

DAIMLER

Social Networks as Data Source for Recommendation Systems

Mathias Bank, Jürgen Franke

mathias.bank@daimler.com, juergen.franke@daimler.com

Review Based Recommendation Systems



Hotelbewertungen Finca/Hotel Ca'n Canals

von allen
 von Paaren
 von Familien

Hotel	5.8	Lage & Umgebung	4.9
Zimmer	5.8	Gastronomie	5.6
Service	5.9	Sport & Unterhaltung	5.5
Gesamt	5.6	Weiterempfehlung 100%	Trend

Gäste loben: gute Fremdsprachenkenntnisse, freundliches Personal, Größe der Zimmer, guter Check-In/Check-Out, allgemeine Sauberkeit, [alle anzeigen](#)

[Reiseangebote für dieses Hotel finden & prüfen](#)

61 Hotelbewertungen sortieren nach

Super-Urlaub in der idyllischen Finca "Can Canals" 5.8

Die Finca ist ein kleines komfortables Hotel mit hohem Verwöhncharakter wo wir uns sofort wohlfühlten. Die Zimmer waren mit sehr viel Liebe eingerichtet und sehr großzügig und wurden jeden Tag, so wie die Terrasse, gereinigt (tägliches Handtuchw... [Hotelbewertung weiterlesen](#)

Mai 10, Dieter & Petra, Alter 51-55, Familie 29 mal gelesen - 100% hilfreich

Maravilloso 5.8

Diese schöne kleine aber großzügig angelegte Finca-Anlage empfehlen wir allen Ruhe Suchenden. Es gibt hier 12 großzügige Junior Suiten sowie weitere Zimmer in einem Nebengebäude. Diese Anlage toppt vieles, was wir bisher gesehen haben, weil sie so per... [Hotelbewertung weiterlesen](#)

Mai 10, Lydia, Alter 36-40, Paar 134 mal gelesen - 100% hilfreich

bimodal

Advertisement, fake reviews, communication, copies

Kundenrezensionen

52 Rezensionen

5 Sterne	(41)
4 Sterne	(6)
3 Sterne	(0)
2 Sterne	(2)
1 Sterne	(3)

Durchschnittliche Kundenbewertung
★★★★☆ (52 Kundenrezensionen)

Rezensent: **have liked more!**, 21. Juli 2000

Rezension hilfreich: (KS United States) - [Alle meine Rezensionen ansehen](#)

Rezensent: **Non-Designer's Design Book, Design and typographic principles for the visual novice (Taschenbuch)**

as an aspiring web page designer (who always believed she had a "good eye" but often needs hours of trial and error

Helfen Sie anderen Kunden bei der Suche nach den hilfreichsten Rezensionen

War diese Rezension für Sie hilfreich?

Ja Nein

[Rezensent unzumutbar?](#) | [Kommentar als Link](#) | [Kommentar](#)

Rating method to...

- Deal with large amount
- prioritize

Review Rating Techniques

Sorting Strategies for Product Reviews

Peer based rating method

Not quality related

Algorithm rating based on...

Objectivity

Document sentiment

User rating

What about...

... quality changes over time?

... products with many variants?

... products with no or only few reviews?

