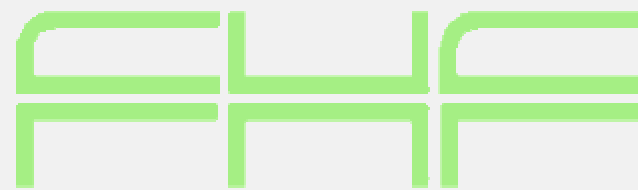


Binäres XML

Fortsetzung oder Ende der <R>Evolution?

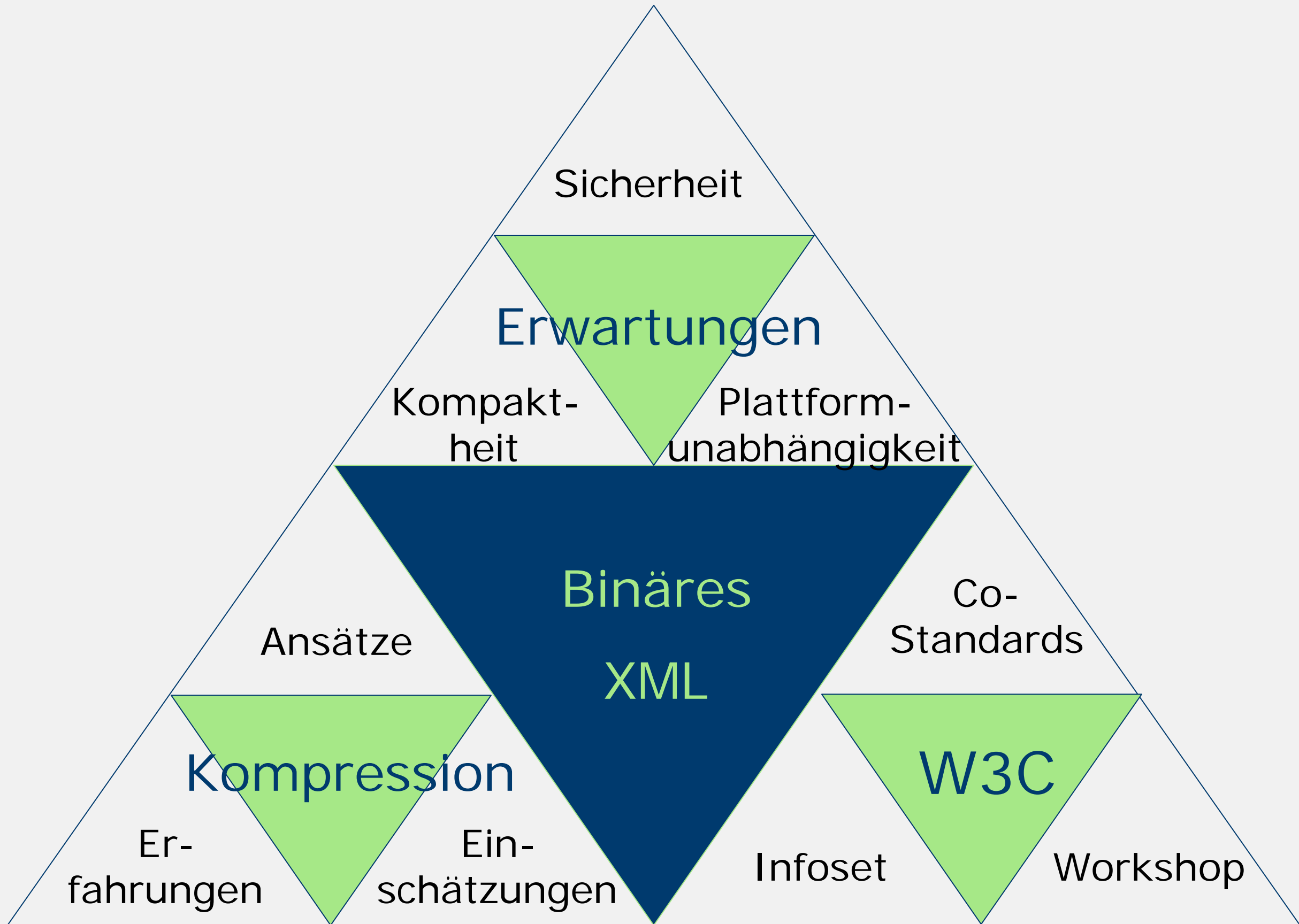


Prof. Mario Jeckle

Fachhochschule Furtwangen

mario@jeckle.de

<http://www.jeckle.de>



XML v1.0/1.1

- I. XML steht für strukturierte Daten
- II. XML sieht ein wenig wie HTML aus
- III. XML ist Text, aber nicht zum Lesen
- IV. XML ist vom Design her ausführlich
- V. XML ist eine Familie von Techniken
- VI. XML ist neu, aber nicht so neu
- VII. XML überführt HTML in XHTML
- VIII. XML ist modular
- IX. XML ist die Basis für RDF und das Semantic Web
- X. XML ist lizenzfrei, plattformunabhängig und gut unterstützt

