

# Webkorpora in Computerlinguistik und Sprachforschung

Workshop am Institut für Deutsche Sprache (IDS) in Mannheim  
R5 6-13, 68161 Mannheim, Großer Vortragssaal

**27. - 28.09.2012**

Der Workshop thematisiert Gewinnung, Aufbereitung und Analyse von Webkorpora für computerlinguistische Anwendungen und für die sprachwissenschaftliche Forschung. Ein Schwerpunkt sind dabei die speziellen Anforderungen, die sich gerade im Hinblick auf deutschsprachige Ressourcen ergeben. Im Fokus steht weiterhin die Nutzung von Webkorpora für empirisch gestützte Sprachforschungen, also beispielsweise für sprachstatistische Analysen, für Untersuchungen zur Sprachlichkeit in der internetbasierten Kommunikation oder für die korpusgestützte Lexikographie.



## Donnerstag, 27.09.2012

- 14:00 **Begrüßung / Eröffnung**  
Roman Schneider (IDS Mannheim)
- 14:10 **Webkorpora als qualitätsgesicherte Forschungsdaten**  
Marc Kupietz, Harald Längen (IDS Mannheim)
- 14:40 **Establishing Gold Standards for Web Corpora**  
Felix Bildhauer, Roland Schäfer (FU Berlin)
- 15:10 **Kaffeepause**
- 15:30 **Die Leipzig Corpora Collection: Datenbeschaffung, Textaufbereitung und Qualitätssicherung für Korpora in 150 Sprachen**  
Uwe Quasthoff, Dirk Goldhahn, Thomas Eckart (Universität Leipzig)
- 16:00 **Kontrastive Analyse der morphosyntaktisch annotierten Wikipedia-Korpora in EuroGr@mm**  
Karl Hagen Augustin (IDS Mannheim)
- 16:30 **Kaffeepause / Postersession**
- 16:50 **Sie fangen dort an, wo wir aufhören: Möglichkeiten der Nutzung von Webkorpora in Peer Production Communities**  
Claudia Müller-Birn (FU Berlin)
- 17:20 **Diskussion: Deutsche Webkorpora: Stand der Kunst – Desiderate – Perspektiven**  
Moderation: Angelika Storrer (TU Dortmund), Roman Schneider (IDS Mannheim), Alexander Mehler (Universität Frankfurt am Main)

## Freitag, 28.09.2012

- 9:00 **Web-basierte Extraktion und Repräsentation linguistischer Netzwerke**  
Alexander Mehler, Vincent Esche (Universität Frankfurt am Main)
- 9:30 **Webkorpora für die automatische Akquisition lexikalisch-semantischen Wissens**  
Sabine Schulte im Walde (Universität Stuttgart)
- 10:00 **Postersession**
- 10:30 **Kaffeepause / Postersession**
- 10:50 **Gewinnung, Aufbereitung und Analyse von Korpora zu Genres internetbasierter Kommunikation: Herausforderungen und Perspektiven**  
Michael Beißwenger (TU Dortmund), Stefan Evert, Torsten Zesch (TU Darmstadt)
- 11:20 **Aufbau und Annotation eines Referenzkorpus zur deutschsprachigen internetbasierten Kommunikation (DeRiK-Projekt)**  
Michael Beißwenger (TU Dortmund), Lothar Lemnitzer (BBAW Berlin)
- 11:50 **Kollaborative Erstellung eines annotierten Korpus als Grundlage für die Anwendung statistischer Ansätze der automatischen Sprachverarbeitung auf internetbasierte Kommunikation**  
Alexander Geyken, Bryan Jurish, Kay-Michael Würzner (BBAW Berlin)
- 12:20 **Kaffeepause, anschl. Abschlussdiskussion und Ausblick**

<http://hypermedia.ids-mannheim.de/webkorpora/>