



## Medical Services

### DocCheck Medical Services GmbH

DocCheck ([www.doccheck.com](http://www.doccheck.com)) ist die größte und am schnellsten wachsende Healthcare-Community für medizinische Fachkreise in Europa. DocCheck verbindet über 800.000 registrierte Nutzer aus dem medizinischen und pharmazeutischen Bereich. Vom Arzt und Apotheker bis zum Studenten ist alles vertreten. Mit dem kostenlosen DocCheck-Passwort erhalten Healthcare-Professionals kostenlos Zugang zu über 2.500 medizinischen Websites.

Im Durchschnitt erreicht DocCheck mehr als 2,3 Mio. Visits und ca. 7 Mio. Page Impressions pro Monat (nach IVW). Damit belegt DocCheck die Spitzenposition für Healthcare-Fachmedien. DocCheck bietet den Mitgliedern geschlossene Communities/Netzwerke sowie vielfältige Services rund um Fortbildung, Karriere und fachlichen Austausch an.

### Abschlussarbeiten bei DocCheck

Wir bieten Studenten interessante Themen zu Abschlussarbeiten mit IT-Schwerpunkt an – natürlich rund um das Thema Healthcare! ;-)

Bei Interesse an den unten aufgeführten Themen können Sie gerne Kontakt zu uns aufnehmen:

DocCheck Medical Services GmbH  
Andrea Krieger  
Human Resources Management

[karriere@doccheck.com](mailto:karriere@doccheck.com)  
[www.doccheck.ag](http://www.doccheck.ag)



## Medical Services

### 1. Datenmanagement DocCheck-Content

DocCheck bietet den Usern auf über 100.000 Seiten eigene redaktionelle Inhalte sowie User-Generated Content an. Dazu gehören und anderem ein Mediawiki (mit med. Fachinformationen), News, Filme, Bilder, Praxis- und Medizinbedarf (eCommerce), Job- und Fortbildungs-Angebote sowie eine Community für Healthcare-Professionals.

Zur Strukturierung der Inhalte, Verbesserung der Wiederauffindbarkeit (SEO) und semantische Verknüpfung von Inhalten („weitere Inhalte zu dem Thema ...“) sollen einheitliche und übergreifende Schnittstellen zur Klassifizierung, Verschlagwortung und Internationalisierung durch Redakteure und User geschaffen werden.

Folgende Themen können in Abschlussarbeiten bearbeitet werden:

- **Definition/Erfassung von Meta-Daten**  
Erarbeitung eines übergreifenden Konzepts zur Klassifizierung von Inhalten durch eine einheitliche Zuordnung von Meta-Daten (inhaltlich/technisch). Dabei sind besonders Konzepte auf Grundlage vorhandener Terminologien, Ontologien, Codierungen und Relationen wie „Medical Subject Headings (MeSH)“ bzw. „Unified Medical Language System (UMLS)“ zur Klassifizierung und Strukturierung von medizinisch-wissenschaftlichen Inhalten zu berücksichtigen.
- **Automatisierte Textanalyse und Verschlagwortung**  
Entwicklung eines Verfahrens zur automatischen Verschlagwortung von Texten/Webseiten. Anhand einer Textanalyse sollen relevante Meta-Daten ermittelt werden.
- **Verschlagwortungshilfen**  
Entwicklung von Hilfen („auto suggest/complete“) bei der Verschlagwortung zur Qualitätssicherung, z.B. durch thematisch passende Vorgaben/Vorschläge (siehe MeSH-Terms, UMLS) oder geführte Prozesse abhängig von Medium (Bild, Text, ...), Fachbereich (Chirurgie, Pharmazie, ..), Organsystemen etc.
- **Owner- vs. Community Verschlagwortung**  
Entwicklung eines Konzepts zur kollaborativen Verschlagwortung/Klassifizierung durch Redakteure, Admins und User („Community“). Ziel ist die Generierung zusätzlicher Inhalte, bzw. Klassifizierung der Inhalte durch zusätzliche User – gleichzeitig muss eine Verunstaltung der Inhalte durch Vandalismus vermieden werden.