XML v1.0/1.1

```

<?xml version="1.0"?>
<PLAY>
  <TITLE>The Tragedy of Macbeth</TITLE>
  <FM>
    <P>ASCII text placed in the public domain by Moby Lexical Tools, 1992.</P>
    <P>SGML markup by Jon Bosak, 1992-1994.</P>
    <P>XML version by Jon Bosak, 1996-1999.</P>
    <P>The XML markup in this version is Copyright ©; 1999 Jon Bosak.
This work may freely be distributed on condition that it not be
modified or altered in any way.</P>
  </FM>
  <PERSONAE>
    <TITLE>Dramatis Personae</TITLE>
    <PERSONA>DUNCAN, king of Scotland.</PERSONA>
    <PGROUP>
      <PERSONA>MALCOLM</PERSONA>
      <PERSONA>DONALBAIN</PERSONA>
      <GRPDESCR>his sons.</GRPDESCR>
    </PGROUP>
    <PGROUP>
      <PERSONA>MACBETH</PERSONA>
      <PERSONA>BANQUO</PERSONA>
      <GRPDESCR>generals of the king's army.</GRPDESCR>
  </PERSONAE>

```

Inhaltliche Eigenschaften:

- Verschiedene Encodierungen, wahlfrei Unicode
- Semistrukturierter Inhalt (*mixed content*)
- Typisierter Inhalt (XML-Schema)
- Entitäten (in verschiedenen Varianten)
- Metazeichen („<“, „>“, „““, „’“, „&“)

XML v1.0/1.1

Typ	XML -Schema	Microsoft CLS	Java	SQL92/SQL3
$-2^7 \leq t \leq 2^7$	byte	-	byte	tinyint
$0 \leq t \leq 2^8 - 1$	unsignedByte	byte	-	
$-2^{15} \leq t \leq 2^{15} - 1$	short	short	short	smallint
$-2^{31} \leq t \leq 2^{31} - 1$	int	int	int	int, integer
$-2^{63} < t < 2^{63} - 1$	long	long	long	byte
Unicode-Zeichen(kette)	string	string	String	
IEEE754	float double	float double	float double	float, real, double
{true,false,0,1}	boolean	bool	boolean	boolean
\mathbb{N}	decimal	decimal	BigDecimal	numeric(p,s), dec, decimal(p,s)

Inhaltliche Eigenschaften:

- Verschiedene Encodierungen, wahlfrei Unicode
- Semistrukturierter Inhalt (*mixed content*)
- Typisierter Inhalt (XML-Schema)
- Entitäten (in verschiedenen Varianten)
- Metazeichen („<“, „>“, „““, „'“, „&“)

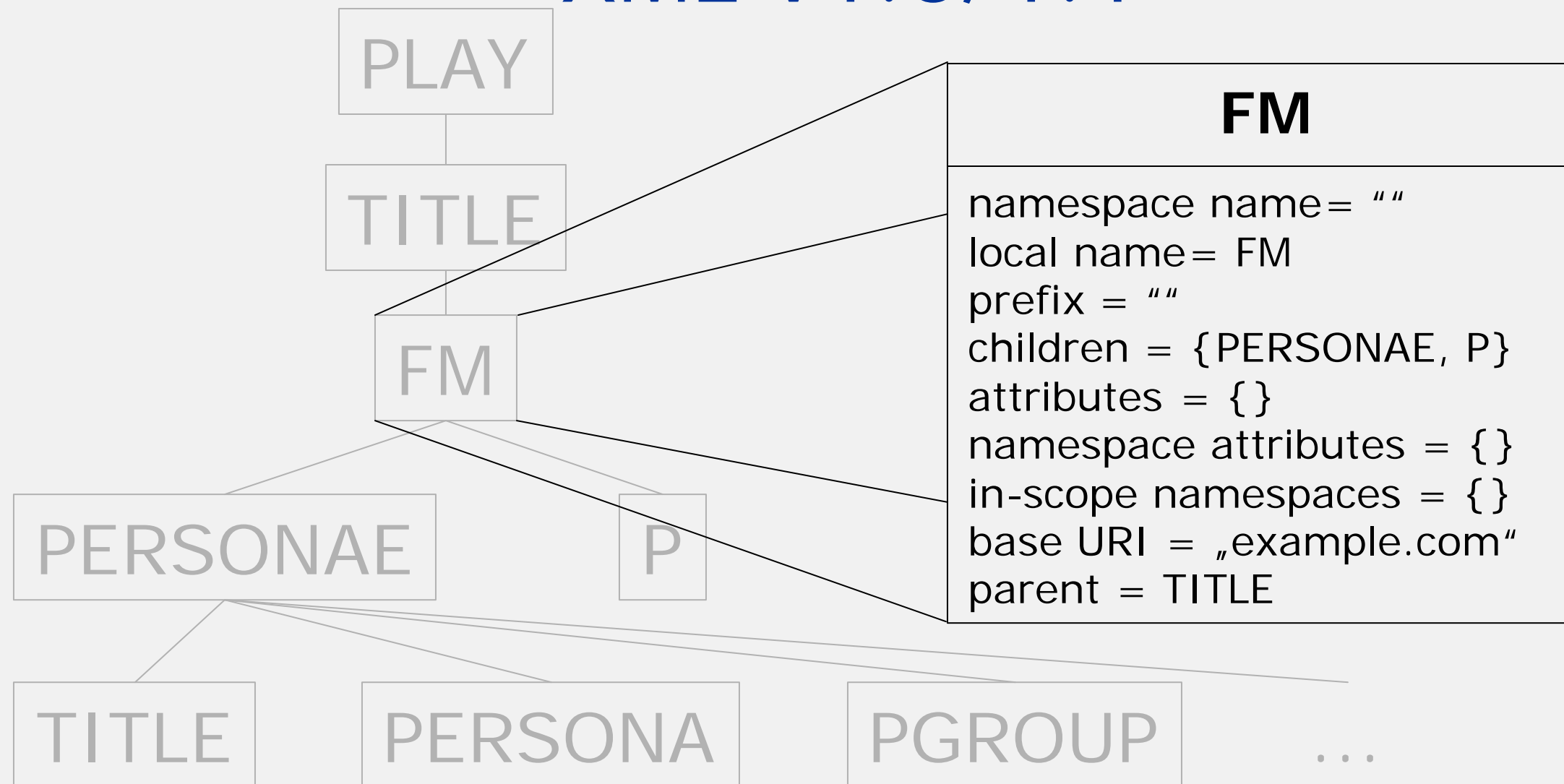
XML v1.0/1.1

```
<?xml version="1.0"?>
<PLAY>
  <TITLE>The Tragedy of Macbeth</TITLE>
  <FM>
    <P>ASCII text placed in the public domain by Moby Lexical Tools, 1992.</P>
    <P>SGML markup by Jon Bosak, 1992-1994.</P>
    <P>XML version by Jon Bosak, 1996-1999.</P>
    <P>The XML markup in this version is Copyright © 1999 Jon Bosak.
This work may freely be distributed on condition that it not be
modified or altered in any way.</P>
  </FM>
  <PERSONAE>
    <TITLE>Dramatis Personae</TITLE>
    <PERSONA>DUNCAN, king of Scotland.</PERSONA>
    <PGROUP>
      <PERSONA>MALCOLM</PERSONA>
      <PERSONA>DONALBAIN</PERSONA>
      <GRPDESCR>his sons.</GRPDESCR>
    </PGROUP>
    <PGROUP>
      <PERSONA>MACBETH</PERSONA>
      <PERSONA>BANQUO</PERSONA>
      <GRPDESCR>generals of the king's army.</GRPDESCR>
```

