

## TOP-Forschungsprojekte 2017

**Immersive Web Observatory –  
Das Web als Datenbasis intelligenter und immersiver Anwendungen**

Juniorprofessur:	Big Data Analytics Prof. Dr. Matthias Hagen Fakultät Medien
Laufzeit:	1. Mai 2017 bis 31. Dezember 2017
Drittmittelgeber:	BMBF
Fördersumme:	2.511.015,05 Euro

**Beschreibung:**

Das World Wide Web dient heute als Grundlage zur Beantwortung zahlreicher unternehmerischer und gesellschaftlicher Fragestellungen. Durch die gezielte Erhebung, Auswertung und Visualisierung von Web-Daten werden im Rahmen der Business Intelligence entscheidungsrelevante Informationen gewonnen, in den „digitalen“ Sozialwissenschaften aktuelle gesellschaftliche Trends und soziale Netzwerke analysiert und im Forschungsfeld der Digital Humanities kulturelle und geschichtliche Fragestellungen mit Hilfe digitaler Medien studiert. Darüber hinaus ist das Web Forschungsgegenstand der Informatik zur Entwicklung von Informationssystemen, die neben Endbenutzern zunehmend auch gezielt auf Experten der vorgenannten Bereiche zugeschnitten werden. Für das maschinelle Lernen stellt das Web eine Sammlung von Fallbeispielen für zahllose alltägliche Problemstellungen dar, die zum Training maschineller Lernverfahren eingesetzt werden. Hieraus entwickelt sich insbesondere das rapide wachsende Forschungsfeld für intelligente Assistenztechnologien. Diese adressieren – oft auf Basis von aus dem Web extrahiertem Wissen – Herausforderungen im Bereich des Cognitive Computing mit dem Ziel, menschlichen Maßstäben genügend urteilen und agieren zu können.

Trotz der breiten Anwendungsmöglichkeiten, das *Web als Korpus* für Wirtschaft, Wissenschaft und Gesellschaft zu nutzen, gibt es bislang kaum Anbieter, die das Web in größerem Umfang hierfür verfügbar machen. Einzig große Internet-Unternehmen wie Google, Microsoft, Yandex und Baidu pflegen eigene Webcrawls als Grundlage ihrer Suchmaschinen sowie zum Zweck der Erforschung und Entwicklung intelligenter Assistenten. Diese Crawls stehen Dritten jedoch nicht zur Verfügung; kleinen und mittleren Unternehmen fehlen die Ressourcen, das Web in großem Umfang selbst zu crawlen und die Crawls aktuell zu halten. Demzufolge fehlt oftmals auch das hierfür notwendige Know-how, so dass Marktchancen, die andernfalls in greifbarer Nähe wären, regelmäßig ungenutzt bleiben. Ähnliches gilt für wissenschaftliche Einrichtungen, in denen die Analyse des Webs zwar in vielen Bereichen zu tiefgreifenden Einsichten führen könnte, die aber nicht die erforderliche Infrastruktur und – mit Ausnahme der Informatik – auch nicht die nötige Expertise besitzen, das Web als Korpus zu nutzen.

**Kontakt:**

Bauhaus-Universität Weimar  
Big Data Analytics  
Prof. Dr. Matthias Hagen  
matthias.hagen@uni-weimar.de

Bauhausstraße 11  
99423 Weimar  
03643 / 58 37 46

## TOP-Forschungsprojekte 2017

Das Immersive Web Observatory (IWO) stellt ein sehr großes Crawl des Webs als Korpus zur Verfügung und erfüllt damit folgende Zwecke:

- Infrastruktur für Projekte in allen oben genannten Bereichen mit besonderem Fokus auf die Querschnittsthemen Retrieval, Mining und Visualisierung. Da für einzelne Projekte eine solche Infrastruktur ad hoc nicht erschaffen werden kann, ermöglicht das IWO Kooperationen mit kleinen und mittleren Unternehmen, die bislang aufgrund fehlender Hardware und zu großen Vorlaufzeiten bei der Beschaffung von Daten nicht möglich waren.
- Ausbildung Studierender und Promovierender zu Data Scientists, also zu Experten in der Verarbeitung, Analyse und Visualisierung von Big Data im Allgemeinen und des Webs im Speziellen. Ein Schwerpunkt wird das immer wichtigere Cognitive Computing sein, um hier in ausgewählten Anwendungsbereichen zu den weltweit führenden internationalen Unternehmen aufzuschließen.
- Plattform für Wissens- und Technologietransfer zu kleinen und mittleren Unternehmen der Region. Der Transfer geschieht hauptsächlich im Rahmen von Projektpartnerschaften, aber auch mittels Demonstratoren und durch Open-Access-Veröffentlichungen.

Das IWO ergänzt die vorhandene Infrastruktur um einen für die fortwährende Bereithaltung und die schnelle Darstellung von sehr großen Datenmengen geeigneten aktuellen Clustercomputer und GPU-Cluster, um entsprechende Displaytechnologie sowie um geeignete Endgeräte. Insbesondere wird die in vorangegangenen Projekten beschaffte Hardware nicht ersetzt, sondern integriert: zu diesem Zweck wird die Netzwerkinfrastruktur so ausgelegt, dass an allen notwendigen Standorten eine auf die Verarbeitung und Visualisierung großer Datenmengen in Realzeit ausgelegte Bandbreite und Geschwindigkeit zur Verfügung steht. Zum Zeitpunkt der Inbetriebnahme laufende Drittmittelprojekte sowie in naher Zukunft startende Projekte werden direkt von der Verfügbarkeit des IWO profitieren.

Weitere Informationen:

<https://www.uni-weimar.de/en/media/chairs/computer-science-and-media/webis/research/>

### Kontakt:

Bauhaus-Universität Weimar  
Big Data Analytics  
Prof. Dr. Matthias Hagen  
matthias.hagen@uni-weimar.de

Bauhausstraße 11  
99423 Weimar  
03643 / 58 37 46