



Betrachtung der Suchmaschine "Verity" des Delivery Servers  
Sabine van Almsick, Miroslawa Santur

Bremen, den 16.02.2009

# Agenda

- » **MEDIAL in Kürze**
- » **Suchmaschinen und ihre Einsatzfelder**
- » **Verity – Systemarchitektur**
- » **Verity – Performance**
- » **Die 10. „goldenen“ Regeln**
- » **Fazit**

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# MEDIAL in Kürze

**MEDIAL**  
GESELLSCHAFT FÜR DIGITALE MEDIEN MBH

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# MEDIAL in Kürze

## » 1995 Gründung

- > der MEDIAL GmbH durch die beiden Gesellschafter und Geschäftsführer Sabine van Almsick & Dietmar Zeschau

## » Standort

- > in Köln Konrad-Adenauer-Ufer 39, Innenstadt Nähe Dom/Hauptbahnhof

## » Seit 1998 Ausbildungsbetrieb

- > für Fachinformatiker und Informatikkaufleute

## » Seit 2000 RedDot-Partner, seit 2006 RedDot-Premium-Partner

- > Spezialisierung auf Full-Service RedDot-Projekt-Umsetzungen
- > Jährlich werden über 30 Projekte mit RedDot-CMS / Live-Server umgesetzt und weiter ausgebaut.
- > Schwerpunkte: Internationalisierung, Performance-Optimierung, Konsolidierung

## » Mitarbeiterzahl 2009

- > 30 feste Mitarbeiter

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Suchmaschinen und ihre Einsatzfelder

**MEDIAL**  
GESELLSCHAFT FÜR DIGITALE MEDIEN MBH

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Suchmaschinen und ihre Einsatzfelder



## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Verity – Systemarchitektur

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Verity – Systemarchitektur



Delivery Server  
Verity

Vorteil:

» Keine Indizierungsprobleme

Nachteil:

» Keine Ausfallsicherheit

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit



# Verity – Systemarchitektur



Delivery Server  
Verity

Vorteil:

» Keine Indizierungsprobleme

Nachteil:

» Keine Ausfallsicherheit



Delivery Server  
Verity



Delivery Server

» Keine Indizierungsprobleme

» Ausfallsicherheit für die Suche nicht gewährleistet

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Verity – Systemarchitektur



Delivery Server  
Verity

Vorteil:

» Keine Indizierungsprobleme

Nachteil:

» Keine Ausfallsicherheit



Delivery Server  
Verity



Delivery Server

» Keine Indizierungsprobleme

» Ausfallsicherheit für die Suche  
nicht gewährleistet



Delivery Server  
Verity



Delivery Server  
Verity

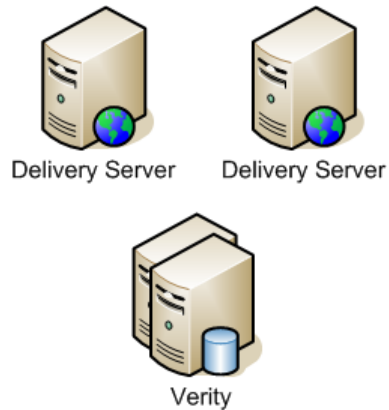
» Ausfallsicherheit auch für  
die Suche sichergestellt

» Ggf. Differenzen bei der  
Indizierung

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Verity – Systemarchitektur



## Vorteil:

- » Keine Indizierungsprobleme
- » Keine Differenzen bei der Indizierung

## Nachteil:

- » Zusätzliche Kosten

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Verity – Performance

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 1. Nur das indizieren, was man wirklich finden möchte

- > Suchkonzept
  - > Was möchte ich suchen?
  - > Wie soll dies aufbereitet werden?
  