Strukturelle Eigenschaften:

- Semistrukturierter Inhalt (*mixed content*)
- Inhärent baumartig
- Wohlgeformtheitsregeln
- Anwenderdefinierte Strukturen (XML-Schema)
- Internal Subsets

XML v1.0/1.1



Strukturelle Eigenschaften:

- Semistrukturierter Inhalt (*mixed content*)
- Inhärent baumartig
- Wohlgeformtheitsregeln
- Anwenderdefinierte Strukturen (XML-Schema)
- Internal Subsets

XML Information Set

- Definiert strukturellen Aufbau einer XML-Quelle
 - ... als abstrakte Datenmenge
 - ... als Definitionssammlung
- Jedes wohlgeformte XML-Dokument besitzt einen InfoSet
 - ... auch wenn es nicht als Datei vorliegt
 - ... daher ist diese Spezifikation Basis weiterer W3C-Spezifikationen (XML-Schema, XMLP/SOAP, WSDL, ...)
- W3C Recommendation seit 2001-10-24
 - ... definiert keinerlei programmierparadigmenspezifische Schnittstellen
 - ... operiert auf abstrakten Baumstrukturen
- Wird durch XML-Schemavalidierungsprozeß
 - ... auf strukturelle Konformität geprüft
 - ... mit Typinformation dekoriert

Konsequenzen von XML v1.0 und dem XML Information Set

- Inhalt
 - mit XPath eindeutig lokalisierbar
 - mit XQuery extrahierbar
 - mit XSLT transformierbar
 - mit XML Schema validierbar
- Strukturen
 - mit XQuery konstruierbar
 - mit XSLT transformierbar
 - mit XML Schema validierbar

Binäres XML

Ziele und Mythen

- Sicherheit
- Kompaktheit
- Schnelle Verarbeitung

binäres XML v1.0/1.1

I. XML steht für strukturierte Daten

~~II. XML sieht ein wenig wie HTML aus~~

~~III. XML ist Text, aber nicht zum Lesen~~

~~IV. XML ist vom Design her ausführlich~~

V. XML ist eine Familie von Techniken

VI. XML ist neu, aber nicht so neu

~~VII. XML überführt HTML in XHTML~~

VIII. XML ist modular

IX. XML ist die Basis für RDF und das Semantic Web

X. XML ist lizenzfrei, ~~plattformunabhängig~~ und gut unterstützt

Binäres XML

Sicherheit

- Mythos: Binär = Sicher ...
 - „klassische“ Sicherheitsmechanismen
 - XML-spezifische Sicherheitsmechanismen

Binäres XML

Sicherheit

- Klar ist: für Menschen nicht (direkt) lesbar bedeutet nicht für die Maschine unlesbar
 - Daher: Binär \neq Sicher ...
- Sicherheit für XML kann mit klassischen Mechanismen, die auf beliebigen Eingaben operieren, erreicht werden.
- Für XML-Sicherheit existieren eigenständige Spezifikationen

Binäres XML

Sicherheit

```
<?xml version="1.0"?>
<PLAY>
  <TITLE>The Tragedy of Macbeth</TITLE>
  <FM>
    <P>ASCII text placed in the public domain by Moby Lexical Tools, 1992.</P>
    <P>SGML markup by Jon Bosak, 1992-1994.</P>
    <P>XML version by Jon Bosak, 1996-1999.</P>
    <P>The XML markup in this version is Copyright &#169; 1999 Jon Bosak.
This work may freely be distributed on condition that it not be
modified or altered in any way.</P>
  </FM>
  <PERSONAE>
    <TITLE>Dramatis Personae</TITLE>
    <PERSONA>DUNCAN, king of Scotland.</PERSONA>
    <PGROUP>
      <PERSONA>MALCOLM</PERSONA>
      <PERSONA>DONALBAIN</PERSONA>
      <GRPDESCR>his sons.</GRPDESCR>
    </PGROUP>
    <PGROUP>
      <PERSONA>MACBETH</PERSONA>
      <PERSONA>BANQUO</PERSONA>
      <GRPDESCR>generals of the king's army.</GRPDESCR>
```

