



**Hasso
Plattner
Institut**

IT Systems Engineering | Universität Potsdam

Context Matching Philipp Maschke

A Novel Context-based Technique for
Web Information Retrieval



Agenda

2

- **Motivational Example**
- **Traditional Scoring**
- **Term Context?**
- **Scoring with Context Matching**
- **Context Matching – Example**
- **Traditional vs. Context Matching**

Motivational Example

3

- **Searching for information about the BMW Z3**
- **Query: ‚bmw z3‘**
- **Google results:**

Motivational Example

4



[Anmelden](#)

Artikel
Diskussion
Seite bearbeiten
Versionen/Autoren

Unterstützen Sie Wikipedia: ein gemeinnütziges Projekt. Spenden Sie jetzt >> [\[Anzeigen\]](#)

BMW Z3

Der **BMW Z3** ist ein [PKW](#) des deutschen Automobilherstellers [BMW](#).

Der zweisitzige [Roadster](#) wurde 1996 auf dem Markt eingeführt und 1998 durch die [Coupé](#)-Variante ergänzt. Im April 1999 brachte BMW ein verändertes Modell der Roadster auf den Markt ([Facelift](#)). Die Motorenpalette reicht von einem 1,8-Liter-[Vierzylindermotor](#) mit 85 kW (115 PS) bis zu einem 3,2-Liter-Sechszylinder-Motor mit 239 kW (325 PS). Bis 18. Juni 2002 liefen in [Spartanburg, South Carolina \(USA\)](#) 279.273 Roadster und 17.815 Coupés vom Band. Der Z3 trägt die interne Bezeichnung [E36/7](#) für den Roadster und [E36/8](#) für das Coupé (E36/7S bzw. E36/8S für die Motorsport-Modelle). Er basiert auf der Plattform des [E36 Compacts](#) und teilt sich mit ihm auch diverse Teile des Cockpits.

Inhaltsverzeichnis [\[Verbergen\]](#)

- 1 Modellgeschichte
 - 1.1 Facelift
- 2 Karosserieversionen
- 3 Verdeck
- 4 Sondermodelle
- 5 Motorisierung
- 6 Weblinks

Modellgeschichte [\[Bearbeiten\]](#)

Am 20. September 1995 lief der erste BMW Z3 vom Band. Die Produktion des [Roadsters](#) begann mit dem Vierzylindermodell *Z3 Roadster 1.8* und *1.9*. Diesen [platzierte](#) BMW in dem [James-Bond](#)-Film „[GoldenEye](#)“. 1996 wurde die Produktpalette durch den *Z3 Roadster 2.8* mit einem Reihensechszylinder erweitert, welcher u. a. an breiteren Kotflügeln hinten und zwei Bearengulasleuchten in der Frontstoßstange zu erkennen ist.

BMW



Z3 Roadster, geschlossen

Z3	
Hersteller:	BMW
Produktionszeitraum:	1996–2002
Klasse:	Sportwagen
Karosserieversionen:	Roadster , Shooting Brake
Motoren:	Ottomotoren : 1,8–3,2 Liter (85–239 kW)
Länge:	4025–4050 mm
Breite:	1692–1740 mm
Höhe:	1266–1306 mm

