

Wie das SDSC-BW die richtigen Algorithmen für erfolgreiche Big Data-Analysen aufspürt.



# Optimierte Werkzeuge für die Big Data-Profis.

„Unser gemeinsames Projekt mit dem SDSC-BW hat uns geholfen, die richtigen Algorithmen für die Weiterentwicklung unseres Angebots zu finden.“

Dr. Matthias Keller, Head of Data-Science

The logo for echobot, consisting of three vertical bars of increasing height followed by the word "echobot" in a lowercase, sans-serif font.

Success-Story



Success-Story online unter [www.sdsc-bw.de/echobot](http://www.sdsc-bw.de/echobot)

## Das Smart Data Solution Center Baden-Württemberg

„Zusammenhänge erkennen. Potentiale nutzen.“ Unter diesem Motto ermöglicht das SDSC-BW kleinen und mittelständischen Unternehmen die Nutzung von Smart Data-Technologien. Und das als neutrale, unabhängige und vom Land Baden-Württemberg geförderte Institution.

In unseren kostenlosen Potentialanalysen zeigen wir Vorteile und Nutzungsmöglichkeiten von Daten – die in jedem Unternehmen jeder Branche anfallen – und helfen, diese zu bewerten.

Wie wertvoll sind Ihre Daten? Finden Sie es heraus unter [www.sdsc-bw.de](http://www.sdsc-bw.de)



**Der Kunde** Echobot Media Technologies GmbH analysiert für ihre Kunden täglich Millionen Texte von Websites, sozialen Medien und Nachrichtenseiten. Die Analysen helfen zum Beispiel bei Marketing, Vertrieb oder der Optimierung der Pressearbeit. Im Jahr 2011 als Start-Up mit nur wenigen Mitarbeitern gegründet, hat sich Echobot nach sechs Jahren zu einem erfolgreichen Unternehmen entwickelt. Heute ist das Unternehmen eine Business Information Intelligence Plattform, die Milliarden digitaler Inhalte analysiert und ihre Kunden in Echtzeit mit geschäftsrelevanten Daten und Informationen versorgt.

**Die Daten** Um digitale Inhalte im Netz zu analysieren, muss Echobot im ersten Schritt eine automatische Klassifikation der erfassten Texte vornehmen. Das im Projekt untersuchte Klassifikationsproblem zeichnet sich durch die große Anzahl möglicher Text-Klassen aus (mehr als 90). Für jede Klasse konnte Echobot eine unterschiedliche Menge an Beispieldokumenten bereitstellen.

**Die Herausforderung** Ein Unternehmen, das sich wie Echobot auf Big Data- und Smart Data-Projekte spezialisiert hat, steht immer wieder vor der Herausforderung, geeignete Methoden und Algorithmen auszuwählen und bei den Analysen einzusetzen. Die Auswahl und optimale Konfiguration dieser Werkzeuge ist dabei entscheidend und beeinflusst die Qualität der späteren Ergebnisse nachhaltig. Ziel des Projekts mit dem SDSC-BW war es, sowohl hardware- als auch softwareseitige Optimierungspotentiale zu ermitteln und Empfehlungen für die zur Anwendung kommenden Analyseverfahren abzuleiten. Die besondere Herausforderung im durchgeführten Projekt lag in der großen Anzahl unterschiedlicher Text-Klassen bzw. Kategorien und der daraus resultierenden großen Anzahl an Trainingsdaten („Beispiel-Texte“). Die hohen Anforderungen machten u.a. Spezial-Hardware erforderlich, die Echobot selbst nicht zur Verfügung stand.

**Die Potentialanalyse** In der kostenlos durchgeführten Potentialanalyse untersuchten die Experten des SDSC-BW die bereitgestellten Daten. Die Erstanalyse konzentrierte sich darauf, die hohe Anzahl an Attributen auf die relevantesten zu reduzieren und eine Modellbildung zur Klassifikation der Inhalte vorzunehmen. Die Analyse wurde durch das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert.

**Das Resultat** Auf Basis von Smart Data-Methoden untersuchte das Expertenteam verschiedene Ansätze zur Klassifikation der Texte und führte für das beste Verfahren eine Parametersuche durch. In einer weiteren Analyse mit einer durch Echobot erweiterten Datenbasis konnte es nachweisen, dass sich die Rate für eine korrekte Klassifikation im Vergleich mit den ursprünglichen Daten erhöhen ließ.

**Wie es weiter geht** Die Analyse des SDSC-BW hat Echobot gezeigt, welche unterschiedlichen Algorithmen und Tools für den Praxiseinsatz erfolgversprechend sind und welche nicht. Mit den Erkenntnissen aus dem Projekt werden sich die Profis nun auf die richtigen Werkzeuge konzentrieren, um noch bessere Ergebnisse aus den eigenen Lösungen herauszuholen.



Kostenlose Potentialanalyse  
gefördert durch  
das Land Baden-Württemberg

Kontakt

0711. 217 2828-0  
info@sdsc-bw.de

oder unter

[www.sdsc-bw.de/potentialanalyse](http://www.sdsc-bw.de/potentialanalyse)