Social Networks as recommendation data source



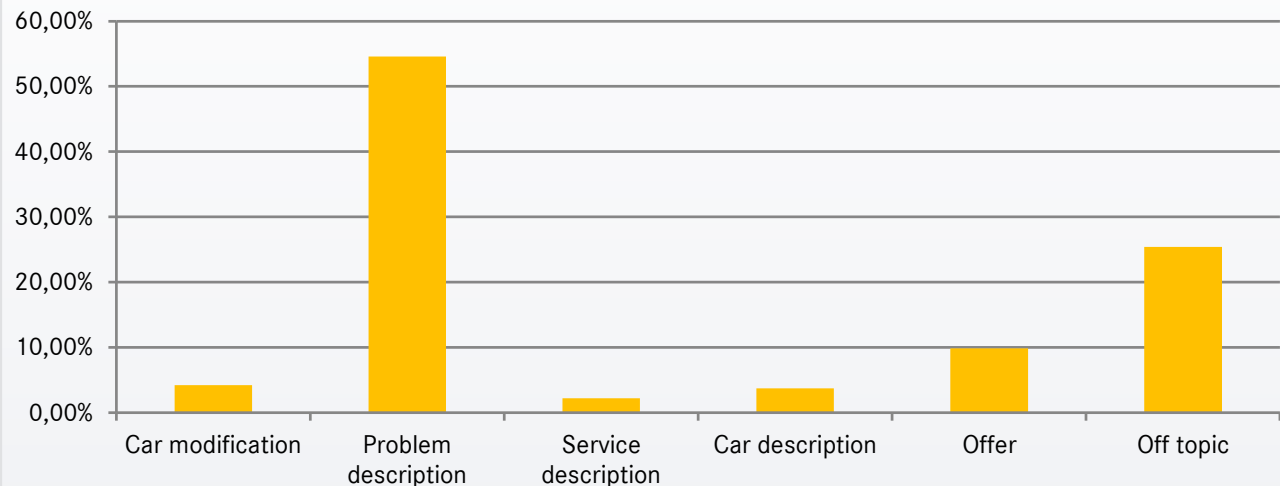
Requirements

- Easy to crawl
- Textual content
- Analysis Topic related
- Quality related



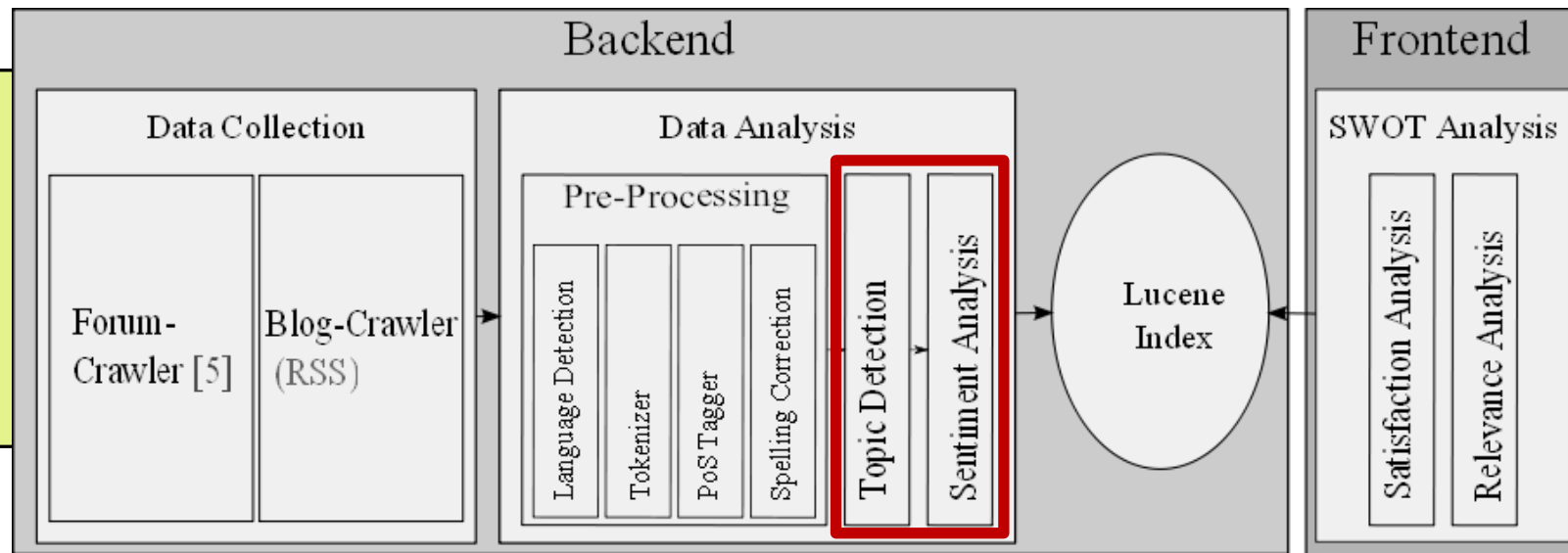
purposeful selection

Discussion Genres in Internet fora (benzworld.org)