□ **Good**

Motivational Example

5



mobile.de Deutschlands größter Fahrzeugmarkt

Hilfe | Übersicht  Deutsch

SUCHEN ANBIETEN INFORMIEREN MEIN MOBILE.DE HÄNDLER

Home » Modellverzeichnis » BMW » Z3

BMW Z3

BMW Z3 als Gebrauchtwagen oder neues Auto

Der BMW Z3 war der erste in den USA produzierte BMW. Ab 1996 lief der Z3 Roadster im eigens dafür gebauten Werk in Spartanburg, South Carolina vom Band. Der BMW Z3 war zunächst nur mit Vierzylindermotoren erhältlich, deren Leistung 115 PS und 140 PS betrug. 1997 wurde die Motorenpalette des BMW Z3 um einen 192 PS starken 2,8-Liter-Reihensechszylinder erweitert. Der BMW Z3 2.8 war äußerlich durch sein breiteres Heck zu erkennen. Zur gleichen Zeit kam auch der 321 PS starke BMW M Roadster in den Handel. 1997 debütierte dann das zweiseitige BMW Z3 Coupé, das im englischen Sprachraum auch unter dem Namen Z3 Shooting Brake bekannt wurde. 1999 bekam der BMW Z3 Roadster ein Facelift, erkennbar an neu geformten Rückleuchten und stärker ausgestellten Kotflügeln, mit Hüftschwung, im Heck. Der Vierzylindermotor mit 140 PS wurde durch einen 2,0-Liter-Sechszylinder mit 150 PS ersetzt. Im Jahr 2000 hat BMW das Motorenkonzept des Z3 nochmals überarbeitet: Der Z3 2.0 mit 150 PS wurde durch den BMW Z3 2.2i mit 170 PS ersetzt, der Z3 2.8 mit 193 PS durch den BMW Z3 3.0i mit 3,0-Liter-Motor und 231 PS. Der M Roadster wurde über die komplette Bauzeit mit altem Heck weitergebaut. 2002 wurde der BMW Z3 schließlich vom BMW Z4 abgelöst. Sie möchten einen BMW Z3 Roadster kaufen oder verkaufen? Nutzen Sie mobile.de, den großen Online-Marktplatz für Fahrzeuge aller Art.

Weitere Autos dieser Klasse

- > Mazda Mx 5 Unplugged
- > Ferrari F430
- > Nissan 350Z
- > Ferrari 330
- > Audi TT
- > Mercedes Benz Slk 280
- > Maserati Spyder
- > Mercedes Benz Slk 32 Amg
- > Porsche 944 S2
- > Mazda Mx 5 Trilogy
- > Noch mehr: Sportwagen

Bmw Händler

- > Autohaus Cloppenburg GmbH
- > BMW Niederlassung Essen
- > BMW AHG mbH St.Georgen
- > BMW Vertragshändler AH Heuchert GmbH & Co. KG
- > Automobile Bellinghausen
- > Armin Kühl GmbH

□ Also good

Motivational Example

6

Fahrbericht BMW Z3

Die Herztransplantation

Von FOCUS-Online-Redakteur Gerd Stegmaier

James Bond hat den Z3 als Dienstwagen abgelegt und fährt inzwischen Z8. Jetzt hat BMW dem Geheimagenten Ihrer Majestät mittels einer Herztransplantation noch einmal einen Anlass gegeben, sein altes Gefährt wieder in die engere Wahl zu nehmen.



Die Herztransplantation: BMW Z3



Fahrbericht BMW Z3

Seite 1/7 ▶

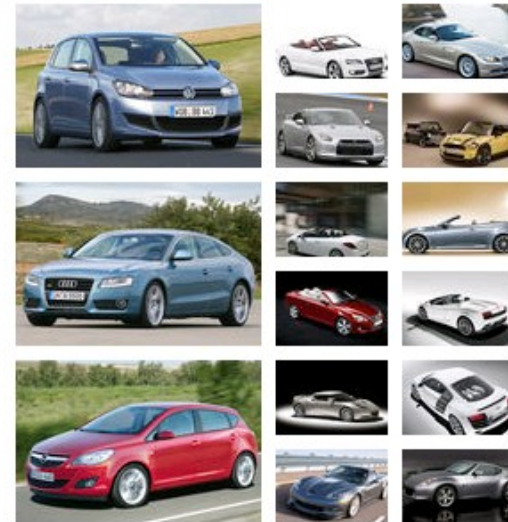
- [Die Herztransplantation](#)
- [Karosserie/Innenraum](#)
- [Fahrwerk/Fahrverhalten](#)
- [Motor/Getriebe](#)
- [Kosten](#)
- [Datenblatt](#)

□ **Yep**

MEHR AUTO

- ✔ Service
- Community
- News
- Fotos
- Videos

AUTONEUHEITEN 2009



NEUESTE FAHRBERICHTE



Alltags-Check Lamborghini Gallardo: Alarmstufe Weiß



Alltags-Check Mazda 2 Sport: So leicht kann es sein





PROBEFAHRT: Gewöhnlicher Rebell

Motivational Example


7

Weltbild.de






- 35%
Schlank im Schlaf
Der 4-Wochen-Power-Plan
9.95 EUR



- 25%
Wende-Bettwäsche "Fresh"
135x200 cm
14.99 EUR

Lifestyle
Computer & Technik
Geld & Finanzen
Nachrichten
mehr ▾

[Mehr MSN](#) | [Hotmail](#) | [Messenger](#) | [Anmelden](#)




[Hilfe & FAQ](#) | [Sitemap](#)

Home
Bekleidung & Schuhe
Haushalt & Wohnen
Elektronik
Handy
Computer & Software
Alle Kategorien

[Home](#) > [Spielzeug](#)

Darda DARDA BMW Z3: Testberichte u. Bewertungen



mytoy

Darda DARDA BMW Z3

von Darda

Produktdetails

Shops und Preise

Testberichte u. Bewertungen

3 2 1 Los! Für spannende Rennen auf den Darda Rennbahnen ist dieser flotte Flitzer bestens geeignet. Ausgestattet mit einem Friktionsmotor in ...