PGP:

```
qANQR1DBwk4DeAMSaxcCQrkQDAC8D2DT6p84JHDuTH
mRa0N9PXqfSBDyEASRBOwG

f8W0pL3gm0NIR/Ph/qHsliPRe1wIpd6FKYPf4uRoPSiDYop
orIezWexHXCmjtLSf

3tP3sQTosFJjYZjwzjl/8+z0ghD2PTsnceXMEk+qGXOFA02
Asor0PGPGA44KVoYq

GCirZ9JSZSF/EIEyP9w5H4nllkIU/SZeOqyp8dp+GvWyQ
5J6ZYiEhNypJkyvyTZ

...
```

Binäres XML

Sicherheit

```
<?xml version="1.0"?>
<PLAY>
  <TITLE>The Tragedy of Macbeth</TITLE>
  <FM>
    <P>ASCII text placed in the public domain by Moby Lexical Tools, 1992.</P>
    <P>SGML markup by Jon Bosak, 1992-1994.</P>
    <P>XML version by Jon Bosak, 1996-1999.</P>
    <P>The XML markup in this version is Copyright © 1999 Jon Bosak.
This work may freely be distributed on condition that it not be
modified or altered in any way.</P>
  </FM>
  <PERSONAE>
    <TITLE>Dramatis Personae</TITLE>
    <PERSONA>DUNCAN, king of Scotland.</PERSONA>
    <PGROUP>
      <PERSONA>MALCOLM</PERSONA>
      <PERSONA>DONALBAIN</PERSONA>
      <GRPDESCR>his sons.</GRPDESCR>
    </PGROUP>
    <PGROUP>
      <PERSONA>MACBETH</PERSONA>
      <PERSONA>BANQUO</PERSONA>
      <GRPDESCR>generals of the king's army.</GRPDESCR>
```

XML Encryption:

```
Moby Lexical Tools, 1992.</P>
</P>
</P>
Copyright © 1999 Jon Bosak.
on condition that it not be modified or altered in any way.</P>
```

```
<EncryptedData Id="ed1" Type="http://www.w3.org/2001/04/xmlenc#Element"
  xmlns="http://www.w3.org/2001/04/xmlenc#">
  <EncryptionMethod Algorithm="http://www.w3.org/2001/04/xmlenc#tripleDES-cbc"/>
  <KeyInfo xmlns="http://www.w3.org/2000/09/xmldsig#">
    <EncryptedKey xmlns="http://www.w3.org/2001/04/xmlenc#">
      <EncryptionMethod Algorithm="http://www.w3.org/2001/04/xmlenc#rsa-1_5"/>
      <KeyInfo xmlns="http://www.w3.org/2000/09/xmldsig#">
        <KeyName>mario</KeyName>
      </KeyInfo>
      <CipherData>
        <CipherValue/>
      </CipherData>
    </EncryptedKey>
  </KeyInfo>
  <CipherData>
    <CipherValue>1XBCH70R3BRVGo8g7adBFfm7xqvOY0PKk3sF/yOIkKI/mQP5t
jB2AAo1ZMp6uPIF4YBLxvtq7Xg=</CipherValue>
  </CipherData>
</EncryptedData>
```


Binäres XML

Kompaktheit

- Erreichbar durch
 - Anderes Serialisierungsformat
 - „klassische“ Kompression
 - XML-spezifische Kompression

