



4.360 Zeichen
Abdruck honorarfrei
Beleg wird erbeten

Dipl.-Inf. Katja Losemann und Prof. Dr. Wim Martens haben kürzlich in Scottsville/Arizona (USA) an der ACM SIGMOD/PODS 2012 teilgenommen, einer internationalen Fachkonferenz der Association for Computing Machinery (ACM). Auf das offizielle Konferenzprogramm folgte ein Ausflug in den Grand Canyon.

Damit die Suche im World Wide Web schneller und präziser wird:

Bayreuther Informatiker erhöhen die Effizienz der Abfragesprache SPARQL

Die meisten Internet-Nutzer erleben das World Wide Web als unerschöpfliche Fundgrube für Informationen und Unterhaltung. Was aber dabei nur wenigen bewusst ist: Hinter allen Texten, Bildern und Musikstücken steht ein hochkomplexes System von Dateiformaten, Programmen und Abfragesprachen. Weltweit arbeiten Mathematiker und Informatiker daran, diese formalen Grundlagen im Interesse der Internet-Nutzer weiterzuentwickeln. Zu ihnen gehören auch Prof. Dr. Wim Martens, Professor für Theoretische Informatik an der Universität Bayreuth, und seine Doktorandin Dipl.-Inf. Katja Losemann. In der noch



jungen Abfragesprache SPARQL haben sie Schwachstellen entdeckt, die den Umgang mit großen Datenmengen verlangsamen und die schnelle, zielgenaue Suche im Internet behindern. Dem World Wide Web Consortium (W3C) haben sie bereits einen Vorschlag unterbreitet, wie dieses Problem beseitigt werden sollte. Eine Konferenz in Arizona/USA bot den Informatikern aus Bayreuth kürzlich die Gelegenheit, ihre Ideen zur Web-Optimierung der internationalen Fachwelt vorzustellen.

SPARQL ist seit 2008 eine vom W3C empfohlene Abfragesprache. Sie ermöglicht die Verarbeitung spezieller Daten, die das World Wide Web besser strukturieren: und zwar dadurch, dass sie Informationsquellen aller Art unter formalen Aspekten kenntlich machen. Dies geschieht mithilfe des Resource Description Framework (RDF), eines Regelwerks, das für die formale Beschreibung von Informationsquellen die einzuhaltenden Standards definiert. Es verhält sich wie mit den Büchern in einer großen Bibliothek. Diese werden viel leichter auffindbar, wenn Informationen über Autoren, Titel und Erscheinungsdatum sowie weitere Schlagworte auf standardisierte Weise in einer Kartei erfasst werden. In ähnlicher Weise – nur auf einem viel höheren Abstraktionsgrad – helfen die mithilfe des RDF erzeugten Daten, Informationsquellen im World Wide Web zugänglich zu machen. Um diese Daten verarbeiten zu können, wurde SPARQL entwickelt. Die neue Sprache kommt heute bei vielen Web-Recherchen zum Einsatz, ohne dass die Nutzer etwas davon bemerken.

Prof. Dr. Wim Martens und Katja Losemann haben nun herausgefunden, dass SPARQL in mancher Hinsicht kontraintuitiv ist. Die Sprache ist anders konstruiert, als mathematisch geschulte Anwender es erwarten würden. Zudem konnten sie nachweisen, dass SPARQL teilweise ineffizient arbeitet. Es dauert in einigen Fällen ungewöhnlich lange, um mittels einfacher, in SPARQL verfasster Programme eine geringe Anzahl von Bytes zu verarbeiten. Und sobald es um Datenmengen im Terabyte-Bereich geht, was bei großen Datenbanken durchaus üblich ist, entstehen sogar mit Hochleistungsrechnern extrem lange Wartezeiten von mehreren Jahren.

„Mit theoretischen Analysen und empirischen Tests ist uns der Beweis gelungen, dass diese Ineffizienz nicht durch eine falsche oder ungeschickte Programmierung zustande



kommt“, erklärt Martens. „Die Ursache liegt vielmehr in der Abfragesprache SPARQL selbst, genauer gesagt: in einem Feature, das erst 2010 eingeführt wurde. Das W3C hat dieses Feature bisher nicht als Standard eingeführt. Und so bestand für uns die Chance, mit einem Optimierungsvorschlag noch rechtzeitig in die weitere Entwicklung einzugreifen.“ Die Zusammenarbeit mit Katja Losemann führte in kurzer Zeit zu einem Vorschlag, der die entdeckten Schwachstellen beseitigt. „Wir sind optimistisch, dass das W3C auf unsere Anregungen aufgeschlossen reagieren wird“, meint die Bayreuther Doktorandin. „Denn der normale Internet-Nutzer, der schnelle und präzise Antworten auf seine Suchanfragen erwartet, wird mit hoher Wahrscheinlichkeit davon profitieren – insbesondere dann, wenn SPARQL und die durch RDF definierten Standards sich im Web weiter durchsetzen.“

Ihren Optimierungsvorschlag haben die Bayreuther Informatiker mit Kollegen an der Universität von Chile und der Päpstlichen Katholischen Universität von Chile – beide in der Hauptstadt Santiago – abgestimmt. Dort hatten sich drei Forscher ebenfalls intensiv mit SPARQL befasst und waren dabei auf die gleichen Probleme gestoßen. „Dass zwei voneinander unabhängige Forschungsteams denselben Optimierungsvorschlag einbringen, wird die Erfolgsaussichten beim World Wide Web Consortium erhöhen“, glaubt Martens, der sich auch in seinen künftigen Forschungsarbeiten für Verbesserungen im World Wide Web engagieren will.

Veröffentlichung:

Katja Losemann, Wim Martens:

The complexity of evaluating path expressions in SPARQL,

in: PODS '12 Proceedings of the 31st Symposium on Principles of Database Systems, Association for Computing Machinery (ACM), New York 2012, pp. 101-112

DOI (Link): [10.1145/2213556.2213573](https://doi.org/10.1145/2213556.2213573)



Kontaktadresse:

Prof. Dr. Wim Martens
Theoretische Informatik
Universität Bayreuth
D-95540 Bayreuth
Tel.: +49 (0)921 55 7750
E-Mail: wim.martens@uni-bayreuth.de

Text und Redaktion:

Christian Wißler M.A.
Stabsstelle Presse, Marketing und Kommunikation
Universität Bayreuth
D-95440 Bayreuth
Tel.: 0921 / 55-5356 / Fax: 0921 / 55-5325
E-Mail: mediendienst-forschung@uni-bayreuth.de

Foto Seite 3: Wim Martens; zur Veröffentlichung frei.

In hoher Auflösung zum Download:
www.uni-bayreuth.de/presse/images/2012/209