

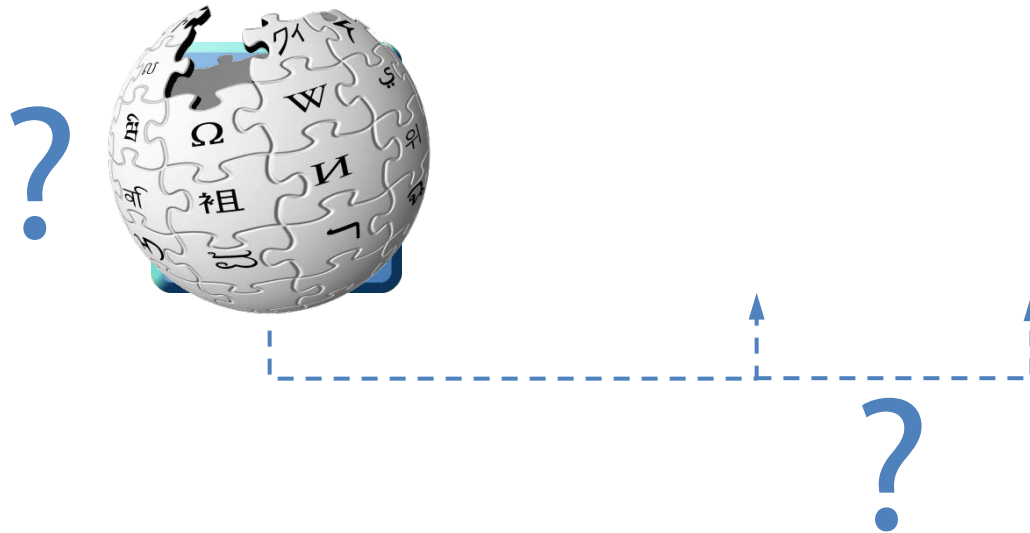
Die Entwicklung von Qualitätsmängeln in der Wikipedia anhand von Wartungsbausteinen

Matthias Busse

Bachelorverteidigung
Studiengang Mediensysteme
Bauhaus-Universität Weimar

27.04.2012

Motivation



[Blumenstock, WWW 2008]

[Dalip et al., JCDL 2009]

[Hu et al., CIKM 2007]

[Lipka und Stein, WWW 2010]

1. Warum verletzt ein Artikel die Qualitätsmaßstäbe?
2. Wie kann ein minderwertiger Artikel an Qualität gewinnen?

Entwicklung der Qualitätsmängel

Identifikation



Extraktion



Analyse

Identifikation

The screenshot shows the Wikipedia article for Charlie Chaplin. A red box highlights a notice at the top of the article body: "This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (February 2012)". The article text begins with "Sir Charles Spencer 'Charlie' Chaplin, KBE (16 April 1889 – 25 December 1977) was an English comic actor, film director and composer best known for his work during the silent film era." To the right of the text is a portrait of Chaplin as "The Tramp" from 1915, with a biographical sidebar below it listing his birth name as Charles Spencer Chaplin and his birth date as 16 April 1889 in Walworth, London, England, United Kingdom.

- ▶ Ansatz: Jeder Wartungsbaustein signalisiert einen Qualitätsmangel
- ▶ > 380 000 Vorlagen verfügbar → automatisierte Erkennung notwendig
- ▶ 444 Wartungsbausteine identifiziert

Identifikation



This article **has no lead section**. Please help by adding an introductory section to this article. For more information, see the [layout guide](#), and Wikipedia's [lead section guidelines](#).



The **neutrality of this article is disputed**. Please see the discussion on the [talk page](#). Please do not remove this message until the [dispute is resolved](#).



This article **does not cite any references or sources**. Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and [removed](#).



It has been suggested that this page **be split into multiple pages accessible from a disambiguation page**.