- > BLOB-Elemente
  - > sinnvoll in Asset Managern aufteilen
  - > Nur auf expliziten Wunsch mit durchsuchen (erweiterte Suche)
  
- > Bereiche von der Suche ausschließen
  - > `<rd-dm:query searchable="true"></rd-dm:query>` bzw.  
`<rd-dm:query searchable="false"></rd-dm:query>`

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 2. Jede Sprachvariante einzeln indizieren

- > Die Bordmittel des Delivery Servers nutzen
  - > Sprachvarianten in die entsprechende Sprache publizieren
  - > Suche mit „`locale`“ einschränken

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 3. Stopplisten pflegen

- > Bei Indizieren der Inhalte werden für die Volltextsuche keine Begriffe indiziert, welche in der Stoppliste zu finden sind
  - > Pronomen, Adverbien, etc
  - > z.B.: der, die, das, ich, du, in, an, um, wo, wie, etc.

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 4. Anzahl der Treffer beschränken

- > Max. Anzahl beschränken
  - > Wer bei 200 Treffern nichts findet, sollte seinen Suchstring anpassen
- > Schnellsuche
  - > max. 20 Treffer
  - > Link „weitere Treffer anzeigen“

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit



## 5. Keine speziellen Such-Funktionalitäten verwenden

- › Thesaurus, phonetische Suche usw. nur bei der erweiterten Suche anbieten
- › Sinnvoll prüfen, welche der Einstellungen gebraucht werden
  - › Bring <THESAURUS> oder <TYPO/2> das gewünschte Ergebnis

```
-<rde-dm:query-verity>  
(<Mult/20000> (pageHeadline <CONTAINS> *info*) <AND> *info* ) <OR> ( (CIS_TYPE <CONTAINS> 'HTML') <AND> <MANY> *info* ) <OR> ( <CIS_TYPE <CONTAINS> 'XML' ) <AND> <MANY> *info* ) <OR> (CIS_TYPE <CONTAINS> 'BLOB') <AND> <Mult/9000> (filename <CONTAINS> *info*) <OR> (<Mult/5000> (CIS_TYPE <CONTAINS> 'BLOB') <AND> <MANY> *info* ) <OR> (CIS_TYPE <CONTAINS> 'BLOB') <AND> <Mult/3000> (<STEM> *info* ) <OR> (CIS_TYPE <CONTAINS> 'BLOB') <AND> <Mult/3000> (<THESAURUS> *info*) <OR> ( <STEM> *info* ) <OR> (<THESAURUS> *info*)  
</rde-dm:query-verity>
```

```
<rde-dm:query-verity>info <AND> (CIS_TYPE <CONTAINS> HTML)</rde-dm:query-verity>
```

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 6. Inhalte, die im Ergebnis dargestellt werden sollen, in Inhaltsattributen vorbereiten

- > Volltextsuche ist schneller als eine Suche über Inhaltsattribute
- > Das Ausliefern der Ergebnisse aus den Inhaltsattributen ist jedoch schneller als das Parsen aus dem HTML-Text.
  - > Überschrift, Teasertext und weitere Informationen über Inhaltsattribute der Seite zu ordnen
  - > Standardattribute der Suche nutzen

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 7. Berechtigung über die Verity abfragen und nicht zusätzlich im Delivery Server überprüfen

- > Die Suche kann beschleunigt werden, wenn man auf das Prüfen der Inhaltsbedingungen verzichtet
- > Bei Projekten mit Zugriffsbeschränkungen müssen diese in der Query abgebildet werden