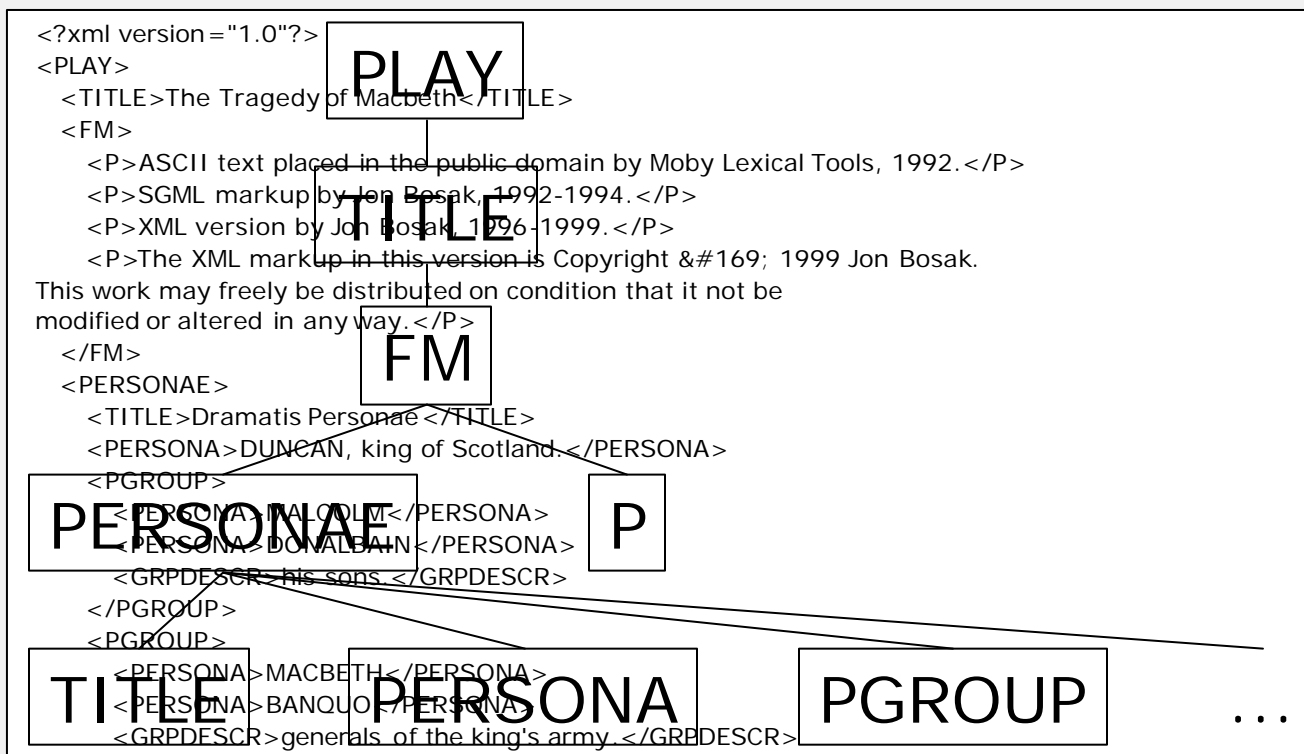
Binäres XML

Kompaktheit

- Andere Serialisierung
 - ASN.1/PER
- Mechanismen und Implementierungen zur Kompression beliebiger Inhalte existieren seit langem
 - Huffman
 - (G)ZIP
 - ...
- XML-spezifische Kompression relativ neu
 - XMill
 - XCQ
 - XMLPPM
 - MPEG7

Binäres XML

Alternativserialisierungen



- Strukturisomophe Kompression

Abstract Syntax Notation One (ASN.1):

```

play PLAY ::=
  title {
    fm {
      personae {
        ...
      }
    }
  }

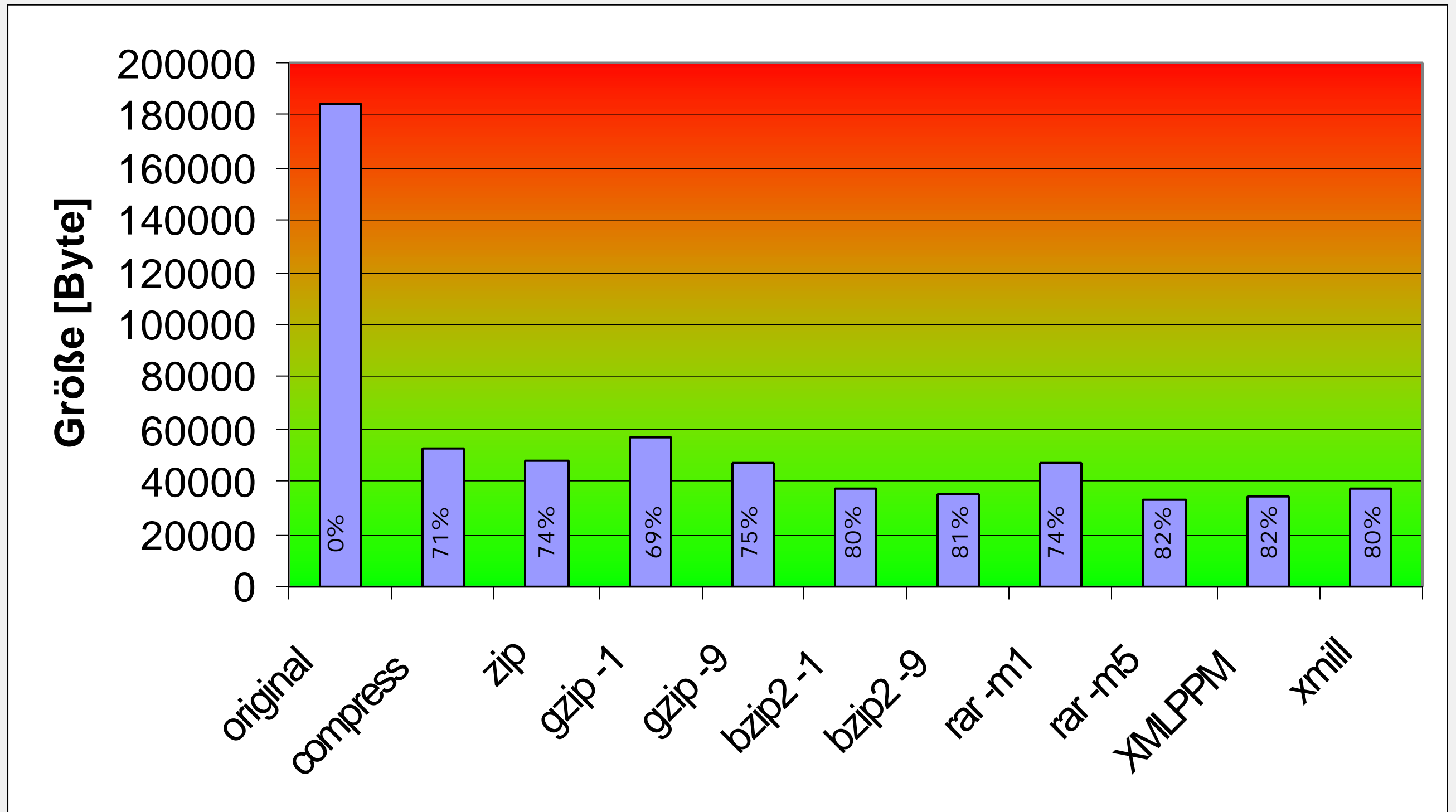
```

