

# Big Text: von Sprache zu Wissen

Gerhard Weikum  
Max-Planck-Institut für Informatik &  
Universität des Saarlandes  
D-66123 Saarbrücken  
weikum@mpi-inf.mpg.de

**Abstract:** Nachrichten, soziale Medien, Webseiten und Dokumente in Unternehmen beinhalten eine Fülle wertvoller Informationen in textueller oder gesprochener Form. Ein entscheidender Schritt, um diese Form von Big Data für Analytik und Entscheidungsfindung zu nutzen, ist die Identifikation von Entitäten wie Personen, Organisationen oder Produkten sowie von Beziehungen zwischen Entitäten. Dieser Schritt ermöglicht auch das Verknüpfen von strukturierten und unstrukturierten Daten für Big-Data-Anwendungen. Allerdings steht dieser Schritt vor dem herausfordernden Problem, dass Namen von Entitäten und relationale Phrasen oft in hohem Maße mehrdeutig sind. Man benötigt eine skalierbare, hochwertige Lösung für die Disambiguierung von Namen in Texten und heterogenen Tabellen. Dieser Vortrag diskutiert den Stand der Kunst, Anwendungen und offene Probleme zu diesem Thema. Er diskutiert außerdem die daraus resultierenden Perspektiven für das umfassendere Thema der Big-Data-Analytik.