Diesen Artikel bewerten

Ziehen Sie Ihre Maus über die Sterne, um eine Bewertung zwischen 1 und 5 Sternen auszuwählen, und klicken Sie, um sie zu speichern. Um doppelte Bewertungen zu vermeiden, melden Sie sich bitte mit Ihren Microsoft Passport-Netzwerk-Zugangsdaten (E-Mail-Adresse und Kennwort) bei MSN Shopping an.


☆☆☆☆☆

□ Hm...


Motivational Example

8

Flüge clever kombinieren und sparen



Alle Flüge mit Geld-zurück-Garantie



Lifestyle | Computer & Technik | Geld & Finanzen | Nachrichten | mehr ▾
Mehr MSN | Hotmail | Messenger | **Anmelden**

msn
Hilfe & FAQ | Sitemap

Home | Bekleidung & Schuhe | Haushalt & Wohnen | Elektronik | Handy | Computer & Software | Alle Kategorien

Suchen
Web

Home > [Uhren & Schmuck](#) > [Kettenanhänger](#)


Autoschlüsselanhänger in 925 Sterling Silber BMW Z3: Testberichte u. Bewertungen

Autoschlüsselanhänger in 925 Sterling Silber BMW Z3

Produktdetails

Shops und Preise

Testberichte u. Bewertungen



Schlüsselanhänger aus Sterling-Silber 925/-. Weitere Modelle auf Anfrage per Email. Achtung !! Da die Schlüsselanhänger individuell gefertigt ...

Diesen Artikel bewerten

Ziehen Sie Ihre Maus über die Sterne, um eine Bewertung zwischen 1 und 5 Sternen auszuwählen, und klicken Sie, um sie zu speichern. Um doppelte Bewertungen zu vermeiden, melden Sie sich bitte mit Ihren Microsoft Passport-Netzwerk-Zugangsdaten (E-Mail-Adresse und Kennwort) bei MSN Shopping an.

☆☆☆☆☆



Motivational Example

9

Fragen & Antworten

■ Frage



? Ist eine BMW R80 Bj.87 (BMW 247) Beiwagenfähig?Wo finde ich das?

chicago3333

0
0

Ich habe eine BMW R80 Bj.87 von der Bundeswehr und finde im KFZ Schein keinen Eintrag ob ich einen Beiwagen dranbauen darf.
Wer gibt mir die notwendige Info.
Frage wurde vor 23 Stunden und 12 Minuten gestellt
Themen: Sonstiges / Quasselücke, BMW

[▶ Antworten](#)

An Freunde versenden | melden

[Bmw E39 Ersatzteile](#)
hochwert. BMW Auto - Ersatzteile
Ersatzteile f. BMW-Modelle bis -60%

[Google-Anzeigen](#)

[Gebrauchtwagen bei pkw.de](#)
Das Schnäppchen-Paradies für alle, die ein neues Auto brauchen.

[Google-Anzeigen](#)

■ Antwort finden

Suchtext...

Themenbereich auswählen

[▶ anzeigen](#)

[Top-Stellenangebote](#)
4.000 Headhunter suchen Sie. Finden Sie Top-Stellen ab 60.000 €.
[Experteer.de](#)

[Neu- und Gebrauchtwagen](#)
KADEA Autohaus Bessemerstraße GmbH
9-3, 9-5, SportLimousine SportCombi
[www.ka-de-a.de/neuwagen/saab.htm](#)

[Google-Anzeigen](#)

■ Stichworte
Beiwagen, Gespann

Neues Stichwort hinzufügen:

[hinzufügen](#)

■ Sie haben eine Frage?