Packet Encoding Rules (PER):

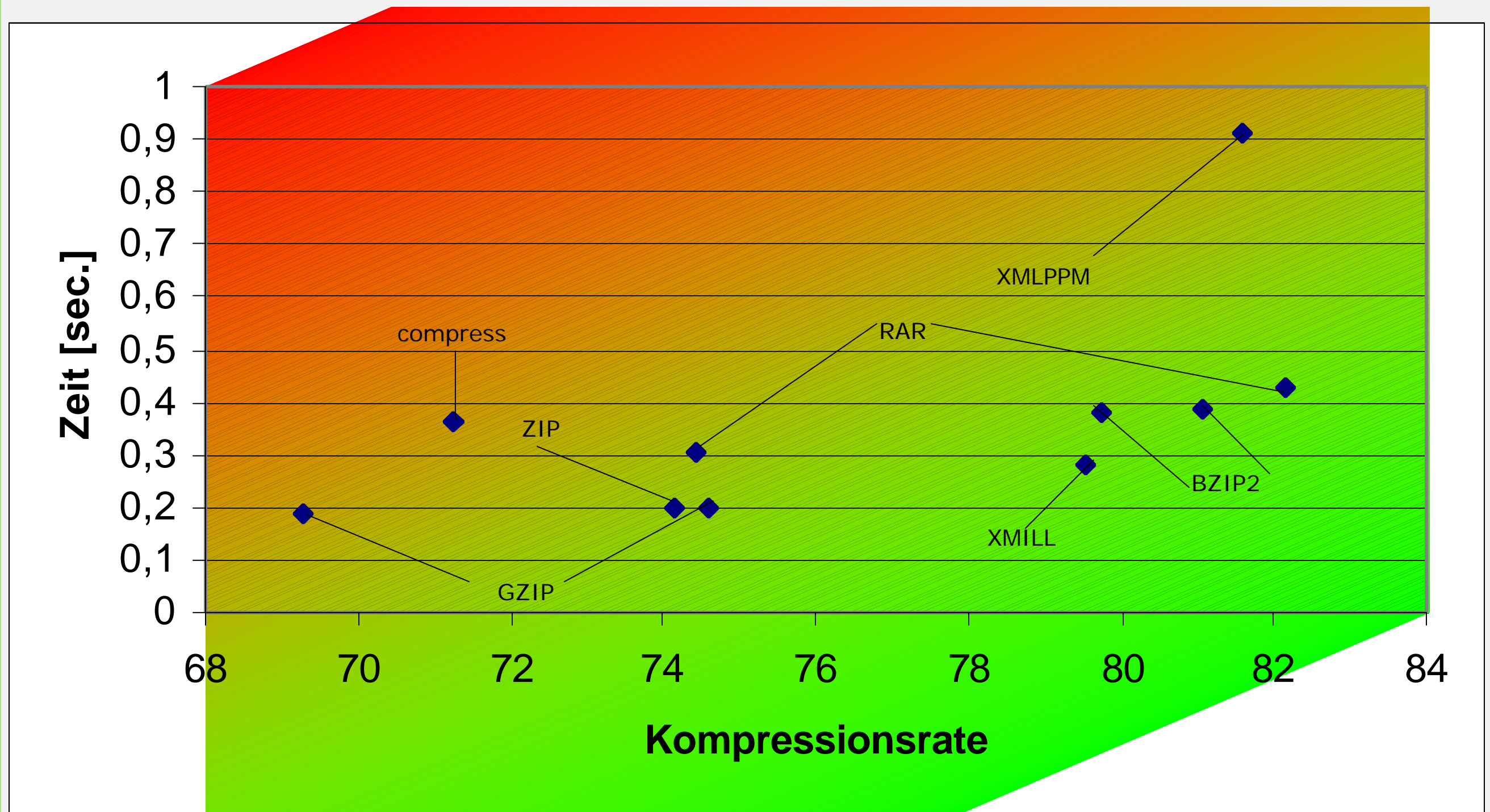
01011...