This article **has not been added to any categories**. Please help out by [adding categories](#) to it so that it can be listed with similar articles. *(April 2012)*

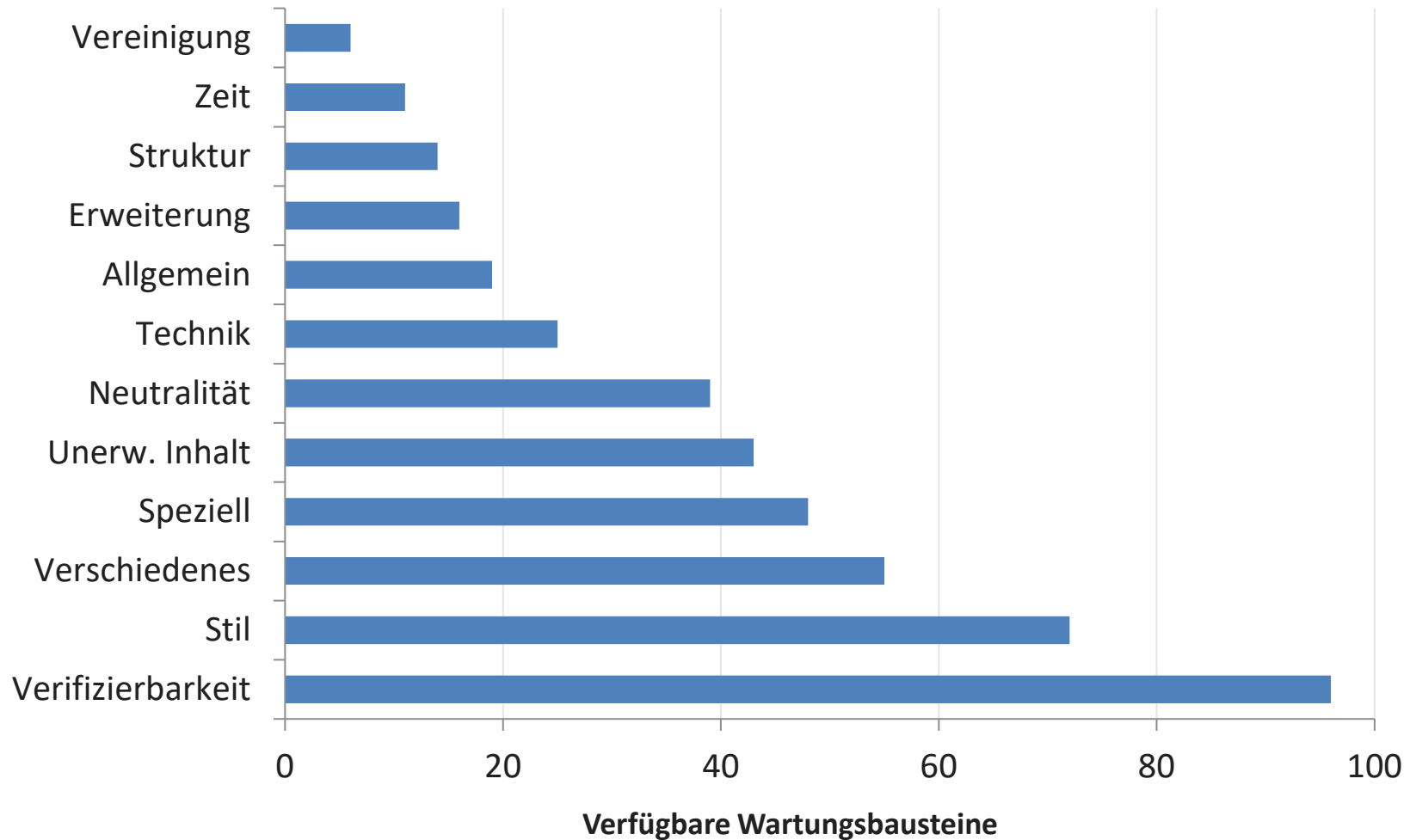


This article's [references may not meet Wikipedia's guidelines for reliable sources](#). Please help by checking whether the references meet the criteria for reliable sources.



This article **needs additional citations for verification**. Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and [removed](#). *(February 2012)*

Identifikation



Identifikation



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Interaction
- Toolbox
- Print/export
- Languages

Article [Talk](#)

[Read](#) [View source](#) [View history](#) [☆](#)

Charlie Chaplin

From Wikipedia, the free encyclopedia



This article **needs additional citations for verification**. Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and [removed](#). (February 2012)