Manually classified using 1.196 chronologically selected discussions

Recommendation System Architecture



Topic Detection via Taxonomy Structure

The image shows a taxonomy browser interface. On the left, a tree structure is displayed under the root 'automotive'. The 'component' folder is expanded, showing sub-folders like 'body', 'Chassis and suspension', 'communication and information', 'Other components', 'Overall vehicle', 'Power generation', 'power transmission', 'Smaller parts', and 'substances'. Under 'communication and information', the 'Warning and control system' folder is selected. On the right, an 'Info: Warning and control system' window is open, displaying multilingual translations and a count of 385 items. The translations include: zh: 警告和控制系统, da: Advarsels- og reguleringssystem, ro: sistem de avertizare și reglare, it: Sistema di segnalazione e di regolazione, tr: İkaz ve kumanda sistemi, ko: 경고 및 제어 시스템, cs: Varovné a regulační systémy, de: Fahrassistenzsysteme, Fahrerassistenzsysteme, Warn- und Regelsystem, Fahrassistenzsystem, FAS, Fahrerassistenzsystem, el: Warning and control system, fi: Varoitus- ja säätöjärjestelmä, pt: Sistema de alarme e controle, pl: system ostrzegawczy i regulacyjny, sv: Varnings- och regler-system, fr: Système d'alerte et de régulation, en: Advanced Driver Assistance System, Driver Assistance System, Advanced Driver Assistance Systems, ADAS, Warning and control system, ru: Предупредительная и управляющая система, es: Sistema de advertencia y de regulación, ja: ウォーニング/コントロール・システム, nl: waarschuwings- en regelsysteem. The query is 'inkTaxID#conceptID:automotive#769' and the count is 385.

automotive

- component
 - body
 - Chassis and suspension
 - communication and information
 - antenna
 - Communication
 - Entertainment
 - Information display
 - Warning and control system
 - data link connector
 - transmitter
 - Other components
 - Overall vehicle
 - Power generation
 - power transmission
 - Smaller parts
 - substances
 - axle pedal
 - car

