

Examining the **Verity** Search Engine in the Delivery Server

London, 19/05/2009

Agenda

- » **MEDIAL in a Nutshell**
- » **Search Engines and Applications**
- » **Verity System Architecture**
- » **Verity Performance**
- » **10 Performance Principles**
- » **Conclusion**

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

MEDIAL in a Nutshell

MEDIAL
GESELLSCHAFT FÜR DIGITALE MEDIEN MBH

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

MEDIAL in a Nutshell

» **Company Foundation 1995**

- > MEDIAL GmbH by the two partners and Managing Directors

Sabine van Almsick

(graduate in Business Information Technology)

Dietmar Zeschau

(graduate in Economics, major: Commercial Information Technology)

» **Location**

- > Cologne, Konrad-Adenauer-Ufer 39
City centre near the Cathedral/Main Station

» **Employees in 2009**

- > 30 employees:
2 executives, 5 project managers, 18 developers,
2 graphic designers, 2 editors, 1 secretary

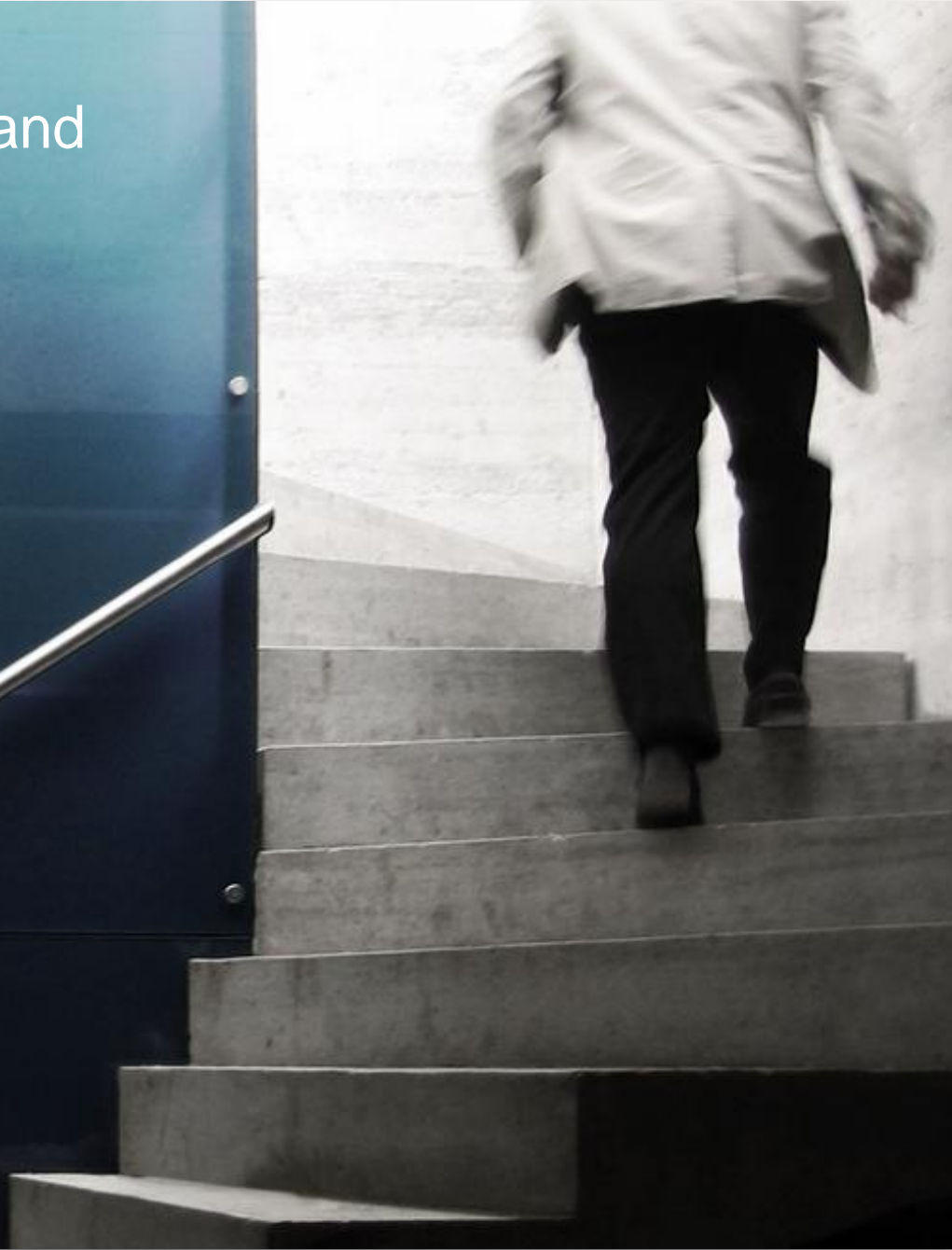
» **Training Company since 1998**

- > For IT specialists and students of business studies

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Search Engines and Applications