■ Antworten



Manni0231

Experte

31
23

Scheint so zu sein,denn es gibt Angebote für Hilfsrahmen.Fotos von Gespannen habe ich auch gefunden von deinem Typ. google mal,da findest du bestimmt was.Ein Beispiel:
<http://www.bauhs-gespannteile.de/>

Antwort vor 22 Stunden und 53 Minuten

[hilfreich \(1\)](#) - melden

0 Kommentare

- **???** (At the **bottom** of the page are **references** to other threads, some of them concerning the z3)

Agenda

10

- Motivational Example
- **Traditional Scoring**
- Term Context?
- Scoring with Context Matching
- Context Matching – Example
- Traditional vs. Context Matching

Traditional Scoring

11

- **One of the reasons for irrelevant query results?**
- **Current approaches only use heuristics on term frequency**
 - „If a term occurs often in a document, then it must be important“
 - **Just with different variations**
 - **plus methods like query expansion, stopword lists, page segmentation, ...**

- **Basic algorithm**

- **For every term q in the query Q**

- **For every document D , in which q occurs**

- » TermConfidence = term frequency (TF) x inversed document frequency (IDF)

- » DocumentScore += TermConfidence

Traditional Scoring

13

- + easy to understand / implement
- + many variations & add-ons developed
- + generally good results
- links, advertisements, auto-generated texts distort results
- context of terms not considered

Agenda

14

- Motivational Example
- Traditional Scoring
- **Term Context?**
- Scoring with Context Matching
- Context Matching – Example
- Traditional vs. Context Matching

Term Context?

15

- **Same word, different meanings**
 - **Golf - car, sports (in German also ‚stream‘ – gulf stream)**
- **Different words, same meaning**
 - **Jet, Airplane**
- **Actual meaning defined by surrounding words**
 - **„VW Golf“, „golf club“**
 - **„The jet landed at the airport and was refueled with kerosene“ (same with ‚airplane‘)**

Term Context ?

16

- **Set of keywords** defines possible meanings of a word
 - **VW, car, drive, volkswagen, engine, gti** → car
 - **Club, hole, car, ball, course, tournament** → sport

- **Open questions**
 - **How many** keywords?
 - **Distance** between term and context keyword?
 - **How** to get keywords?
 - **Mixed sets? (same keyword(s), multiple sets)**
 - Try it and search for ‚golf car‘ (Even works with ‚auto‘)

Agenda

17

- Motivational Example
- Traditional Scoring
- Term Context?
- **Scoring with Context Matching**
- Context Matching – Example
- Traditional vs. Context Matching

Scoring with Context Matching

18

■ Basic algorithm

- Construct **query context QC**

- For every term **q** in the query **Q**

- For every document **D**, in which **q** appears

- » Compute context matching confidence CMC of q using QC

- » TermConfidence **TC** = TF + CMC

- » DocumentScore += (TC x IDF)

Construct query context QC

19

- Reused **query expansion** technique
 - **QR** - set of terms related to a query
 - **R_t** – relatedness of **t** concerning **Q**
 - Range [0,1]
 - Usually computed, here 1 by default
 - **Initial search over documents**
 - Out of **n** most relevant documents ...
 - QR = **m** highest ranked terms (usually TF x IDF)
 - **Traditional scoring used for initial search**

- **QC = QR + Q**

Construct query context QC

20

- **Q = ‚Parkinsons Disease‘**
 - **QR = {dopamine, neurons, brain, levodopa, alzheimer_s, dementia, disorder, patients, dyskinesia, substantia}**
 - **Taken from paper**

- **Q = ‚bmw z3‘**
 - **QR = {roadster, coupe, cabrio, gebrauchtwagen, gebraucht, fahrbericht}**
 - **Rough estimate using first 10 results from Google**

Scoring with Context Matching

21

■ Basic algorithm

- Construct **query context QC**

- For every term **q** in the query **Q**

 - For every document **D**, in which **q** appears

 - » Compute context matching confidence CMC of q using QC

 - » TermConfidence = TF + CMC

 - » DocumentScore += (TC * IDF)