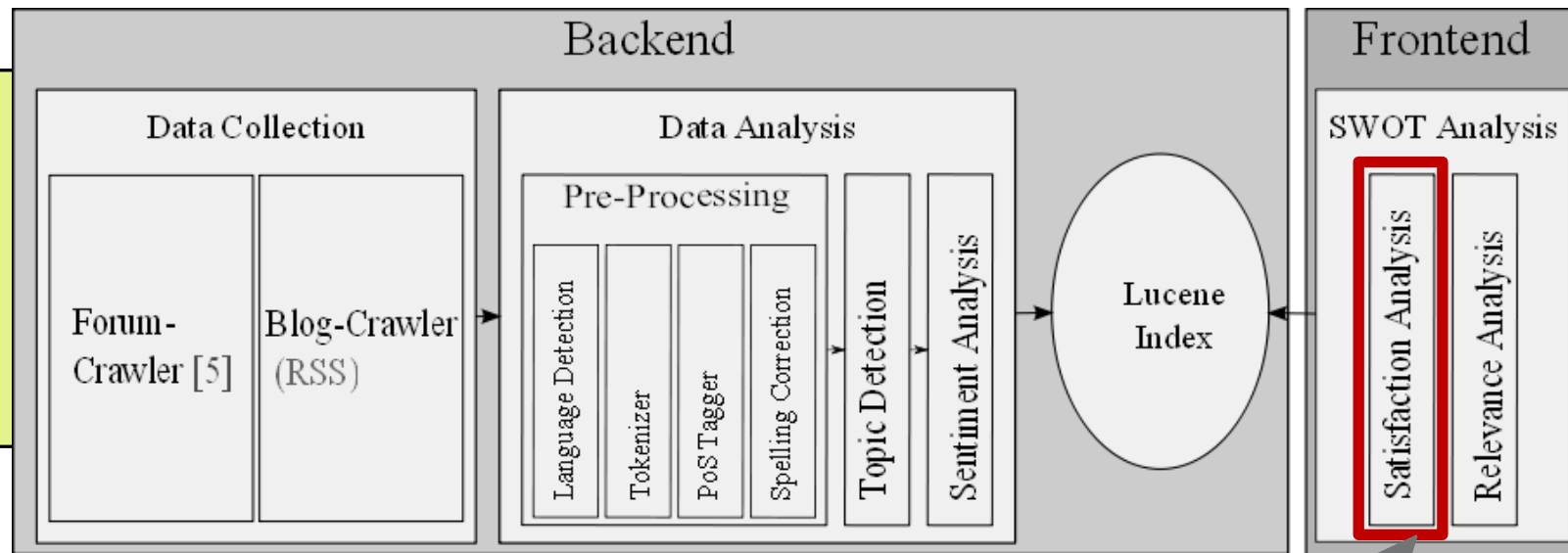
Info: Warning and control system

zh: 警告和控制系统
 da: Advarsels- og reguleringssystem
 ro: sistem de avertizare și reglare
 it: Sistema di segnalazione e di regolazione
 tr: İkaz ve kumanda sistemi
 ko: 경고 및 제어 시스템
 cs: Varovné a regulační systémy
 de: Fahrassistenzsysteme, Fahrerassistenzsysteme, Warn- und Regelsystem, Fahrassistenzsystem, FAS, Fahrerassistenzsystem
 el: Warning and control system
 fi: Varoitus- ja säätöjärjestelmä
 pt: Sistema de alarme e controle
 pl: system ostrzegawczy i regulacyjny
 sv: Varnings- och regler-system
 fr: Système d'alerte et de régulation
 en: Advanced Driver Assistance System, Driver Assistance System, Advanced Driver Assistance Systems, ADAS, Warning and control system
 ru: Предупредительная и управляющая система
 es: Sistema de advertencia y de regulación
 ja: ウォーニング/コントロール・システム
 nl: waarschuwings- en regelsysteem

Query: inkTaxID#conceptID:automotive#769
 Count: 385

Ok

Recommendation System Architecture



Are customers satisfied with a product feature?

Satisfaction Analysis

Base Concept:

Is a product feature discussed positively or negatively

$$p(xy^+|xy) = \frac{\#xy^+}{\#xy}$$

$$p(xy^-|xy) = \frac{\#xy^-}{\#xy}$$

$$s_1(xy) = \frac{p(xy^+|xy)}{p(xy^-|xy)} = \frac{\#xy^+}{\#xy^-}$$

Problem:

Natural language is not neutral

Sentiment lexicon is not balanced

Satisfaction Analysis II

Solution:

Measure relative to the market (the market defines „neutral“)

$$p(xy^+|xy) = \frac{\#xy^+}{\#xy}$$

$$p(xy^-|xy) = \frac{\#xy^-}{\#xy}$$

$$p(y^+|y) = \frac{\#y^+}{\#y}$$

$$p(y^-|y) = \frac{\#y^-}{\#y}$$

$$s_2(xy) = \frac{\frac{p(xy^+|xy)}{p(y^+|y)}}{\frac{p(xy^-|xy)}{p(y^-|y)}} = \frac{\frac{\#xy^+}{\#y^+}}{\frac{\#xy^-}{\#y^-}}$$

Problem:

Not symmetrical

$\in [0; \infty]$

Satisfaction Analysis III

Solution:

