

Pressemitteilung 12. Dezember 2011

ca. 553 Worte: ARD-PAN führt Textmining-System von picturesafe ein

# ARD-PAN führt Textmining-System von picturesafe ein

**Die sieben ARD-Rundfunkanstalten des PresseArchivNetzwerks (PAN) nutzen ein hochleistungsfähiges dio:semantic-System für die inhaltliche Erschließung ihres Presse-Archivs.**

Am Standort Köln betreiben die Rundfunkanstalten (DW, NDR, RB, rbb, SR, SWR und WDR) gemeinsam die PAN-Textdatenbank für die Dokumentation und Recherche von Presseartikeln und anderen Texten. Im PAN erfolgt eine arbeitsteilige Dokumentation von ca. 100 Publikationen für den gemeinsamen Dokumentenpool. Zusätzlich werden in jeder PAN-Rundfunkanstalt die Publikationen des jeweiligen Sendegebiets bearbeitet. Dieser Wissensfundus steht allen Mitarbeitern im Intranet zur Verfügung. Die fundierte Erschließung nach PAN-Klassifizierung und Regelwerk ermöglicht eine schnellere und präzisere Recherche.

Das PresseArchivNetzwerk hat Ende 2010 unter Federführung des Westdeutschen Rundfunks (WDR) eine neue, von allen Mitgliedern gemeinsam genutzte und mehrere Millionen Artikel umfassende Textdatenbank eingeführt. Ergänzend wurde die von picturesafe gelieferte Textmining-Komponente dio:semantic installiert und nach Abschluss der Abnahmetests Mitte 2011 in den Produktionsworkflow übernommen.

Das neue picturesafe-System umfasst die maschinelle Vorerschließung auf Basis der PAN-Klassifikation und des Regelwerks, sowie die Realisierung eines Statistik-Moduls für Qualitätsauswertungen. Darüber hinaus leistet es die Pflege der für das Textmining relevanten Filter, Parameter und das Update/Retraining der Klassifikation und der verwendeten Named Entities.

Das hierarchische PAN-Klassifikationssystem ist Basis für eine intelligente Recherche in der Textdatenbank. Über ausgereifte Suchfunktionalitäten können Artikel in den Sprachen Deutsch, Englisch, Französisch und Spanisch treffsicher und mit geringem Zeitaufwand gefunden werden. Relevantere Suchergebnisse werden auch dadurch erzielt, dass das PAN-Klassifikationssystem mit Datenfeldern der Sacherschließung, Aspekten und freien Deskriptoren in Beziehung steht. Weiterhin werden relevante Orte, Personen und Geografika anhand von Regeln identifiziert. Das dio:semantic Textmining-System bildet diese Inhaltsstruktur vollständig ab. Um einen Artikel im PAN-Klassifikationssystem vollständig zuzuordnen, werden vier Wissensmodelle pro Sprache verwendet.

Digitale Inhalte  
weltweit kommerziell  
nutzen

Die picturesafe GmbH ist einer der führenden Anbieter von Full Service Solutions rund um Digital Asset und Brand Management mit Sitz in Hamburg, Hannover und Boston, USA.

Seit 1993 konzipiert und realisiert picturesafe internetbasierte Lösungen für weltweit agierende Großkonzerne und Verlage.

[picturesafe media/data/bank](http://picturesafe.media/data/bank)  
Pressekontakt: Herr Oliver Bauer

Simon-von-Utrecht-Str. 31-37  
20359 Hamburg  
E-Mail: [obauer@picturesafe.de](mailto:obauer@picturesafe.de)

 picturesafe

## Pressemitteilung 12. Dezember 2011

ca. 553 Worte: ARD-PAN führt Textmining-System von picturesafe ein

Bei vier unterstützten Sprachen im PAN-System werden insgesamt 16 Klassifizierung-Wissensmodelle eingesetzt, um eine optimale Analyse der Presseartikel zu gewährleisten.

Trotz der komplexen Aufgabe ist die notwendige Infrastruktur für die Anbindung des picturesafe-Textmining-Systems verhältnismäßig einfach gehalten, was zur Kostensenkung beiträgt. Das speicher- und performance-optimierte Verfahren ermöglicht es dem dio:semantic pro Stunde über 300.000 Artikel zu analysieren oder in der Wissensbasis zu trainieren.

**Die Ergebnisqualität des Textmining-Systems ist von zentraler Wichtigkeit. Schon im Vorfeld wurden die Pressedaten des PAN vom Fraunhofer Institut wissenschaftlich analysiert, und damit die Qualitätsstandards für die Auftragsvergabe festgelegt. picturesafe erfüllte diese Standards und konnte sie im Produktivbetrieb nochmals steigern.**

„Besonders erfreulich ist die produktive und angenehme Zusammenarbeit zwischen den Mitarbeitern des PAN und dem picturesafe Team“ sagt Entwicklungsleiter Gerhard Wendorf. „Neben der technischen Leistungsfähigkeit des Systems, ist es wichtig das wir in allen Bereichen einen guten Draht zum Kunden haben. Auch im dokumentarischen Umfeld funktioniert die Abstimmung reibungslos. Unsere Diplom-Bibliothekare und Dokumentare sprechen die gleiche Sprache wie die Mitarbeiter in den Fachabteilungen des Kunden.“

Apropos Sprache: dio:semantic beinhaltet mehrsprachige Linguistikmodule zur Analyse und Aufbereitung unstrukturierter Dokumente, welche die Basis für Klassifizierung, Clusterung, Ähnlichkeitsanalyse und zur Generierung von Tag Clouds und Textsummarys bilden.

### Über picturesafe:

Die picturesafe media/data/bank GmbH entwickelt seit 1992 Softwarelösungen zum Management von Mediendaten und bietet Full Service von der Aufnahme bis zur Auslieferung digitaler Daten in Form von Texten, Bildern, Video- und Tondaten. Dazu werden marktetaillierte Produkte deren Online- sowie DTP-Schnittstellen sowie die Entwicklung unternehmensspezifischer Individuallösungen angeboten. Die Produkte und Services von picturesafe gestalten Informations- und Medienproduktionsprozesse und steigern nachhaltig deren Produktivität. picturesafe hat seinen Firmensitz in Hannover und eine Niederlassung in Hamburg, Deutschland. Mehr Informationen: [www.picturesafe.de](http://www.picturesafe.de)

picturesafe media/data/bank GmbH  
Pressekontakt: Oliver Bauer ([obauer@picturesafe.de](mailto:obauer@picturesafe.de))  
Simon-von-Utrecht-Str. 31-37, 20359 Hamburg

## Digitale Inhalte weltweit kommerziell nutzen

Die picturesafe GmbH ist einer der führenden Anbieter von Full Service Solutions rund um Digital Asset und Brand Management mit Sitz in Hamburg, Hannover und Boston, USA.

Seit 1993 konzipiert und realisiert picturesafe internetbasierte Lösungen für weltweit agierende Großkonzerne und Verlage.

picturesafe media/data/bank  
Pressekontakt: Herr Oliver Bauer

Simon-von-Utrecht-Str. 31-37  
20359 Hamburg  
E-Mail: [obauer@picturesafe.de](mailto:obauer@picturesafe.de)