## Medical Services

- **Qualitätssicherung**  
Kontrolle der Inhalte und Meta-Daten durch User (peer group reviews, Studenten-Quizzes, kollaborative Ansätze wie Amazon Mechanical Turk) sowie Internationalisierung mit Übersetzungshilfen und semantischen Relationen.
- **Automatisierte Ermittlung „ähnlicher Inhalte“**  
Implementierung eines Verfahrens zur automatisierten Verknüpfung ähnlicher, d.h. thematisch passender Inhalte (=Webseiten) aus dem DocCheck Content Pool, die dem Besucher einer Content-Seite angeboten werden können (Stichwort: „Weitere Inhalte zu dem Thema“).

## 2. Verfahren zur Bewertung von User-Eingaben: Datenqualifizierung und -validierung

Entwicklung von Methoden zur Sicherstellung von Konsistenz, Aktualität, Validität und Qualität von nutzergenerierten Daten, z.B.:

- Klassifizierung von Adressdaten (passt PLZ zu Straße/Ort, Schreibfehler; ist die Adresse veraltet, ist ein Nutzerdatensatz ein Zweitaccount eines bestehenden Nutzers)
- Bewertung der Kompetenz des Autors von fachlichen Beiträgen (wie groß ist die Wahrscheinlichkeit, dass ein Beitrag des Autors fachlich korrekt und hilfreich ist/sein wird; auch zwecks "Scouting" von Autoren/Fachratgebern etc.)
- Bewertung der Qualität von Marktforschungsdaten (wurde die Umfrage sorgfältig oder "nebenbei" ausgefüllt)
- Erkennung von "auffälligem" und unerwünschten Verhalten (störendes Verhalten in der Community; betrügerische Teilnahme an geldwerten Aktionen)

Gefragt sind dabei sowohl vollautomatisierte Verfahren zur dynamischen Kontrolle von Massendaten (Beispiel: automatische Adress-Korrektur) als auch Assistenzsysteme zur Vorqualifizierung der Ergebnisse für die Entscheidung durch einen Menschen (Beispiel: Moderation in der Community, Verschmelzung von doppelten Datensätzen, "Rauswurf" eines Nutzers).



## Medical Services

Berührte Fachgebiete:

- Fuzzy Logic
- Data Mining
- Statistik
- Geotagging

Voraussetzungen:

- Kenntnisse der objektorientierten Programmierung (optimal: Erfahrung mit PHP)
- Grundkenntnisse in SQL
- Grundkenntnisse in Statistik (alternativ: sehr gutes mathematisches Verständnis)

### 3. Entwicklung eines Filter für Kommentar Spam

Täglich gehen hunderte Spam Einträge über die DocCheck Kontakt-, und Bestell-Formulare, Foren-Beiträge, ... ein. Es werden - wie aus E-Mail-Spam bekannt - Medikamente, teure Uhren und Autoersatzteile beworben. Kommentar-Spam verstopft Ticket-Systeme, nervt die User und führt zu einer SEO-Abwertung (Stichwort „Bad Neighborhood“).

Wir setzen bereits einen selbst entwickelten Spam-Filter ein, der verschiedene Parameter und statistische Methoden kombiniert. Die bereits vorhandenen Daten können zur Analyse, bzw. Training und Testing des Filters verwendet werden.

Ziel ist die Entwicklung eines verlässlichen Filters (d.h. statistischer Methoden zur Textanalyse und Klassifizierung) zur Erkennung vom Kommentar Spam und die Bereitstellung einer API (Webservice) zur Einbindung des Filters in bestehende interne und externe Auftritte.

Berührte Fachgebiete:

- Fuzzy Logic
- Data Mining
- Statistik

Voraussetzungen:

- Kenntnisse der objektorientierten Programmierung OOP/OOD (optimal: Erfahrung mit PHP)
- Grundkenntnisse in SQL
- Grundkenntnisse in Statistik (alternativ: sehr gutes mathematisches Verständnis)