Compute context matching confidence

22

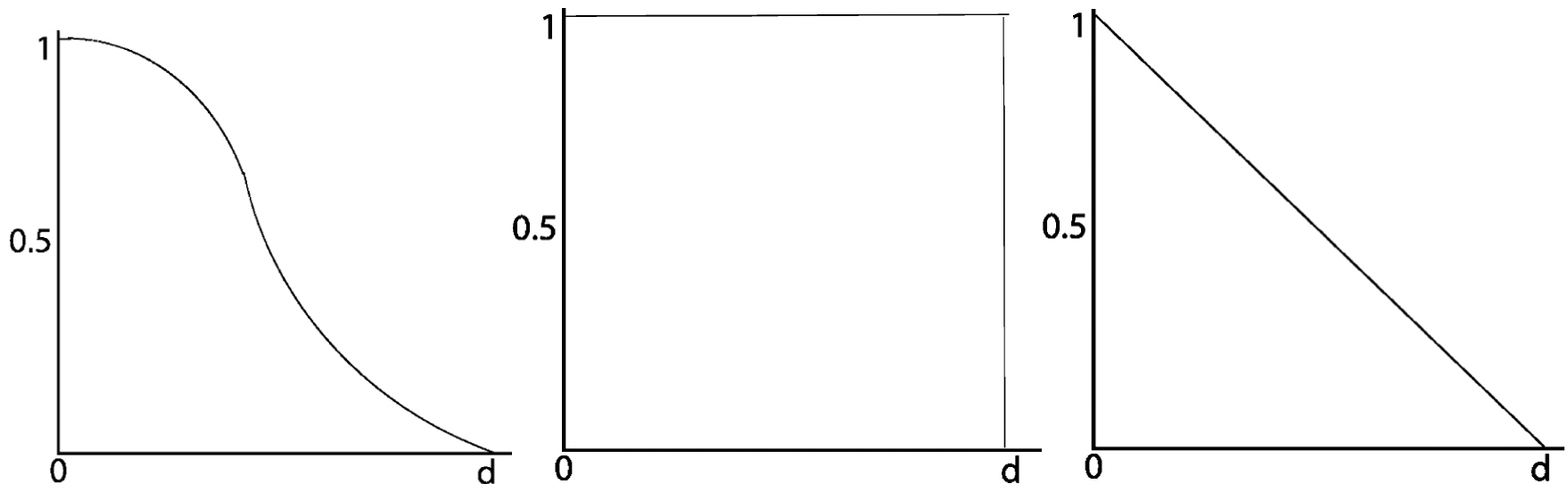
- **ContextDistance** ($CD_{a,b}$) = min. distance between all occurrences of two words in a text

- $$CMC_q = \frac{\sum_{c \in QC, c \neq q} Dist(CD_{q,c}) \times R_c}{\sum_{c \in QC, c \neq q} R_c}$$

Compute context matching confidence

23

- $Dist(CD)$ – function of **distance importance**
 - $Dist: \mathbb{N} \rightarrow [0,1]$
 - Gaussian, hard limiter or linear functions possible



Scoring with Context Matching

24

■ Basic algorithm

- Construct **query context QC**

- For every term **q** in the query **Q**

- For every document **D**, in which **q** appears

- » Compute context matching confidence CMC of q using QC

- » TermConfidence = TF + CMC

- » DocumentScore += (TC * IDF)

Agenda

25

- Motivational Example
- Traditional Scoring
- Term Context?
- Scoring with Context Matching
- **Context Matching – Example**
- Traditional vs. Context Matching

Context Matching - Example

26

- **Let's look at 2 examples**

- **Dist(CD) linear**

- **d = 10**

Context Matching - Example

$$\text{CMC}_{\text{bmw}} = \frac{(1 + 0,9 + 0,9 + 0 + 0,8 + 0 + 0)}{(1 + 1 + 1 + 1 + 1 + 1 + 1)} \approx 0,514$$