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

## 8. Ausgabe im XML nur auf das wesentliche beschränken

- › Nur das Ausgeben, was man braucht
  - › context-tags nicht mit ausgeben
  - › Alles über Inhaltsattribute abbilden

```
-<pages chunk="1" chunksize="10" context-mode="mixed" context-tags="title,error" external-only="ignore" highlight="no" include-mode="context-tags;content;content-info;metainfo" lastchunk="1" locale="de" maxhits="100" metainfo="score,xml;pageTitle,xml;pageURL,xml;pageType,xml;VDKSUMMARY,xml" nextchunk="2" project="wago" searchengine="verity">
  <page constraints="no-constraints" hit="1" locale="de" name="html/112.htm" project="wago">
    <locale>de</locale>
    <type>HTML</type>
    <name>112.htm</name>
    <released>true</released>
    <guid>
      standardcontentimpl-c0a89ab4-11db3e7c31e-35f9f9cf8a905087
    </guid>
    <group>html</group>
    <description/>
    <last-updated>2008-12-15 09:33:53.543</last-updated>
    <context>
      Suche function actPrint() { vorschau = window.open(document.URL
      (window_name="printversion"); document.getElementById("main
      /#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/#/-/!
      </context>
    <title>Suche</title>
    <VDKSUMMARY>
      www.wago.com/index.html Diese Seite drucken Sie sind hier: Hilfe
    </VDKSUMMARY>
    <pageTitle>Suche</pageTitle>
    <score>0.8658</score>
    <pageType/>
    <pageURL/>
    /#/#RDE-REQUEST:rdePrefix/#/xchg/#RDE-REQUEST:rdeSessionID/#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/
    </pageURL>
    <page/>
    </page>
  </page>
  <page constraints="no-constraints" hit="2" locale="de" name="html/557.006.htm" project="wago">
    <locale>de</locale>
    <type>HTML</type>
    <name>112.htm</name>
    <released>true</released>
    <guid>
      standardcontentimpl-c0a89ab4-11db3e7c31e-35f9f9cf8a905087
    </guid>
    <group>html</group>
    <description/>
    <last-updated>2008-12-15 09:33:53.543</last-updated>
    <VDKSUMMARY>
      www.wago.com/index.html Diese Seite drucken Sie sind hier: Hilfe Suche 2008 WAGO GmbH & Co.KG, Minden
    </VDKSUMMARY>
    <pageTitle>Suche</pageTitle>
    <score>0.8658</score>
    <pageType/>
    <pageURL/>
    /#/#RDE-REQUEST:rdePrefix/#/xchg/#RDE-REQUEST:rdeSessionID/#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/
    </pageURL>
    <page/>
    </page>
  </page>
</pages>
<!-- rde-html-section -->
<!-- rde-rd:content="html/112.htm" rde-rd:datallanguage="de" rde-rd:locale="de" rde-rd:project="wago" rde-rd:xmlSeparator="/" -->
<!-- rde-rd:content="html/557.006.htm" rde-rd:datallanguage="de" rde-rd:locale="de" rde-rd:project="wago" rde-rd:xmlSeparator="/" -->
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<meta http-equiv="content-type" content="text/html; charset=utf-8" />
<title>Suche</title>
<meta name="keywords" content="" />
<meta name="description" content="Suche" />
<meta name="author" content="WAGO Kontakttechnik GmbH & Co. KG" />
<meta http-equiv="pragma" content="no-cache" />
<meta http-equiv="cache-control" content="no-cache" />
<meta name="revisit-after" content="30d" />
<meta http-equiv="cache-control" content="no-cache" />
<meta name="shortcut icon" href="/#/#RDE-REQUEST:rdePrefix/#/xchg/#RDE-REQUEST:rdeSessionID/#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/#/-/static/favicon.ico" type="image/x-icon" />
<link rel="stylesheet" type="text/css" href="/#/#RDE-REQUEST:rdePrefix/#/xchg/#RDE-REQUEST:rdeSessionID/#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/#/-/static/special.css" media="screen" id="spec" />
<link rel="stylesheet" type="text/css" href="/#/#RDE-REQUEST:rdePrefix/#/xchg/#RDE-REQUEST:rdeSessionID/#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/#/-/static/navigation.css" media="screen" id="navicss" />
<link rel="stylesheet" type="text/css" href="/#/#RDE-REQUEST:rdePrefix/#/xchg/#RDE-REQUEST:rdeSessionID/#/#RDE-REQUEST:rdeProjectID/#/#RDE-REQUEST:rdeXsID/#/-/static/navigation.css" media="screen" id="navicss" />
</head>
<body>
</body>
</html>
</rde-html-section>
```

### Agenda

- › MEDIAL in Kürze
- › Suchmaschinen und ihre Einsatzfelder
- › Verity – Systemarchitektur
- › Verity – Performance
- › Die 10. „goldenen“ Regeln
- › Fazit