Sir Charles Spencer "Charlie" Chaplin, KBE (16 April 1889 – 25 December 1977) was an English [comic actor](#), film director and composer best known for his work during the [silent film](#) era.^[2] He became the most famous film star in the world before the end of World War I. Chaplin used [mime](#), [slapstick](#) and other [visual comedy](#) routines, and continued well into the era of the [talkies](#), though his films decreased in frequency from the end of the 1920s. His most famous role was that of *The Tramp*, which he first played in the [Keystone](#) comedy *Kid Auto Races at Venice* in 1914.^[3] From the April 1914 [one-reeler](#) *Twenty Minutes of Love* onwards he was writing and directing most of his films, by 1916 he was also producing them, and from 1918 he was even composing the [music](#) for them. With [Mary Pickford](#), [Douglas Fairbanks](#) and [D. W. Griffith](#), he co-founded [United Artists](#) in 1918.^[citation needed]

Chaplin was one of the most creative and influential personalities of the silent-film era. He was influenced by his predecessor, the French silent film comedian [Max Linder](#), to whom he dedicated one of his films.^[5] His working life in entertainment spanned over 75 years, from the [Victorian](#) stage and the [music hall](#) in the United Kingdom as a child performer, until close to his death at the age of 88. His high-profile public and private life encompassed both adulation and controversy. Chaplin was identified with [Left-wing politics](#) during the [McCarthy era](#) and he was ultimately forced to resettle in Europe from 1952.

In 1999, the [American Film Institute](#) ranked Chaplin the 10th [greatest male screen legend of all time](#).^[6] In 2008, Martin Sieff, in a review of the book *Chaplin: A Life*, wrote: "Chaplin was not just 'big', he was gigantic. In 1915, he burst onto a war-torn world bringing it the gift of comedy, laughter and relief while it was tearing itself apart through World War I. Over the next 25 years, through the [Great Depression](#) and the rise of [Adolf Hitler](#), he stayed on the job. ... It is doubtful any individual has ever given more entertainment, pleasure and relief to so many human beings".^[7]