BMW Z3 als Gebrauchtwagen oder neues Auto

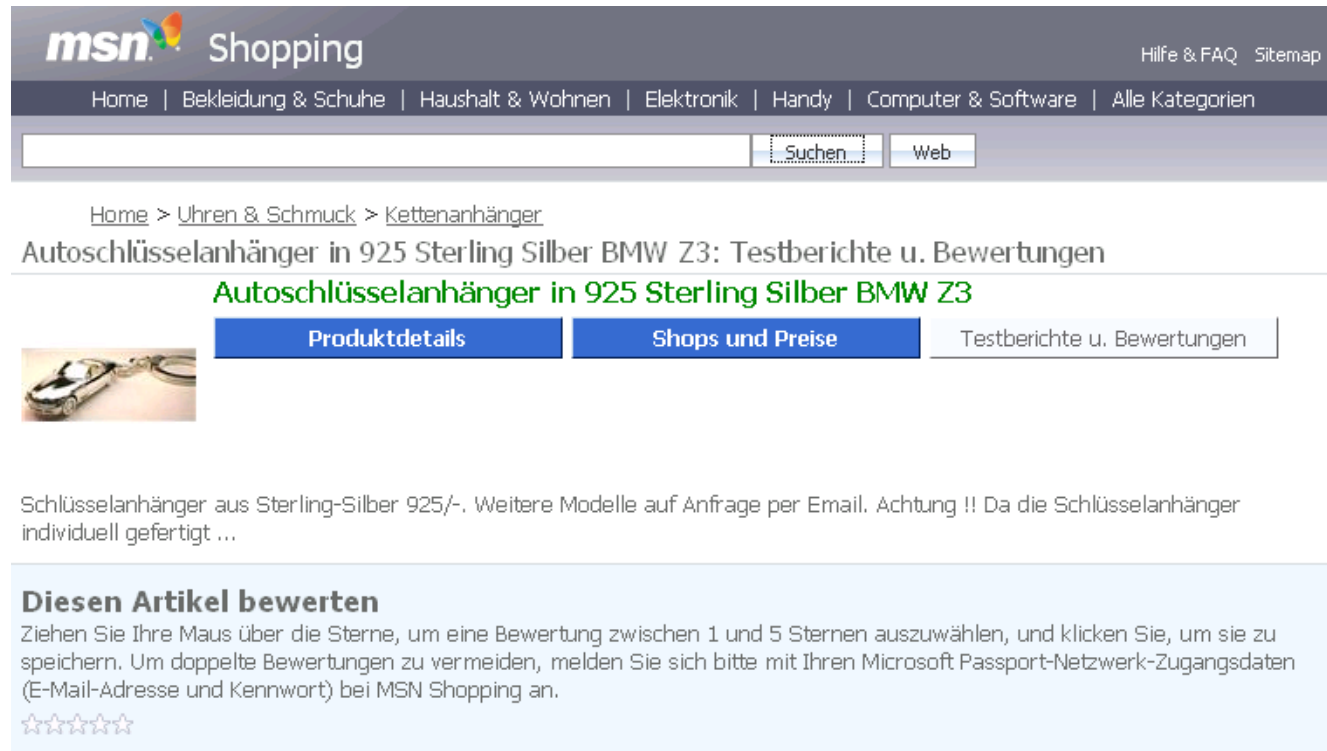
Der BMW Z3 war der erste in den USA produzierte BMW. Ab 1996 lief der Z3 Roadster im eigens dafür gebauten Werk in Spartanburg, South Carolina vom Band. Der BMW Z3 war zunächst nur mit Vierzylindermotoren erhältlich, deren Leistung 115 PS und 140 PS betrug. 1997 wurde die Motorenpalette des BMW Z3 um einen 192 PS starken 2,8-Liter-Reihensechszylinder erweitert. Der BMW Z3 2.8 war äußerlich durch sein breiteres Heck zu erkennen. Zur gleichen Zeit kam auch der 321 PS starke BMW M Roadster in den Handel. 1997 debütierte dann das zweiseitige BMW Z3 Coupé, das im englischen Sprachraum auch unter dem Namen Z3 Shooting Brake bekannt wurde. 1999 bekam der BMW Z3 Roadster ein Facelift, erkennbar an neu geformten Rückleuchten und stärker ausgestellten Kotflügeln, mit Hüftschwung, im Heck. Der Vierzylindermotor mit 140 PS wurde durch einen 2,0-Liter-Sechszylinder mit 150 PS ersetzt. Im Jahr 2000 hat BMW das Motorenkonzept des Z3 nochmals überarbeitet: Der Z3 2.0 mit 150 PS wurde durch den BMW Z3 2.2i mit 170 PS ersetzt, der Z3 2.8 mit 193 PS durch den BMW Z3 3.0i mit 3,0-Liter-Motor und 231 PS. Der M Roadster wurde über die komplette Bauzeit mit altem Heck weitergebaut. 2002 wurde der BMW Z3 schließlich vom BMW Z4 abgelöst. Sie möchten einen BMW Z3 Roadster kaufen oder verkaufen? Nutzen Sie mobile.de, den großen Online-Marktplatz für Fahrzeuge aller Art.