MEDIAL
GESELLSCHAFT FÜR DIGITALE MEDIEN MBH

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Search Engines and Applications



Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity System Architecture

MEDIAL
GESELLSCHAFT FÜR DIGITALE MEDIEN MBH

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity System Architecture



Delivery Server
Verity

Pro:

» No indexing problems

Con:

» No reliability

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity System Architecture



Delivery Server
Verity

Pro:

» No indexing problems

Con:

» No reliability



Delivery Server
Verity



Delivery Server

» No indexing problems

» No search engine reliability

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity System Architecture



Delivery Server
Verity

Pro:

» No indexing problems

Con:

» No reliability



Delivery Server
Verity



Delivery Server

» No indexing problems

» No search engine reliability



Delivery Server
Verity



Delivery Server
Verity

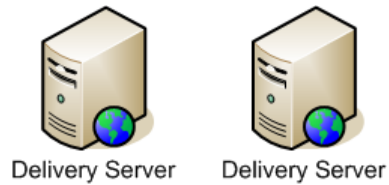
» Search engine reliability
guaranteed

» Potentially minor indexing
differences

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity System Architecture



Pro:

- » No indexing problems
- » No index differences

Con:

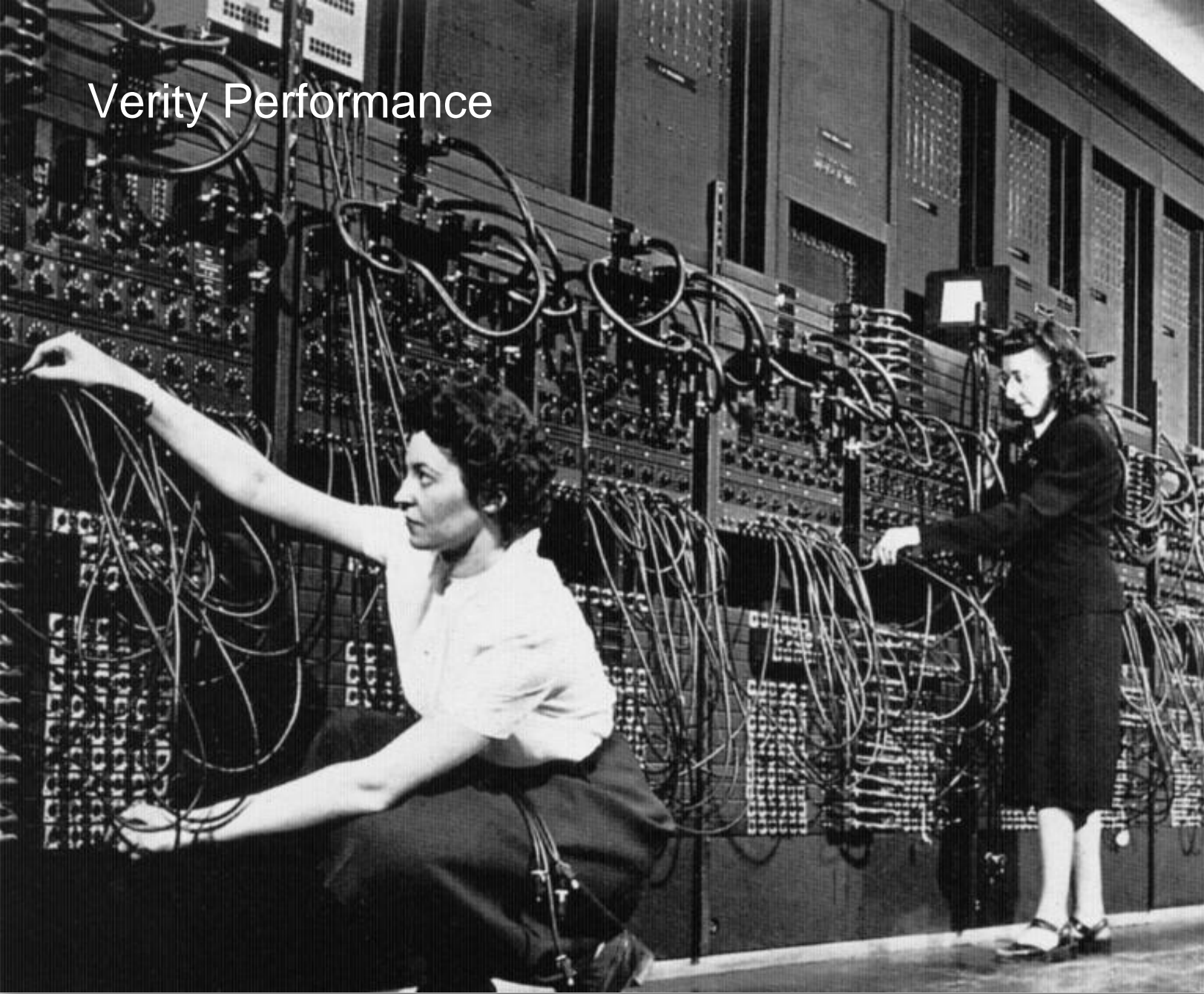
- » Extra costs



Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity Performance



Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

1. Index only what you want to find

- > Search strategy
 - > What to find?
 - > How to present it?
- > BLOB items
 - > Reasonable partitioning of Asset Managers
 - > Search only on explicit requests (extended search)
- > Exclude content areas from searching
 - > `<rd-dm:query searchable="true"></rd-dm:query>`
respectively
`<rd-dm:query searchable="false"></rd-dm:query>`

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity Performance

2. Index language variants separately

- > Utilise the Delivery Server
 - > Publish language variants to their according language
 - > Narrow search by "locale"

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

3. Maintain stop word lists

- > Stop words will not be indexed for the full-text search
 - > Pronouns, adverbs or definite and indefinite articles etc.
 - > Examples: *the, it, and, you, me, where, why, this, that...*

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

4. Limit the number of hits

- > Define a maximum number of hits
 - > “Modify your search term if you can’t find it within 200 hits“

- > Quick search
 - > max. 20 hits
 - > Link to [More hits...](#)

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

5. Do not apply extra search features by default

- › Thesaurus, phonetic search etc. are provided only in the extended search
- › Evaluate the settings `<THESAURUS>` or `<TYPO/2>`
Do they provide the appropriate results?

Agenda

- › MEDIAL in a Nutshell
- › Search Engines and Applications
- › Verity System Architecture
- › Verity Performance
- › 10 Performance Principles
- › Conclusion

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

```
-<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>
```

6. Prepare contents for search results by adding content attributes

- > *Full-text search* is faster than searching content attributes
- > But *rendering* results from content attributes is faster than parsing the HTML contents
 - > Apply content attributes to pages for headline, teaser text and other information
 - > Use standard attributes

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Verity Performance

7. Check authorisation via Verity, not via the Delivery Server

- > Accelerate the search by skipping the constraints for contents
- > Projects with access constraints should map these constraints to the query