## 9. Highlighting ausschalten

- > Keine grafische Aufbereitung von Suchtreffern im Summary-Text
- > Das Versehen des HTML-Codes mit einer zusätzlichen CSS-Klasse kostet Performance

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

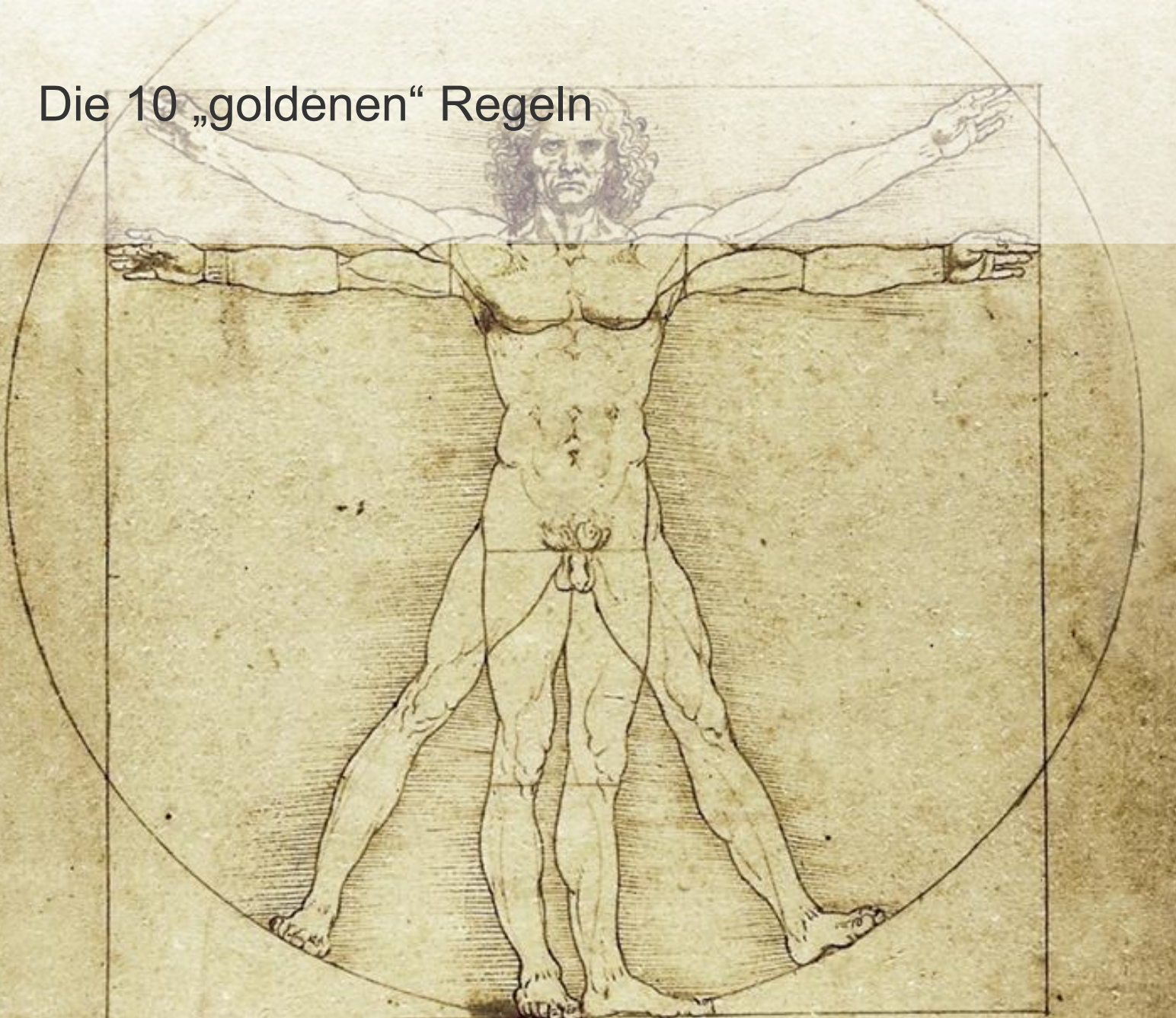
## 10. Caching einschalten und Suche über PSX-Module einbinden