- **QC = {bmw, z3, roadster, coupe, cabrio, gebrauchtwagen, gebraucht, fahrbericht}**

Context Matching - Example

28

$$CMC_{bmw} = \frac{(1 + 0 + 0 + 0 + 0 + 0 + 0)}{(1 + 1 + 1 + 1 + 1 + 1 + 1)} \approx 0,143$$



The screenshot shows the MSN Shopping interface. At the top, there's a navigation bar with 'msn Shopping' and links for 'Hilfe & FAQ' and 'Sitemap'. Below that, a category menu includes 'Home', 'Bekleidung & Schuhe', 'Haushalt & Wohnen', 'Elektronik', 'Handy', 'Computer & Software', and 'Alle Kategorien'. A search bar with 'Suchen' and 'Web' buttons is present. The breadcrumb trail reads 'Home > Uhren & Schmuck > Kettenanhänger'. The main heading is 'Autoschlüsselanhänger in 925 Sterling Silber BMW Z3: Testberichte u. Bewertungen'. Below this, there's a sub-heading 'Autoschlüsselanhänger in 925 Sterling Silber BMW Z3' and three buttons: 'Produktdetails', 'Shops und Preise', and 'Testberichte u. Bewertungen'. An image of a car keychain is shown. The text below the image reads: 'Schlüsselanhänger aus Sterling-Silber 925/-. Weitere Modelle auf Anfrage per Email. Achtung !! Da die Schlüsselanhänger individuell gefertigt ...'. At the bottom, there's a section 'Diesen Artikel bewerten' with instructions on how to rate the product and a star rating display.

- QC = {bmw, z3, roadster, coupe, cabrio, gebrauchtwagen, gebraucht, fahrbericht}

Agenda

29

- Motivational Example
- Traditional Scoring
- Term Context?
- Scoring with Context Matching
- Context Matching – Example
- **Traditional vs. Context Matching**

Traditional vs. Context Matching

30

- **Experiments run on a benchmark system**
 - **TREC web document collection WT2g**
 - **247.491 documents**
 - **50 pre-defined queries**
 - **Relevance judgment for each query**
 - **Standard inverted index with**
 - Document ID
 - Term frequencies
 - **Positions of all terms**

Traditional vs. Context Matching(CM)

31

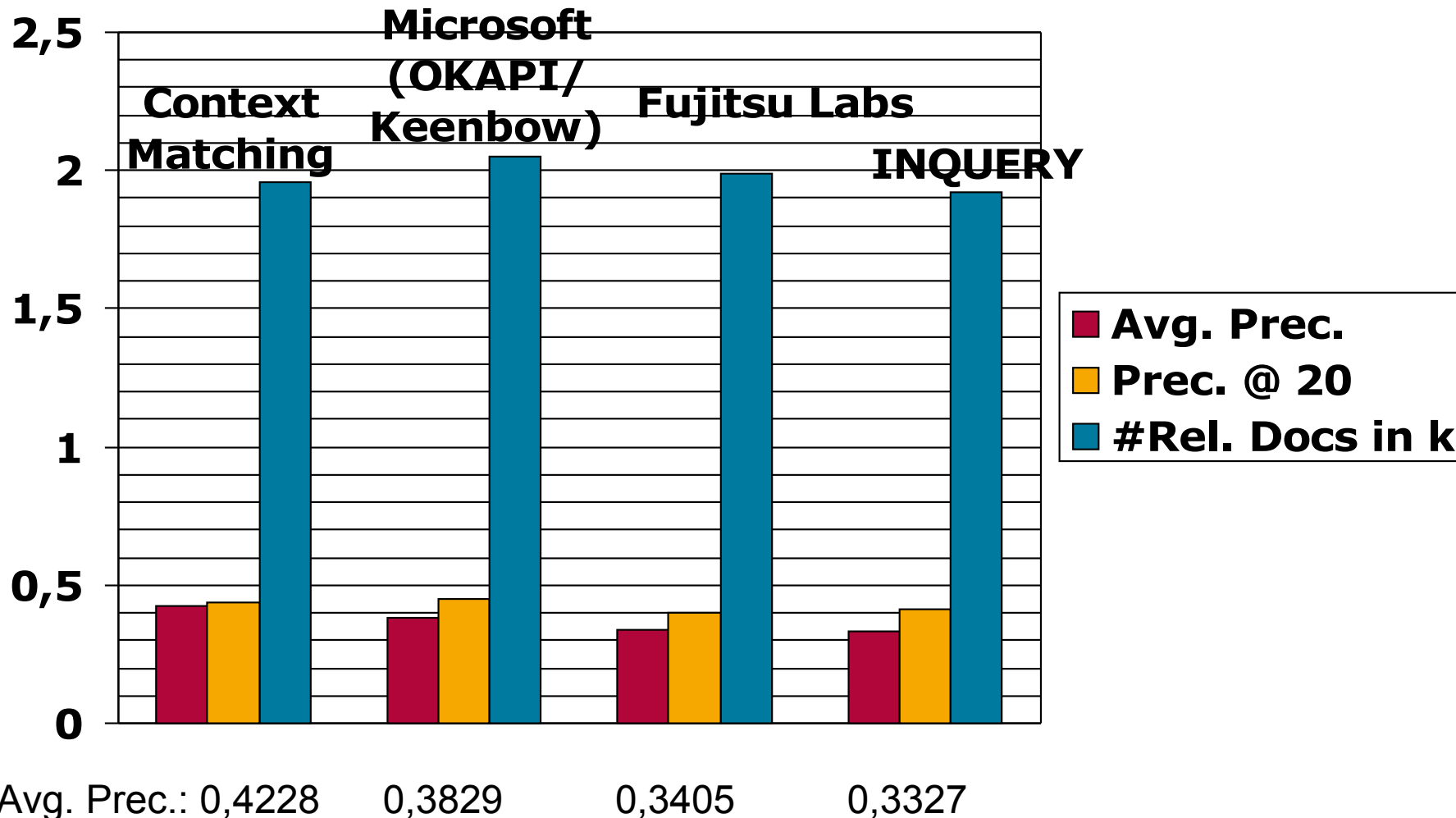
- **Experimental runs** to configure parameters of CM
 - **Weighting factor for TF & CMC in TC**
 - **Weighting factor for Q & QR in QC**
 - **d**
 - **Dist(CD)**
 - **m**
 - **n**
 - **R**