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

8. Limit XML output to the essentials

- Export only what you need
 - Dot not export context-tags
 - Map everything with content attributes

```
<pages chunk="1" chunksize="10" context-mode="mixed" context-tags="title" error="0" external-only="ignore" highlight="none" include-mode="content-tags,content;content-info;metainfo" lastchunk="2" locale="de" maxhits="100" meta-info="score,xml;pageName,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,"printversion",
      (window.name=="printversion"), document.getElementById("maincss"),href="#"#RD
      /#/#RDE-REQUEST.rdeProjectID/#/#RDE-REQUEST.rdeXsIID/#/-/html/print.css");
  </context>
  <title>Suche</title>
  <VDKSUMMARY>
    www.wago.com/index.html Diese Seite drucken Sie sind hier: Hilfe Suche 2008 WAG
  </VDKSUMMARY>
  <pageName>Suche</pageName>
  <score>0.8658</score>
  <pageType/>
  <pageURL>
    /#/#RDE-REQUEST.rdePrefix/#/xchg/#RDE-REQUEST.rdeSessionID/#/#RDE-REQUEST
  </pageURL>
  <rde-html rde-rd:content="html/112.htm" rde-rd:datalanguage="de" rde-rd:dat
  rde-rd:locale="de" rde-rd:project="wago" rde-rd:xs:xmlSeparator="/">
    <rde-html-section>
      <IDCTYPE Html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.or
      PageID 112 - published by RedDot 7.5 - 7.5.1.103 - 25689 --> <meta http-equiv="
      <title>Suche</title> <meta name="keywords" content="" /> <meta name="description" content="
