

# BACHELORARBEIT:

## „ENTWICKLUNG VON METHODEN UM DEN WESENTLICHEN (TEXT-)INHALT WEB-BASIERTER ARTIKEL VON NEBENBESTANDTEILEN UNTERSCHIEDEN ZU KÖNNEN“

### DEINE AUFGABE:

**in4s** ist ein junges, frisch gegründetes Technologie-Unternehmen direkt auf dem Campus der OVGU mit Wurzeln in der Data & Knowledge Engineering Arbeitsgruppe. Wir entwickeln unter anderem ein Werkzeug um Webportale und Newsletter zu crawlen und ihre Inhalte in aufbereiteter Form zur Verfügung zu stellen.

**Deine** Aufgabe ist die Entwicklung von Methoden, um den „Hauptartikel“ aus web-basierten Artikeln extrahieren zu können. Nebenbestandteile, z. B. Navigationselemente, Werbeblöcke oder eingblendete dynamische Inhalte wie Newsticker, sollen zuverlässig gefiltert werden können.

### RELEVANTE TECHNOLOGIEN / RELEVANTE VORLESUNGEN

- ✓ **Java / Apache Tika**
- ✓ **Information Retrieval**

### INTERESSIERT?

Sende Deine Bewerbung an: [internships@in4s.de](mailto:internships@in4s.de)