Sir Charlie Chaplin

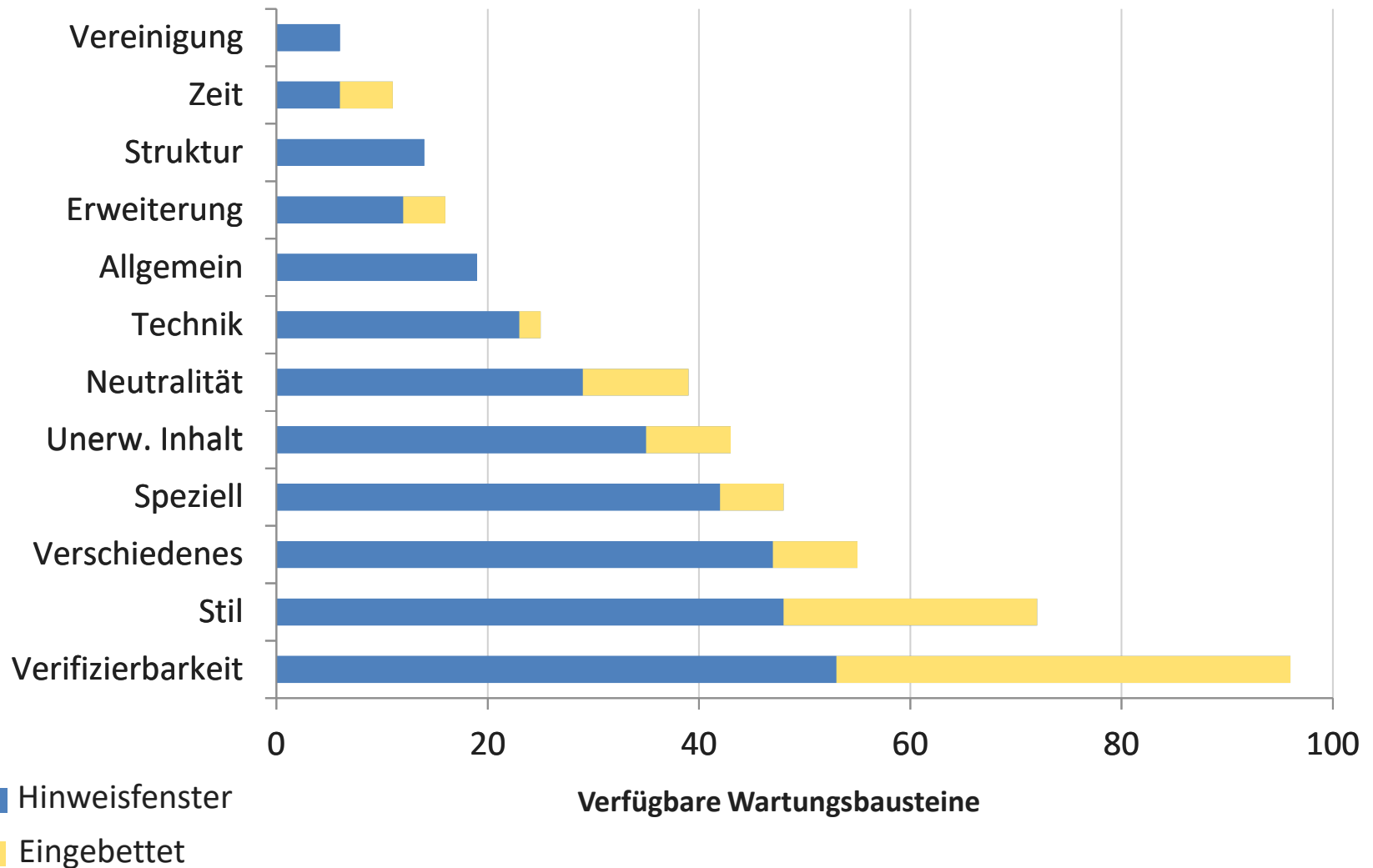


Chaplin as *The Tramp*, 1915

Birth name	Charles Spencer Chaplin
Born	16 April 1889 Walworth , London, England, United Kingdom

[citation needed]

Identifikation



Entwicklung der Qualitätsmängel

Identifikation



Extraktion



Analyse

Extraktion



- ▶ 277 Mio. Revisionen
- ▶ 6,3 TB

Vandalismus

126 Mio. Revisionen

Entwicklung der Qualitätsmängel

Identifikation



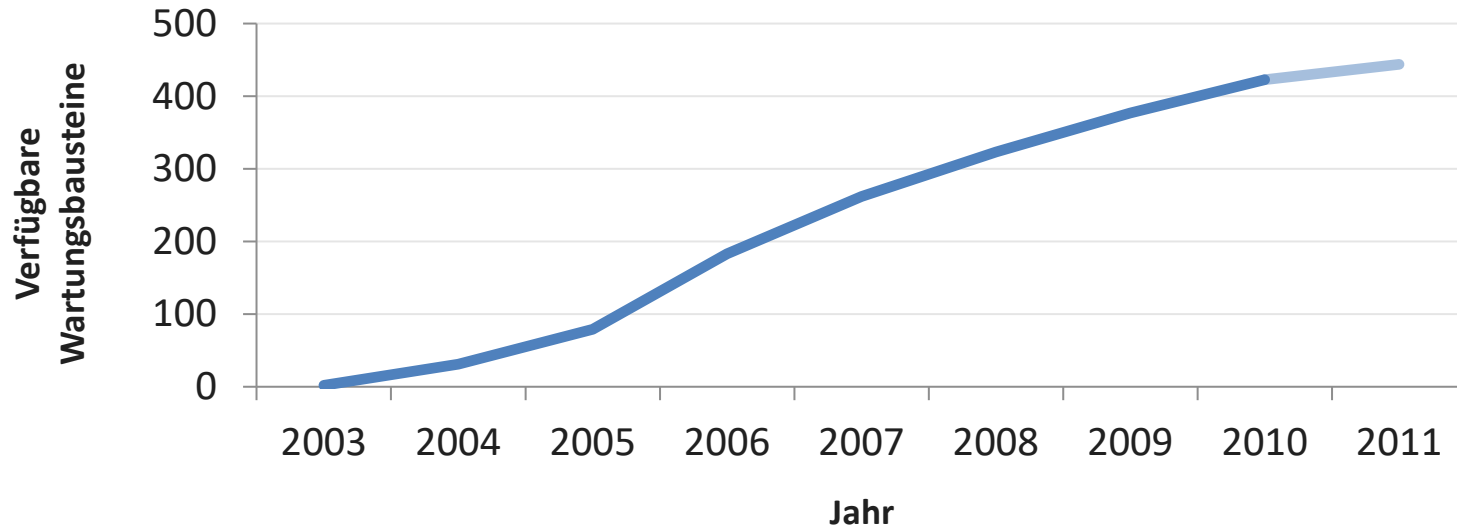
Extraktion



Analyse

Wie haben sich die verfügbaren
Wartungsbausteine entwickelt?