Rescale

$$f_{positive} = \frac{\#xy^+}{\#y^+}$$

$$f_{negative} = \frac{\#xy^-}{\#y^-}$$

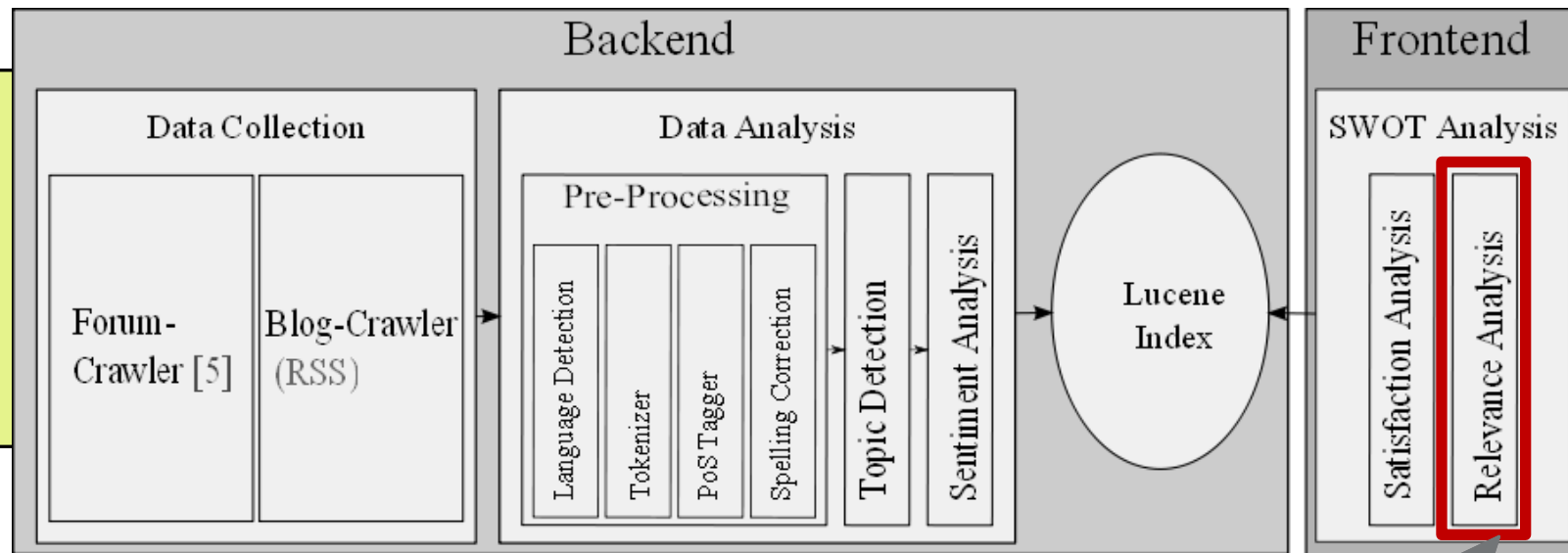
$$s_3(xy) = \begin{cases} \frac{f_{positive}}{f_{negative}} - 1 & f_{positive} \leq f_{negative} \\ 1 - \frac{f_{negative}}{f_{positive}} & \text{else} \end{cases}$$

Advantages:

symmetrical: $\in [-1; 1]$

Normally distributed (ensured with Shapiro-Wilk Normality Test)

Recommendation System Architecture



Is this feature relevant to other customers?

Relevance Analysis

Already applied relevance measures:

Absolute number of comments



Fluctuation not taken into account (e.g. holiday)

Normalization to the market



Not marketing resistant

Market penetration not taken into account



Valid measure for marketing analysis

Quality related relevance measure:

Normalization to the product

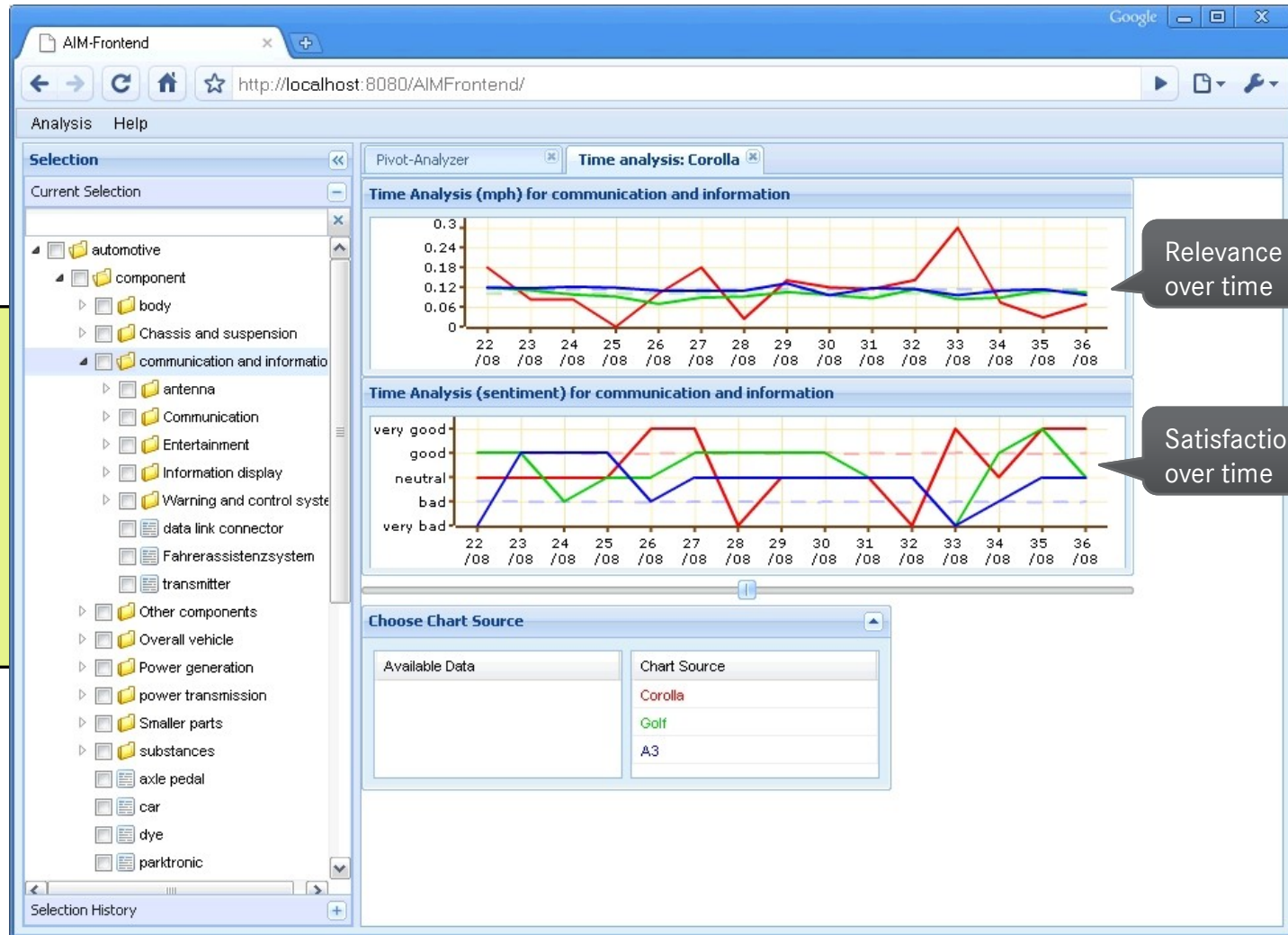
$$p(y|x) = \frac{\#xy}{\#x}$$

AIM Prototype

The screenshot shows the AIM-Frontend web application interface. The browser address bar indicates the URL is `http://localhost:8080/AIMFrontend/`. The application has a menu bar with 'Analysis' and 'Help'. On the left, there is a 'Selection' sidebar with a tree view of components. The main area is a 'Pivot-Analyzer' table. A callout labeled 'Product' points to the top row of the table. A callout labeled 'Feature' points to the 'component' row. A callout labeled 'Relevance' points to the 'communicat' row. A callout labeled 'Satisfaction-Classification' points to the '3894' value in the 'A3' column of the 'communicat' row.

