XML und Semantik

Eckhardt Schön

Technische Universität Ilmenau, Institut für Praktische Informatik und Medieninformatik

Abstract. Der Beitrag beschreibt, in welcher Weise XML, XML -Schema, RDF und RDF -Schema geeignet sind, Semantik in maschinenverständliche r Weise zu kodieren.

Im Zeitalter des weltweiten Nachrichten - und Informationsaustausches wird es immer wichtiger, dass die Nachrichten nicht nur korrekt von einer Applikation zu einer anderen übertragen werden, sondern dass diese Nachrichten auch von der Gegenseite „verstanden“ werden und zwar ohne menschliche Eingriffe. Die Semantik muss also maschinenverständlich sein.

Die Maschinenverständlichkeit kann dadurch erreicht werden, dass die Nachrichten in einer sehr strengen Syntax übermittelt werden, die g enau auf die korrespondierenden Applikationen zugeschnitten ist. Das hat den Nachteil, der geringen Flexibilität. Beispiele sind EDI-FACT und viele binär kodierte Austauschformate.

XML bietet eine weit größere Flexibilität und hat sich deshalb schnell als ein bevorzugtes Nachrichtenaustauschformat etabliert. Nachrichten können damit in stark strukturierter Form und doch flexibel kodiert werden. Applikationen „verstehen“ jedoch den Inhalt einer beliebigen XML -Nachricht nicht, da sie die Bedeutung der Tags (a uch wenn diese mittels XML -Namensräumen weltweit eindeutig sind) nicht kennen. Auch die neu eingeführten XML -Schemata ändern daran nichts, da sie lediglich eine Art Grammatik darstellen, die die syntaktische Korrektheit der Nachrichten garantiert.

Die oft notwendige Transformation von XML mit Hilfe von XSL(T) erfordert eine hohe Aufmerksamkeit, damit das XML -Dokument nicht nur formal richtig transformiert wird (d.h. einem anderen XML -Schema genügt), sondern auch die eventuell kodierte Semantik der Informati onen nicht verfälscht wird. Der semantische Gehalt eines XML-Dokumentes nimmt bei Transformationen im allgemeinen ab.

Einen Ansatz zur Kodierung von maschinenverständlichen Informationen stellt das Resource Description Framework (RDF) dar. Es wurde geschaf fen, um Ressourcen ganz allgemeinen Charakters durch ihre Eigenschaften und die Beziehungen untereinander zu beschreiben. Die Aussagen (*Statements*) können in XML -Syntax serialisiert werden. Ergänzend stellen RDF -Schema quasi Wörterbücher für bestimmte Domä nen zur Verfügung, die die verwendeten Begriffe maschinenverständlich definieren. RDF und RDF -Schema bilden die Basis für das vom W3C angestrebte Semantic Web, das die Ressourcen des bisherigen Internets besser erschließen soll. Dazu ist eine Kodierung der Semantik notwendig, die weltweit von Computern „verstanden“ werden kann.