Verfügbare Wartungsbausteine



- ▶ Regression neu hinzukommender Wartungsbausteine

Mängelklasse

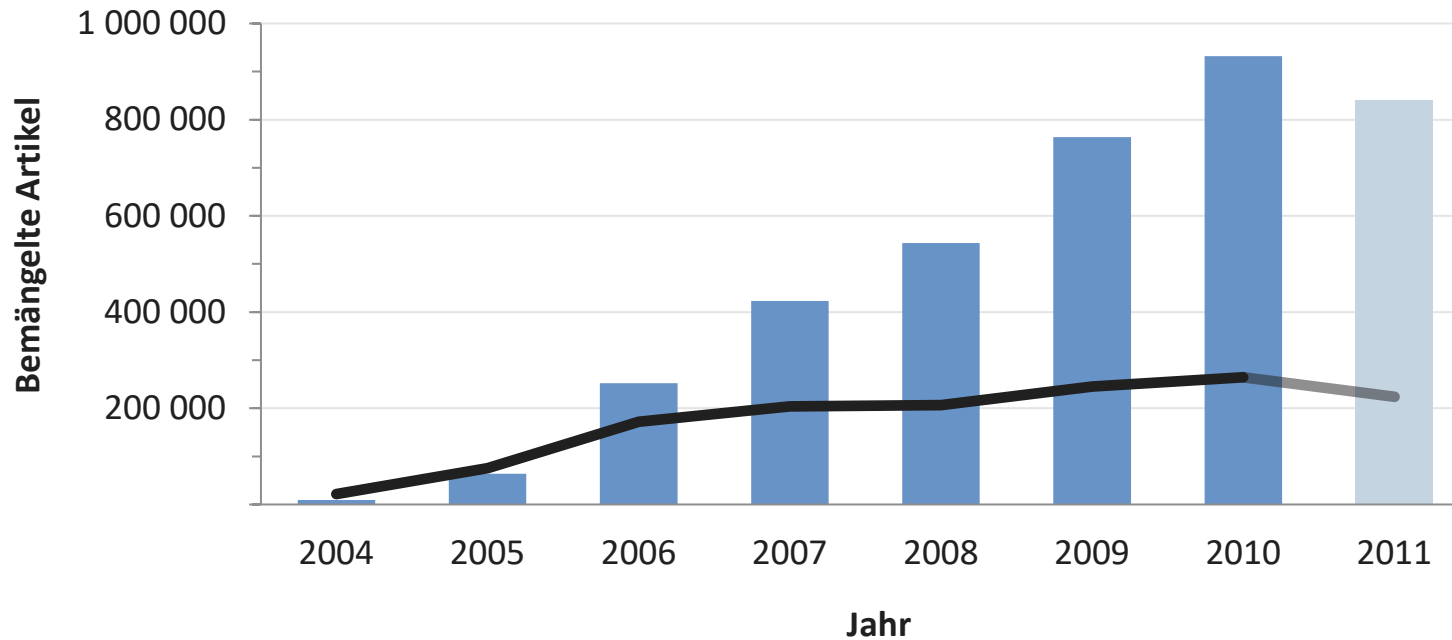
- ▶ Deutliche Zunahme nur von Verifizierbarkeit & Stil

Geltungsbereich

- ▶ Immer mehr eingebettete Wartungsbausteine vorhanden

Wie viele Artikel werden bemängelt?