      content="index, nofollow" /> <meta name="author" content="WAGO Kontakttechnik GmbH & am
      content="WAGO Kontakttechnik GmbH & Co. KG" /> <meta http-equiv="pragma" content="n
      content="30" /> <meta http-equiv="cache-control" content="no-cache" /> <meta name="revisi
      rel="shortcut icon" href="/#/#RDE-REQUEST.rdePrefix/#/xchg/#RDE-REQUEST.rdeSessionID/#/#RDE
      /favicon.ico" type="image/x-icon" /> <link rel="stylesheet" type="text/css" href="/#/#RDE-REQUE
      REQUEST.rdeSessionID/#/#RDE-REQUEST.rdeProjectID/#/#RDE-REQUEST.rdeXsIID/#/-/static/ma
      <link rel="stylesheet" type="text/css" href="/#/#RDE-REQUEST.rdePrefix/#/xchg/#RDE-REQUEST.rc
      REQUEST.rdeProjectID/#/#RDE-REQUEST.rdeXsIID/#/-/static/special.css" media="screen" id="spe
      type="text/css" href="/#/#RDE-REQUEST.rdePrefix/#/xchg/#RDE-REQUEST.rdeSessionID/#/#RDE-R
      REQUEST.rdeXsIID/#/-/static/navigation.css" media="screen" id="navicss"/> <link rel="stylehee
```

Agenda

- MEDIAL in a Nutshell
- Search Engines and Applications
- Verity System Architecture
- Verity Performance
- 10 Performance Principles
- Conclusion

9. Disable highlighting

- > No emphasising of search terms in the hit summary
- > Omit additional CSS classes in the HTML code

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

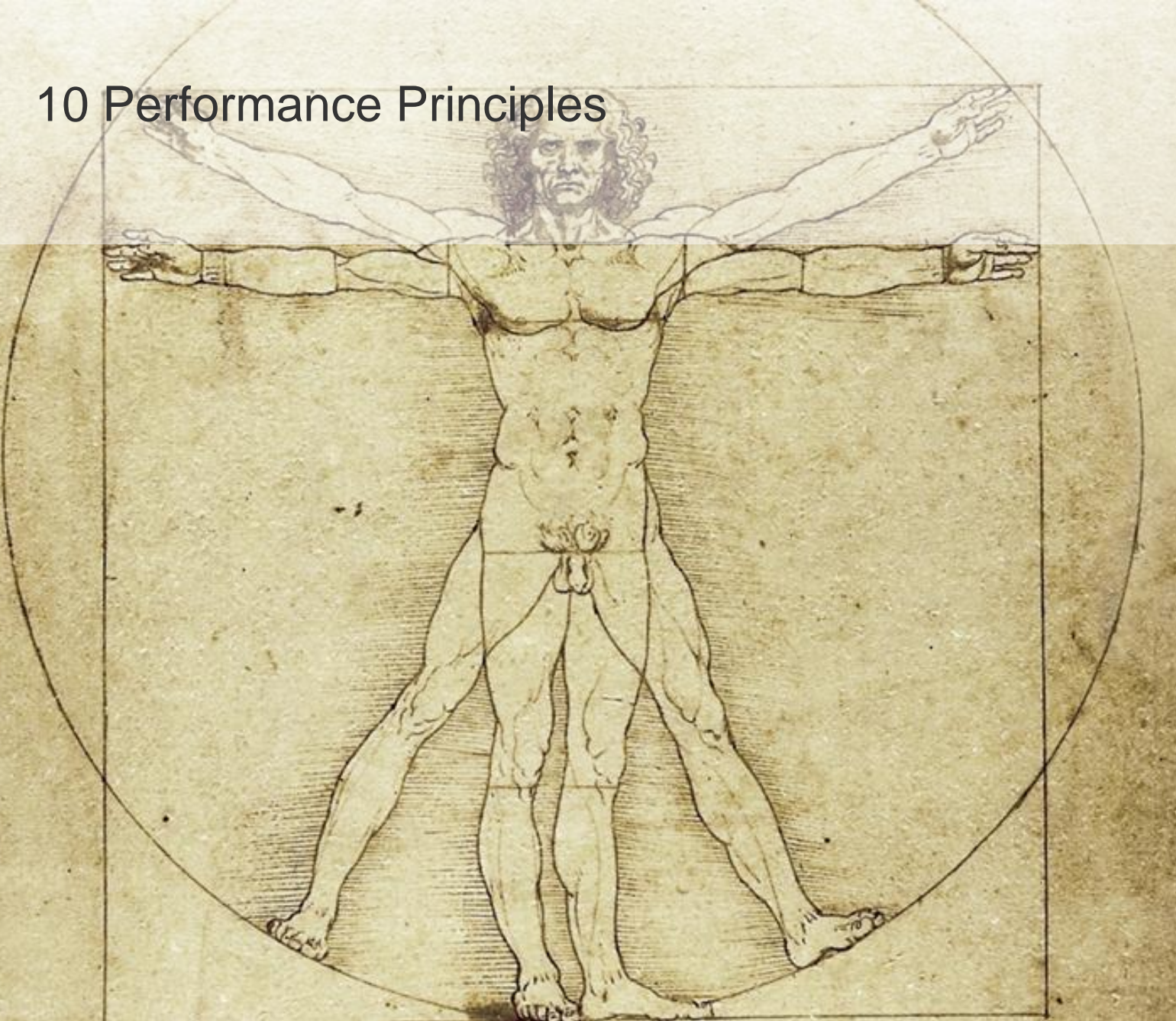
10. Enable caching and include PSX modules