	Corolla	Golf	A3
component	72.64 mph (1418)	76.17 mph (19359)	70.28 mph (22467)
communicat	10.66 mph (208)	10.87 mph (2762)	3894
Chassis and suspension	15.98 mph (312)	19.73 mph (5015)	18.72 mph (5986)
Power generation	22.69 mph (443)	24.33 mph (6183)	18.34 mph (5864)

AIM Prototype – Quality Analysis over Time







AIM Prototype – Detail Analysis

The screenshot displays the AIM-frontend web application interface, which is used for analyzing forum posts. The interface is divided into several sections:

- Selection Panel (Left):** A tree view showing a hierarchy of categories. The 'cars' category is selected, and the 'Toyota' sub-category is highlighted.
- Search Query (Top):** A search query is entered: `inkTaxID#conceptID:automotive#34 AND inkTaxID#conceptID:automotive#34 AND inkTaxID#conceptID:automotive#34`.
- Search Results (Middle):** A list of forum posts is displayed. The first post is selected, and its content is shown in the main view. The post title is "90AIItrac - Leaking Trans Fluid".
- Analysis Panel (Right):** A detailed analysis of the selected post. It shows the post's content with highlighted terms and a list of related terms. The analysis includes:
 - Category:** Motorsport/Toyota Passenger and Sports Car Forums/Corolla Forum/6th Generation 1988-1992
 - Subject:** 90AIItrac - Leaking Trans Fluid!
 - Post Content:** A series of forum posts with highlighted terms. The first post is from user 'eage8' on 30.05.2008, discussing a leak and mentioning terms like 'fluid', 'cooler', 'radiator', and 'basket'. The second post is from user '90AIItrac93kor' on 31.05.2008, discussing the location of the leak and mentioning terms like 'leak', 'pan', 'basket', 'check', 'oil', 'level', 'never', 'moved', 'diff', 'frontright', 'brake', 'shaft', 'problem', 'know', 'slow', 'less', 'help'. The third post is from user '90AIItrac93kor' on 31.05.2008, discussing the location of the leak and mentioning terms like 'small', 'radiator', 'sitting', 'between', 'the', 'radiator', and 'grill', 'oil', 'cooler', 'nose', 'runs', 'from', 'the', 'side', 'of', 'the', 'oil', 'pan', 'and', 'the', 'other', 'connects', 'right', 'between', 'the', 'block', 'and', 'oil', 'filter', 'update'. The fourth post is from user 'rrc88' on 09.06.2008, discussing the car and mentioning terms like 'corolla', 'alltrack', '4WD', 'corolla', 'project', '4age20v', 'black', 'top', 'know', 'transmission', 'will', 'work', 'with', 'the', '20v', 'where', 'can', 'check', 'that', 'don't', 'know', 'the', 'code', 'of', 'the', 'transmission', 'engine', 'do', 'you', 'have', 'thanks'.

Satisfaction as Quality Measure

Does the Satisfaction measure represent quality relevant information?

Model	Used Car Value Range	Predicted Reliability Rating	Performance and Design Rating	Initial Quality Rating
 2008 Mercedes C-Class Sedan	Price It ▶	●●●●○	●●●●●	●●●●○
 2008 Mercedes CL-Class Coupe	Price It ▶	●●●●○	Not Rated	Not Rated
 2008 Mercedes CLK-Class Convertible <i>Award Recipient</i>	Price It ▶	●●●●○	●●●●●	●●●●●
 2008 Mercedes CLK-Class Coupe <i>Award Recipient</i>	Price It ▶	●●●●○	●●●●●	●●●●●



