

Dr. Dirk Ahlers

Geographically Focused Web Information Retrieval

Suchmaschinen sind heute das meistgenutzte Mittel, um Zugang zu Informationen im World Wide Web zu gewinnen. Geografisches Information Retrieval erweitert das Konzept der Suche, um auch den Zugang zu ortsbasierten Informationen zu ermöglichen. Während moderne Suchmaschinen bereits für eine Vielzahl von Anfragen befriedigende Ergebnisse liefern, so gilt dies nur bedingt für ortsbasierte Informationen. Doch 20% aller Anfragen an Suchmaschinen haben einen Ortsbezug. Im Web findet sich eine Vielzahl hochwertiger Informationen mit Ortsbezug. In der fokussierten Aufbereitung von geographischen Inhalten zur gezielten Beantwortung entsprechender Anfragen liegt daher ein großes Potential.

Ziel der Dissertation ist die Untersuchung von effizienten Verfahren zum Aufbau einer geographisch fokussierten Suchmaschine, um den Ortsbezug von frei verfügbaren, allgemeinen Webseiten zu nutzen. Dabei werden die Aspekte Crawling, Parsing und Ranking berücksichtigt. Die Arbeit baut auf Grundlagen des Geographischen Information Retrieval auf, dessen Verfahren genutzt und weiterentwickelt werden. Dabei werden relevante Ortsinformationen explizit in ihrer geographischen Semantik berücksichtigt. Durch die Erschließung von im Volltext von Webseiten enthaltenen Informationen, die Erkennung und Extraktion von Ortsbezügen und deren genaue geographische Positionierung erlaubt die Arbeit die Nutzung der geographischen Dimension als Ordnungskriterium neben dem Volltext der Webseiten und ermöglicht somit eine geographische Sicht auf das WWW. Der zentrale Beitrag der Arbeit ist die Entwicklung einer geographischen Fokussierung in zwei Teilbereichen. Zum einen werden effiziente Strategien entwickelt, um einen Webcrawler fokussiert auf ortsbezogene Seiten zu leiten und damit verstärkt relevante Dokumente aufzufinden. Hierzu wird das so genannte Focused Crawling auf die geographische Suche adaptiert, um die Suchmaschine auf einen Ausschnitt des Web mit den gewünschten Ortsinformationen einschränken zu können. Dies basiert auf der Bewertung und Prognose des Auftretens relevanter Seiten. Zum anderen wird ein geographischer Parser für die Extraktion von Ortsinformationen in einer hohen Granularität entwickelt, der das präzise, adressgenaue Identifizieren von Ortsinformationen und die Verortung von Ergebnissen ermöglicht. Es wird dabei ein verifizierender Parser realisiert, der unter anderem durch die Einbindung externen Domänenwissens eine hohe Ergebnisqualität gewährleistet. Analyseverfahren zur Nutzung der geographischen Dimension der Daten für die Anfrageverarbeitung sowie die Darstellung einer geeigneten Architektur für die geographische Suchmaschine runden die Arbeit ab.

XVI, 214 S., Edewecht 2012, € 39,80

ISBN-13: 978-3-939704-69-0

OlWIR Verlag –
Oldenburger Verlag für Wirtschaft,
Informatik und Recht
Rudolf-Kinau-Str. 54, 26188 Edewecht
Bestellungen an: mail@olwir.de