- **More information about ideal parameter values available, if time permits**

Traditional vs. Context Matching(CM)

32

■ Comparison with top 3 algorithms



Traditional vs. Context Matching(CM)

33

- **41% improvement over simple TF-IDF scoring**
- **But \approx 30% slower than TF-IDF**
 - **Not counting QR computation**
 - **Closest distance computation rather time-consuming**
- **Slightly less recall than previous best results**
- **However, average precision 10% better than previous best results**

Conclusion

34

- **Very interesting & promising approach**
- **First try already better than sophisticated implementations of TF-IDF**
- **Integrates TF-IDF natively**
 - **Able to take advantage of many experiences there**

- **Zakos, J. and Verma, B. 2006. A Novel Context-based Technique for Web Information Retrieval. *World Wide Web* 9, 4 (Dec. 2006), 485-503.**
- **Robertson, S.: On term selection for query expansion. *J. Doc.* 46(4):359–364 (1990)**
- **Robertson, S., Walker, S.: Okapi/Keenbow at TREC-8. In: *Proceedings of the 8th Text Retrieval Conference (TREC-8)*, Gaithersburg, USA, pp. 151–161 (1999)**
- **Yu, S., Cai, D., Wen, J., Ma, W.: Improving pseudo-relevance feedback in web information retrieval using web page segmentation. In: *Proceedings of the 12th International Word Wide Web Conference* (2003)**
- **„BMW Z3“, Google, 04.01.09, www.google.de**

Backup

36

Ideal Parameter Values

37

- **Weighting factor for TF & CMC in TC = 0,5**
- **Weighting factor for Q & QR in QC = 0,5**
- **D = 250**
- **Dist(CD) = linear**
- **m = 7**
- **n = 20**
- **R_t = 1 by default**

- TC – term confidence, score indicating the importance of a query term in a given document
- CMC – context matching confidence
- TF – term frequency
- IDF – inverse document frequency
- DC – document collection
- ID – documents retrieved in initial run
- Q – query on the DC (e.g. „bmw z3“)
- QR – expanded terms (m most frequent terms in n most relevant documents out of ID)
- QC – query context = Q + QR
- R – relatedness, for each term in QR for how related it is to Q

- **TSV_t** – term significance value, used for retrieving QR ($=w*r$)
- **w** – weight of term t , IDF over DC
- **r** – number of documents in ID, where t appears
- **m** – number of words used for QR (taken from ID)
- **n** – size of ID collection
- **$CI_{q,C,D}$** – contextual importance of q in D , with context C
- **$CD_{q,c,D}$** – minimum distance between all occurrences of q and c in D
- **Dist(CD)** – function for distance importance
- **d** – maximum context distance
- **$CMC_{q,D}$** – gewichtet $CI_{q,Q,D}$ + gewichtet $CI_{q,QR,D}$
- **$TC_{q,D}$** – gewichtet TF + gewichtet CMC