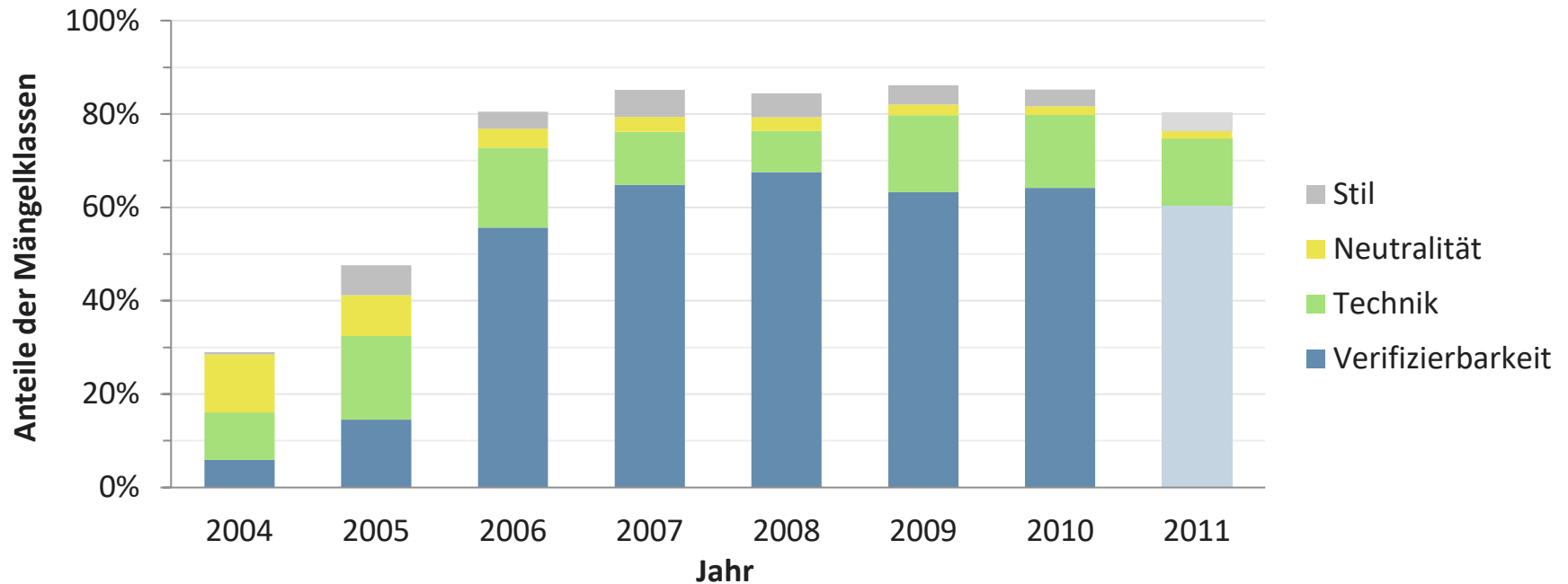
Bemängelte Artikel



- ▶ Seit 2007 jeder fünfte Artikel bemängelt
- ▶ Über 95 % der gesetzten Mängel mit Wartungsbausteinen markiert, die vor 2007 erstellt wurden

Was wird bemängelt?

