

# Inventory Management: Die Vermessung des deutschen E-Commerce

Das unabhängige Shopping- und Vergleichsportal idealo hilft seinen Kunden die optimale Entscheidung beim Kauf von Konsumgütern zu treffen. Dazu gehört es unter anderem, einen Überblick über die Angebote auf dem E-Commerce-Markt zu schaffen, sie vergleichbar zu machen und auch direkt bei idealo kaufen zu können. Damit der Kunde das für ihn optimale Angebot findet, muss das Angebotsinventar bei idealo **möglichst vollständig** und dadurch neutral sein.

Als Deutschlands größtes Preisvergleichsportal sind bei idealo subjektiv alle relevanten Angebote gelistet. Eine aktuelle, objektive Messung gibt es jedoch nicht. Der Anteil der gelisteten Angebote, die auch bei idealo direkt gekauft werden können, muss hingegen noch deutlich gesteigert werden. Das Ziel von idealo ist es dem Kunden einen vollständigen Überblick über den deutschen E-Commerce-Markt zu verschaffen (Neutralität, Markttransparenz) und jedes Angebot auch direkt bei idealo kaufen zu können (Vertrauen, Einfachheit).

## Projektbeschreibung

Das Ziel des Bachelorprojektes ist es die Angebote deutscher Onlineshops für Konsumgüter zu **inventarisieren** und abzugleichen mit dem von ihnen bei idealo gelisteten Inventar. Dabei soll unterschieden werden zwischen Angeboten, die nur im Preisvergleich gelistet sind, und solchen, die auch für den Direktkauf freigegeben sind. Die Lösung soll für beide Kategorien entdecken, welches Inventar idealo fehlt, bzw. gegebenenfalls welche Shops komplett fehlen. Da das Angebot auf dem deutschen Online-Konsumgüter-Markt sehr starken Schwankungen unterliegt, muss die Analyse jederzeit wiederholbar sein und in beschränkter Zeit durchgeführt werden können. Dazu ist eine stark **parallelisierte Verarbeitung** notwendig. Die Daten über das bei idealo vorhandene Inventar (zurzeit ca. 2 Mio. Produkte in Deutschland), macht idealo natürlich zugänglich.



Im Projekt sollen die nachfolgend aufgeführten Anforderungen umgesetzt werden. Bis zu welcher Tiefe dies jeweils erfolgt, stimmen wir gemeinsam im Projektverlauf ab.

## Projektziele

- **Crawling von Shopseiten/Extraktion von Angebotsdaten:** Hochgradig paralleles Crawlen von Shopseiten, Extraktion derer Angebotskataloge, Verifikation, dass das Inventar Konsumgüter umfasst, Vermeidung von Beeinträchtigungen des Shop-Betriebs durch das Crawling.
- **Angebots-Matching/Produkt-Zuordnung:** Matching der inventarisierten Angebote des Shops mit dem Preisvergleichs- und Direktkauf-Inventar von idealo auf Produktebene (In welche idealo-Kategorie fällt das Angebot? Hat idealo das Produkt gelistet?) und Angebotsebene (Hat idealo dieses Angebot dieses Händlers gelistet?)
- **Katalogabgleich:** Persistierung des analysierten Inventars, Zusammenfassung der gematchten Angebote: Welcher Teil des Inventars fehlt idealo im Preisvergleich und im Direktkauf? Unterscheidung nach Händlern ist notwendig, nach Kategorie wäre es ein zusätzlicher Mehrwert. Hierbei ist besonders die Datenmenge zu beachten: idealo listet derzeit in Deutschland rund 330 Mio. Angebote.
- **Reporting:** Auswertungen, die performant und benutzerfreundlich folgende Fragen beantworten: Welches Inventar fehlt idealo je Händler im Preisvergleich und im Direktkauf? Welche Shops fehlen idealo komplett? Gibt es Abdeckungs-Unterschiede zwischen einzelnen Kategorien? Weitergehende Analysemöglichkeiten sind wünschenswert.

Weiterhin sollen die üblichen Anforderungen in der Softwareentwicklung beachtet werden: Wiederverwendung (ggf. existierender idealo-Module), Modularität (Erweiterbarkeit), Einfachheit, Effizienz, Robustheit (Fehlertoleranz), Transparenz (Monitoring). Die technische Lösung muss **hochgradig parallel** konzipiert sein und sollte moderne Matchingverfahren, z.B. mittels **Machine Learning** umsetzen. Die Implementierungssprache sollte auf der JVM basieren, z.B. Scala auf Kafka und Spark. Um eine gemeinsame und flexible Konkretisierung der zu erreichenden Ziele zu ermöglichen, ist ein **agiles Vorgehen** erwünscht und wird von idealo unterstützt.

## Projektpartner

idealo ist Deutschlands großes unabhängiges Shopping- und Vergleichsportal. Seit seiner Gründung im Jahr 2000 hat sich das Unternehmen vom Drei-Mann-Startup zum Marktführer unter den Produktpreisvergleichen entwickelt. In den Büros in Berlin-Kreuzberg arbeiten über 700 Menschen täglich daran, Verbraucher umfassend, aktuell und genau über Online-Angebote und -Händler zu beraten.

idealo arbeitet eng mit tausenden Onlineshops zusammen. Egal ob Global Player wie Amazon oder spezialisierter Anbieter mit kleinem Produktsortiment: Bei idealo hat jeder Händler die Chance, mit dem günstigsten Angebot ganz oben zu stehen.

idealo ist der erste und einzige Produktpreisvergleich mit dem Siegel des TÜV Saarland für geprüfte Vergleichsportale. Das Unternehmen legt Wert auf Nachhaltigkeit und kooperiert u. a. mit Verbraucherschutz-Vertretern wie der Stiftung Warentest, HTV-Life und Der Blaue Engel.

Das Projekt beginnt am 9. Oktober 2017 und wird durch Prof. Dr. Felix Naumann und sein Team betreut. Fragen können gerne an [felix.naumann@hpi.de](mailto:felix.naumann@hpi.de) gerichtet werden.