

## Text Mining bei der Deutschen Nationalbibliothek



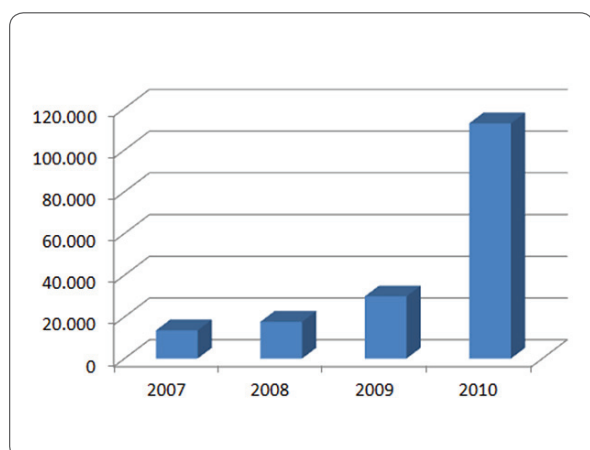
Ein Beispiel für den Einsatz Information Discovery ist die automatische Metadaten-Generierung. Hier bieten wir ein Werkzeug zur gezielten Extraktion von Informationen aus Dokumenten. Die Lösung identifiziert einzelne Informationseinheiten sowie relevante Fakten und Zusammenhänge, Dokumente werden automatisch in Fachbereiche klassifiziert und relevante Schlagworte identifiziert.

### Deutsche Nationalbibliothek

Die Deutsche Nationalbibliothek hat die für Deutschland einzigartige Aufgabe, lückenlos alle deutschen und deutschsprachigen Publikationen ab 1913 zu sammeln, dauerhaft zu archivieren, umfassend zu dokumentieren und bibliografisch zu verzeichnen sowie der Öffentlichkeit zur Verfügung zu stellen. Insgesamt beläuft sich der Gesamtbestand der Deutschen Nationalbibliothek derzeit auf rund 26 Millionen Einheiten.

### Ausgangslage

Die Aufnahme elektronischer Publikationen in bibliothekarische Sammlungen hat zu stark anwachsenden Dokumentmengen geführt (Grafik), die allein durch Formen intellektueller Bearbeitung nicht zu bewältigen sind.



Ziel der Zusammenarbeit war daher, maschinelle Erschließungsverfahren zur Kategorisierung und Verschlagwortung der Bestände einzuführen, um

- Erschließungsaufwände drastisch zu verringern
- bisher auftretende Erschließungslücken in hoher Qualität zu überwinden
- das Finden von Informationen besser zu unterstützen

### Schlagworterkennung

Durch den Einsatz von frei wählbaren Wortlisten und Terminologien werden automatisch beschreibende Schlagworte aus den Texten zur inhaltlichen Strukturierung und zur besseren Wieder auffindbarkeit generiert.

### Dokumentenklassifizierung

Artikel und Textdokumente werden zur Indexierung automatisch anhand eines frei definierbaren Kategoriensystems klassifiziert. Beispielsweise können Werke so automatisch den jeweiligen Ressorts (z.B. „Wirtschaft“, „Politik“) zugewiesen werden.

### Ergebnis

„Die Deutsche Nationalbibliothek stellt die Anforderung, große Mengen an textbasierten, inhaltlich und formal unterschiedlichen, mehrsprachigen, digitalen Objekten in hoher Qualität automatisch zu erschließen“, sagt die Projektverantwortliche Christa Schöning-Walter. Mit der Lösung von Averbis konnten diese Aufgaben erfolgreich in die Praxis umgesetzt werden.