Bemängelte Bereiche



Mängelklasse

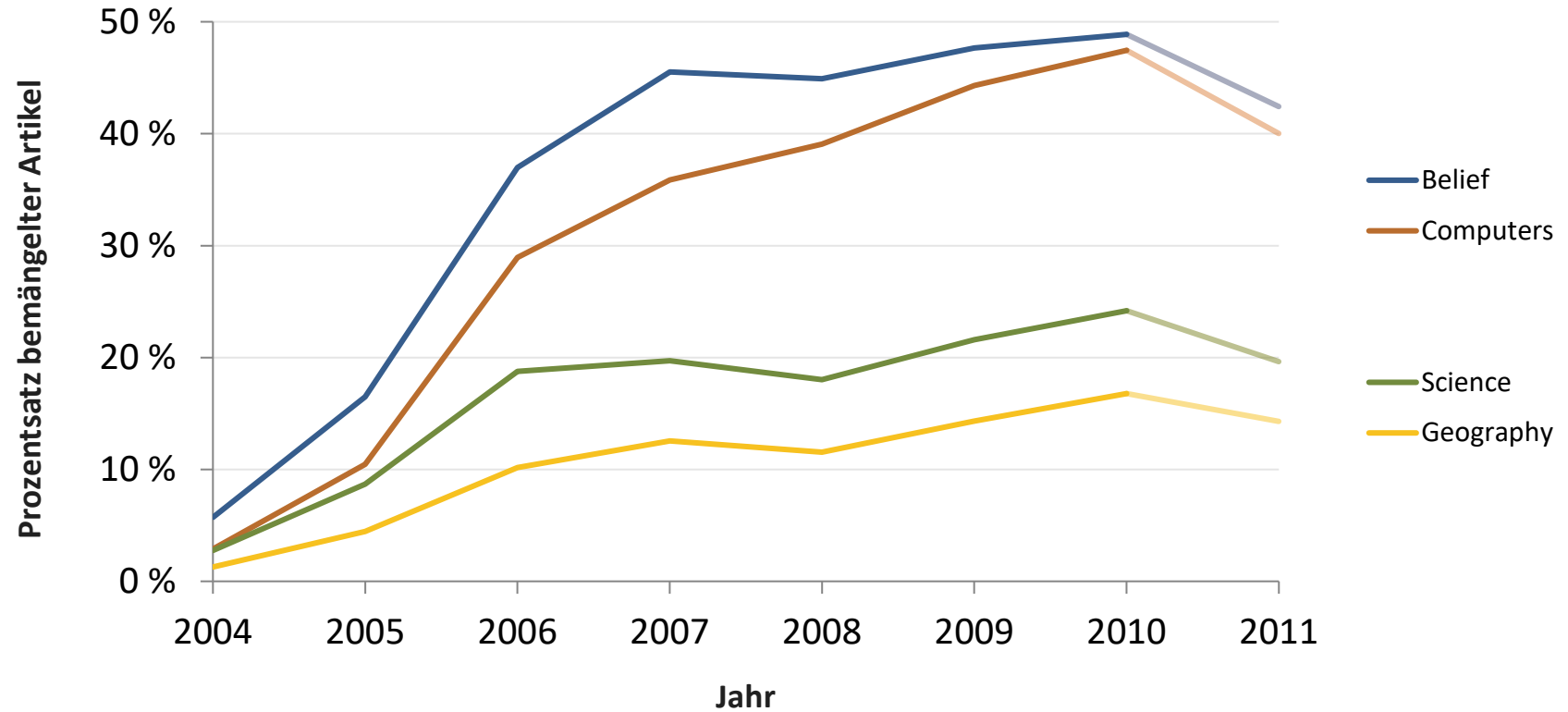
- ▶ Mehrheit der markierten Mängel betreffen Verifizierbarkeit
- ▶ Technische und stilistische Mängel häufig anzutreffen
- ▶ Hohe Aktualität der Artikel