Binäres XML

Kompressionseffizienz

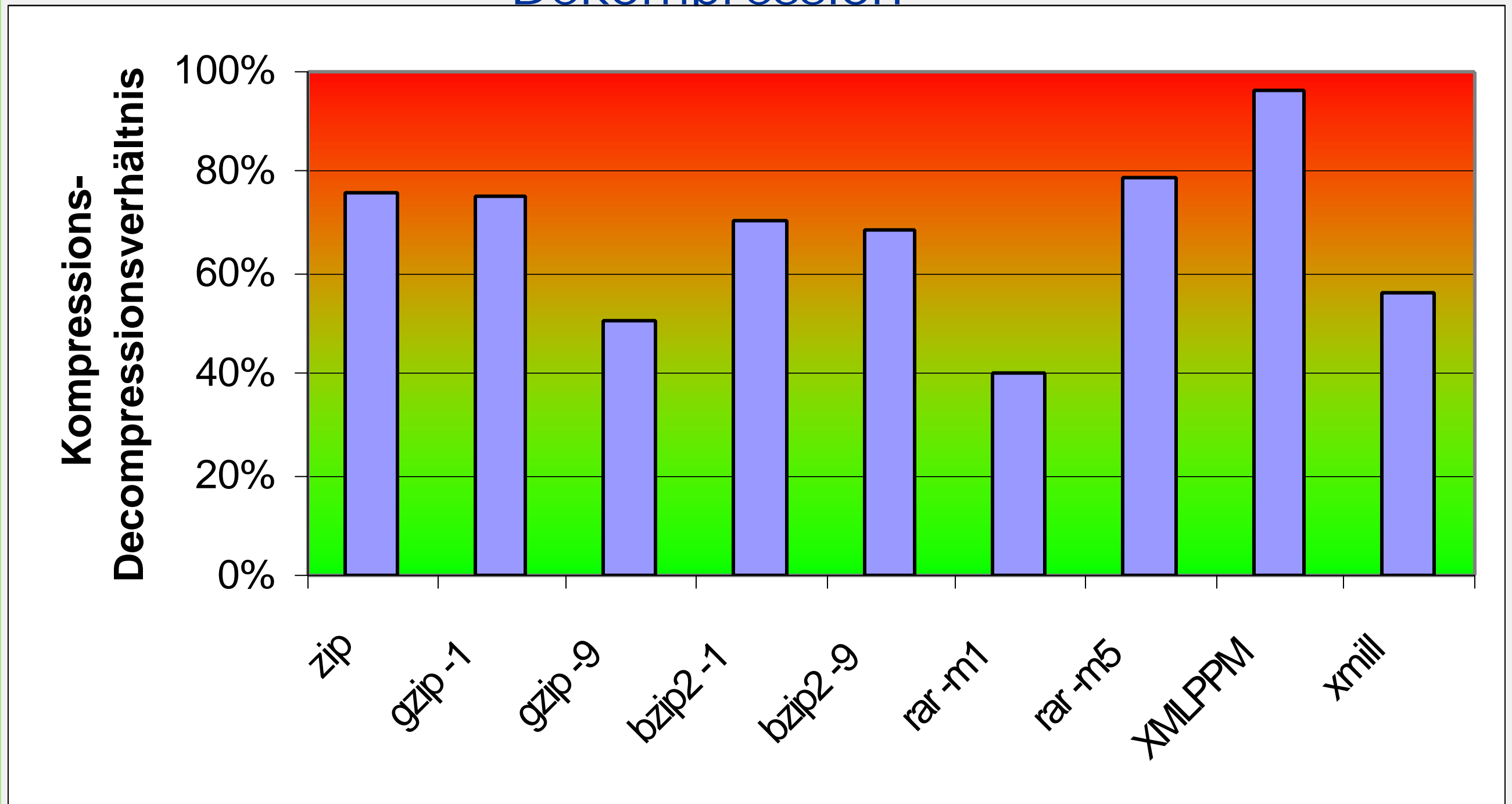


Binäres XML Kompression



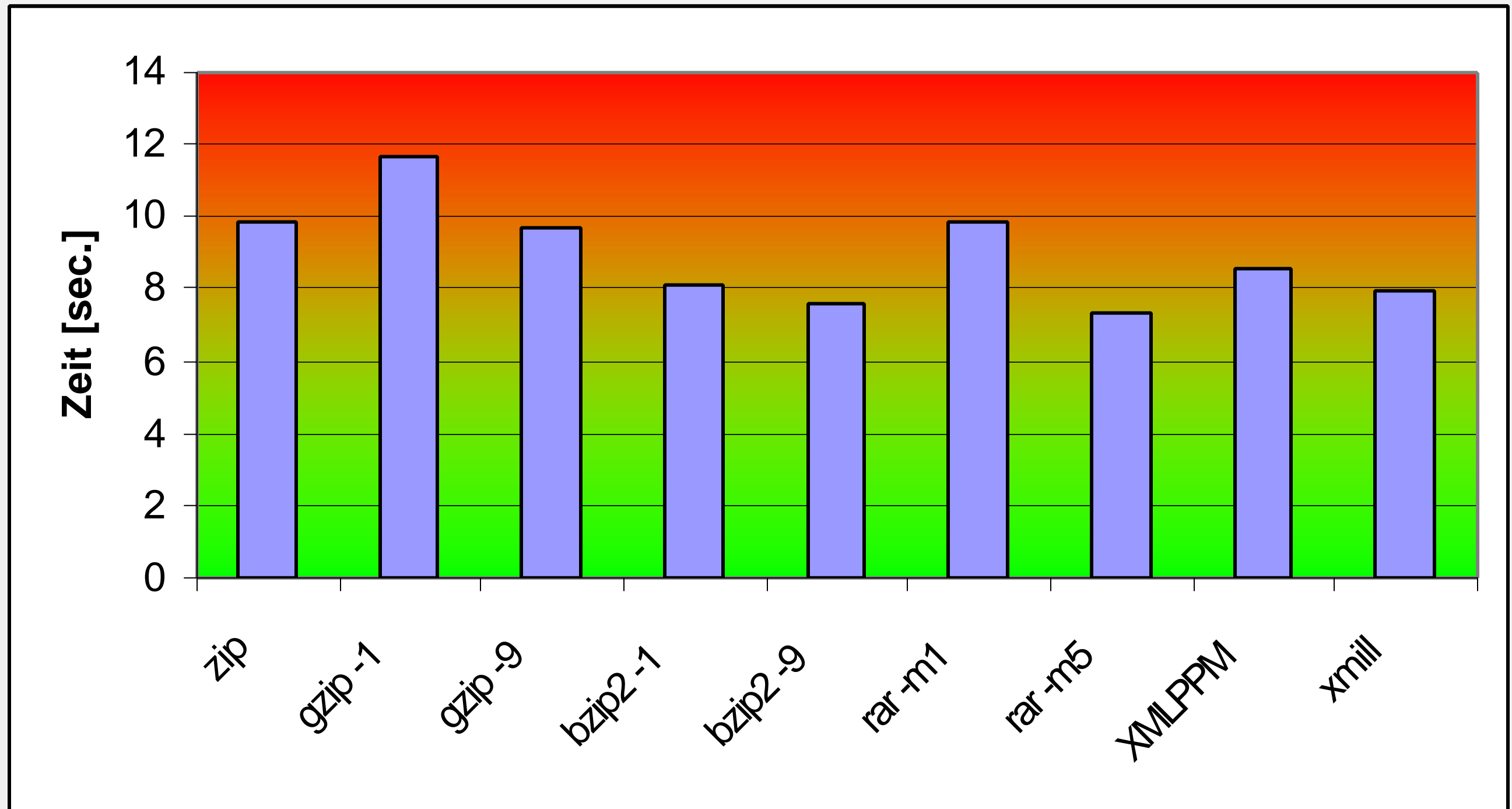
Binäres XML

Dekompression



Schnelle Dekompression insbesondere für Ressourcenbeschränkte Endgeräte notwendig

Binäres XML in der Praxis



Kompression, Übertragung und Dekompression
des Testdatenpaketes unter Nutzung eines 11Mbit-Netzes

Binäres XML

Schnelle Verarbeitung

- Zeichenkettenoperationen sind generell teuer
- Standard-XML-Schnittstellen (SAX, DOM) sind prinzipiell vergleichsweise inperformant
- Optimierte Varianten davon stimmen positiv (z.B. in XSLT-Prozessoren)

Binäres XML

Schnelle Verarbeitung

- Binäres XML erzwingt Entscheidungen
 - Byte Ordering
 - Binärrepräsentation der lexikalischen Datentypen
 - Evtl. Ansatz mit generischem Code (ähnlich Java Bytecode/MSIL) sinnvoll

Fazit: Komprimiertes XML

- Ausschließlich verlustfrei komprimierende Verfahren interessant
- Lauflängencodierung nicht nutzbar
- Auswirkungen auf Applikationen: XSLT, DOM, SAX, Browser
- XML-inhärente Trennung von Inhalt und Struktur muß auch bei Kompression vollzogen werden.
- Komprimierte Strukturen müssen isomorph zum InfoSet sein.
- Falls Strukturesemantik (InfoSet) verloren
 - => XPath, XSLT, XQuery etc. nicht mehr nutzbar

Binäres XML: Workshop des W3C

Fragen an die Web-Gemeinde:

- Erfahrungen und Implementierungen (Messungen)
- Ziele?
 - Bandbreitennutzungsminimierung
 - Reduktion der Parsing-Time
 - Einfachere APIs
 - ...
- Typische
 - Dokumente
 - Applikationen
- Internationalisierung und Barrierefreiheit
- Unterschied zu GZIP
- Abhängigkeit zu bestimmten Schemasprachen
- Interoperabilität mit anderen XML-Standards

