiben lassen.

Das europäische *Network of REWERSE* schafft mit der Entwicklung von Regelsprachen, Systematiken, Methoden und ihrer Anwendung auf Beispielszenarien einen wichtigen Beitrag zur Entstehung des *Semantic Web*. Wie Koordinator François Bry anmerkt, sind die REWERSE-Forscher dem W3C sogar um zwei Jahre voraus. Im April

► Um ein „Web der Bedeutungen“ als ein *Semantic Web* zu errichten, wird eine Schicht-Architektur zugrunde gelegt, deren Schichten von unten nach oben aufgebaut werden. Derzeit wird die Regel-Schicht konzipiert und aufgebaut. Die Web-Anfragesprache Xcerpt ermöglicht es, Daten aus dem Web wie aus einer weltweiten Datenbank zu suchen. Xcerpt strebt Beiträge zu den markierten Schichten und somit eine einheitliche Lösung zu komplexeren Problemen an.



Targeted by REVERSE 14 Query Language 

dieses Jahres haben zwei Autoren des Netzwerks in Washington auf einem Workshop des Konsortiums die regelbasierten Sprachen für das *Semantic Web* vorgestellt. Das Ergebnis: „Die Sicht von REVERSE kam sehr gut an und es ist davon auszugehen, dass damit unsere Forschung eine positive Rolle bei der Entwicklung und Gestaltung von Regelsprachen für das Web durch das W3C spielen wird“, fasst François Bry die Resonanz zusammen. Unumstößlich ist jedoch die gesamte Idee vom semantischen Netz nicht. Kritiker sehen vor allem in der Komplexität, die mit der Klassifizierung verbunden ist, einen möglichen Grund zum Scheitern. Doch François Bry lässt Schwierigkeiten nicht gelten. „Das *Semantic Web* ist eine Vision. Das ist ein Problem, aber auch das Gute. Denn sie bietet viel Raum für Entwicklungen.“ Er erklärt: „Im Jahr 1974, unmittelbar nach meiner Abiturprüfung habe ich eine Computermesse besucht und eine Vorführung der Maus-Windows-Schnittstelle gesehen, die ein Forschungszentrum von Xerox konzipiert hatte. Noch Jahre nach dieser für mich beeindruckenden und wegweisenden Erfahrung habe ich gehört, dass dies eine Spinnerei ohne jegliche wirtschaftliche Relevanz sei.“ Er schließt: „Die Anwendungen auf dem Gebiet *Semantic Web* sind nicht weniger viel versprechend als eine Computermaus.“ Der LMU-Professor verweist auf die vielen konkreten Anwendungen, die heute schon erkennbar sind. „Das semantische Web ist eine Vision, die unsere Forschung nährt und den Ausgang offen lässt.“ In punkto vernetztes Arbeiten ist REVERSE ohnehin ein Erfolg. Bis jetzt, nach nur einem Jahr, hat jede Forschungsgruppe bereits mindestens zwei Veröffentlichungen von Wissenschaftlern präsentieren können, die bis zu diesem Zeitpunkt noch nichts gemeinsam publiziert hatten.

Prof. Dr. François Bry ist seit 1994 Professor am Institut für Informatik. Gemeinsam mit Prof. Dr. Hans Jürgen Ohlbach leitet er die Lehrinheit Programmier- und Modellierungssprachen an der LMU. Beide koordinieren das EU-Projekt REVERSE zur Entwicklung des *Semantic Web*.  
 francois.bry@ifi.lmu.de  
<http://reverse.net>