	Corolla	Golf	A3
component	72.64 mph (1418) ▲	76.17 mph (19359) ▲	70.28 mph (22467) ▲
communication and information	10.66 mph (208) ▲	10.87 mph (2762) ▼	12.18 mph (3894) ▼
Chassis and suspension	15.98 mph (312) ▲	19.73 mph (5015) ▼	18.72 mph (5986) ▼
Power generation	22.69 mph ▲	24.33 mph ▼	18.34 mph ▼

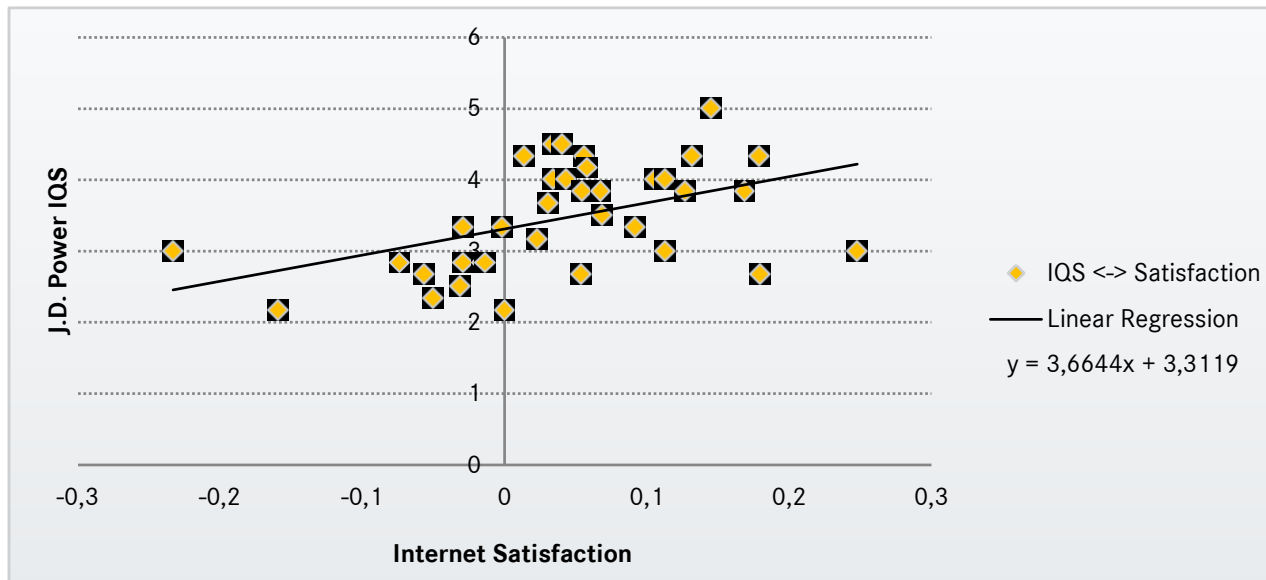
J.D. Power IQS-Rating
∈ [1; 5]

Questionnaire

Satisfaction „component“
∈ [-1; 1]

Open Sensor

Satisfaction as Quality Measure II



Correlation coefficient: 0.46

➔ Question: Is correlation significant?

Satisfaction as Quality Measure III

Significance test for the correlation

Transformation to identical space

Linear regression (no correlation influence):

$$s_4 = 3.66 * s_3(xy) + 3.31$$

→ RMSE: 0.65

T-test analysis

$$p(\mu_0 = 0) = 0.988 \text{ (two-tailed p-value)}$$

→ There is no systematic difference between J.D. Power's IQS and the Internet Satisfaction measure.

Conclusion

Product Reviews

- Increase sales volume
- No reliable quality content
- Text comments numerous
- Ratings bimodal
- Not available in all ecommerce systems or for every product

- Quality analysis on different abstraction levels
- Time related quality analysis
- Feature related
- Large number of comments
- Search & Filtering functionality
- Difficult to control

Product Discussions In Social Networks

Thank you for your attention