Links

- <http://developer.java.sun.com/developer/technicalArticles/WebServices/fastWS/index.html>
- <http://www.w3.org/2003/07/binary-xml-cfp.html>
- <http://www.xml.com/pub/a/2003/08/13/deviant.html>
- <http://www.w3.org/TR/2003/WD-soap12-mtom-20030721/>
- <http://www.cafeconleche.org/books/effectivexml/chapters/50.html>
- <http://xml.coverpages.org/xmlAndCompression.html>
- <http://www-106.ibm.com/developerworks/xml/library/x-matters13.html>
- <http://www-106.ibm.com/developerworks/xml/library/x-matters19/>
- <http://www.ictcompress.com/downloadxml.html>

XML-Kompressionswerkzeuge:

- *Millau*: <http://www9.org/w9cdrom/154/154.html>
- *XCQ*:
http://www2003.org/cdrom/papers/poster/p059/WWW2003_Poster_59.html
- *XMLPPM*: <http://sourceforge.net/projects/xmlppm>
- *XMill*: <http://xmill.sourceforge.net>

jeckle.de - Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop <http://www.jeckle.de/> Search Print

[Unified Modeling Language \(UML\)](#)
[eXtensible Markup Language \(XML\)](#)
[XML Metadata Interchange \(XMI\)](#)
[Web Services](#)
[XML Acronym Demystifier Project](#)
[XML-Strategie](#)

[Vorträge und Publikationen](#)
[Vorlesungen](#)
[Diplomarbeiten](#)
[GOOAL.net](#)
[XML-Arbeitskreis](#)
[Software & Downloads](#)

[Web Services Workshop WS-RSD'02](#)
[ICWS'03-Europe](#)
[Web Services @ Berliner XML-Tage](#)
[Internet > Search Engines](#)
[Mersennesche Primzahlen](#)
[Feedback](#)
[Rotkreuz Mitgliederverwaltung](#)


[Mario Jeckle ...](#)
[Dialog ...](#)
[Über diese Seiten ...](#)
[suchen ...](#)
[SiteMap](#)
[RSS Newsfeed XML](#)
[Was gibt's hier Neues?](#)

jeckle.de

Transferring data from www.jeckle.de...