- > Das Anlegen des Cache-Eintrags hat keine Auswirkungen auf das Antwortverhalten
- > Große Vorteile besonders bei häufig gesuchten Begriffen und beim Paging

### Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Die 10 „goldenen“ Regeln



## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Die 10 „goldenen“ Regeln

1. Nur das indizieren, was man wirklich finden möchte
2. Jede Sprachvariante einzeln indizieren
3. Stopplisten pflegen
4. Anzahl der Treffer beschränken
5. Keine speziellen Such-Funktionalitäten verwenden
6. Inhalte, die im Ergebnis dargestellt werden sollen, in Inhaltsattributen vorbereiten
7. Berechtigung über die Verity abfragen und nicht zusätzlich im Delivery Server überprüfen
8. Ausgabe im XML nur auf das wesentliche beschränken
9. Highlighting ausschalten
10. Caching einschalten und Suche über PSX-Module einbinden

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit



# Fazit



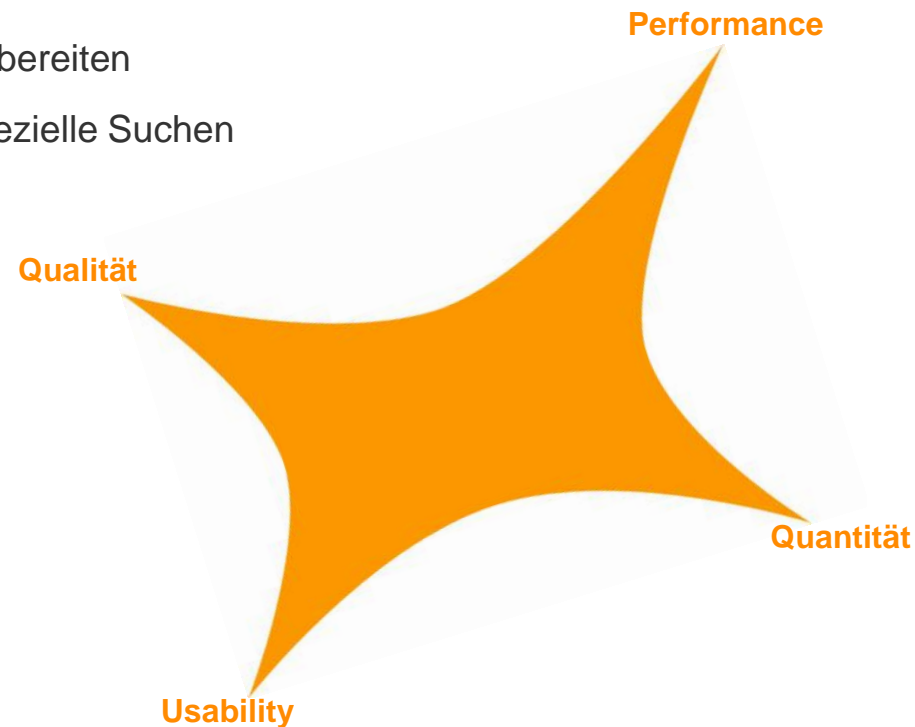
## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » **Fazit**

# Fazit

## Es gibt kein Allheilmittel

- » Suche immer auf das Wesentliche beschränken
- » Dokumente für die Suche aufbereiten
- » Aufteilung in einfache und spezielle Suchen



## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

# Vielen Dank

## Agenda

- » MEDIAL in Kürze
- » Suchmaschinen und ihre Einsatzfelder
- » Verity – Systemarchitektur
- » Verity – Performance
- » Die 10. „goldenen“ Regeln
- » Fazit

### Ansprechpartnerin:

Sabine van Almsick  
MEDIAL – Gesellschaft für digitale Medien mbH  
Konrad-Adenauer-Ufer 39  
50668 Köln  
Tel: 0221 399603-0  
Sabine.vanAlmsick@medial.info