- > Creating cache entries does not affect response times
- > Both accelerate frequently searched terms and paging

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

10 Performance Principles



Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

10 Performance Principles

1. Index only what you want to find
2. Index language variants separately
3. Maintain stop word lists
4. Limit the number of hits
5. Do not apply extra search features
6. Prepare contents for search results by adding content attributes
7. Check authorisation with Verity not the Delivery Server
8. Limit XML output to essentials
9. Disable highlighting
10. Enable caching and include PSX modules

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Conclusion



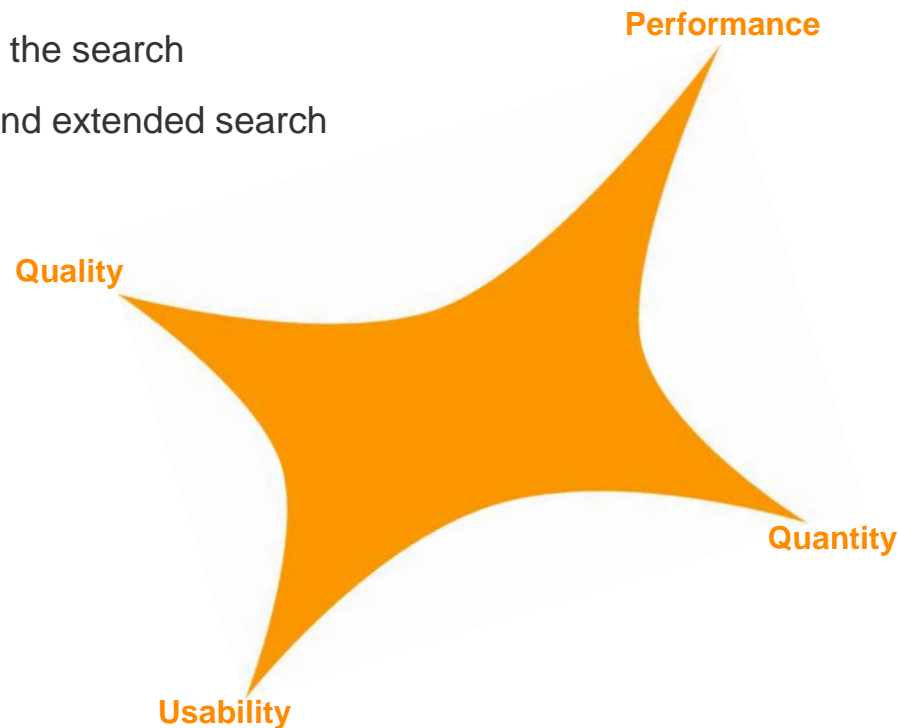
Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Conclusion

There is no silver bullet

- » Limit the search to the basics
- » Prepare your documents for the search
- » Distinguish between quick and extended search



Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Thank you!

Agenda

- » MEDIAL in a Nutshell
- » Search Engines and Applications
- » Verity System Architecture
- » Verity Performance
- » 10 Performance Principles
- » Conclusion

Contact:

Philip Knott
MEDIAL – Gesellschaft für digitale Medien mbH
Konrad-Adenauer-Ufer 39
D-50668 Köln
Tel: +49-221 399603-0
Philip.Knott@medial.info