Geltungsbereich

- ▶ Ausgeglichenes Verhältnis von eingebetteten Wartungsbausteinen zu Hinweisfenstern

Welche Themen werden bemängelt?

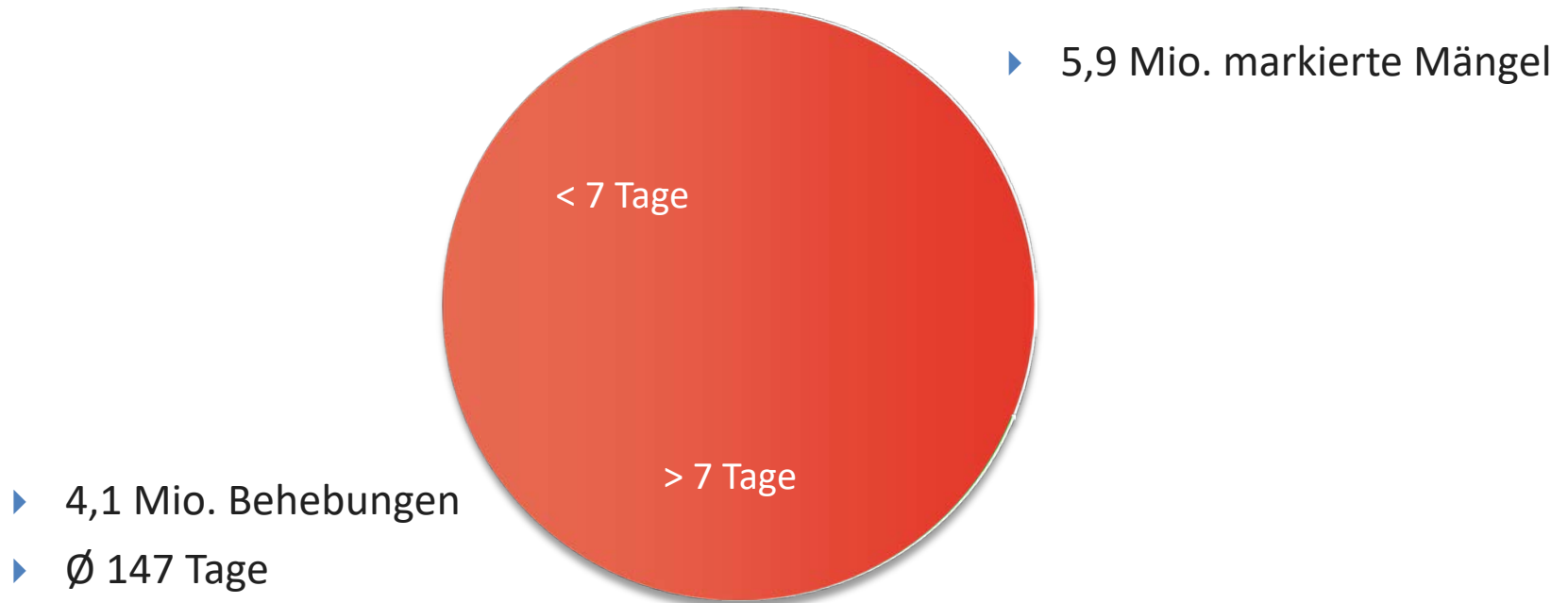
Bemängelte Themen



- ▶ Kontroverse Themen neigen eher zu Beanstandungen

Wie viele Mängel konnten ausgebessert werden
und wie lange dauert bis zur Behebung?

Behebungsdauer



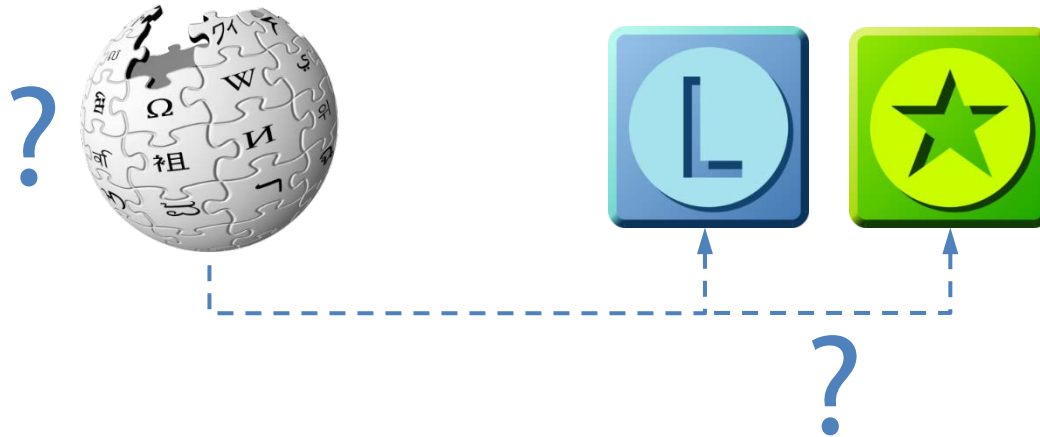
Mängelklasse

- ▶ Oberflächliche Klassen wie Erweiterung (220 Tage) deutlich aufwendiger als technische (97 Tage)

Geltungsbereich

- ▶ Eingebettete Mängel schneller behoben als Hinweisfenster (122 Tage < 165 Tage)

Zusammenfassung



1. Warum verletzt ein Artikel die Qualitätsmaßstäbe?
 2. Wie kann ein minderwertiger Artikel an Qualität gewinnen?
 - ▶ Große Mehrheit der Mängel betrifft Verifizierbarkeit
 - ▶ Hoher Anteil technischer und stilistischer Mängel
-
- ▶ Qualitätsmängel durch Wartungsbausteine spezifizierbar
 - ▶ Nach Möglichkeit eingebettete Wartungsbausteine nutzen
 - ▶ Zu komplexe und allgemeine Hinweisfenster in eingebettete Wartungsbausteine umdefinieren

Ausblick

- ▶ Verbesserte Vandalismuserkennung
- ▶ Automatische Erkennung der Qualitätsmängel
- ▶ Untersuchung der Nutzer, die Wartungsbausteine setzen und beheben

Vielen Dank!

XML-Format

```
<mediawiki xml:lang="en">
  <page>
    <title>Charlie Chaplin</title>
    <ns>0</ns>
    <id>22194</id>
    <revision>
      <id>485344292</id>
      <timestamp>2011-09-01T14:53:11Z</timestamp>
      <contributor>
        <username>JohnDoe</username>
        <id>14383484</id>
      </contributor>
      <comment>Reverting possible vandalism by 12.34.56.78</comment>
      <text bytes="74953">{{Unreferenced|date=October 2011}} Sir
        Charles Spencer "Charlie" Chaplin, KBE (16 April 1889 – 25
        December 1977) was an English comic actor, film director and ...
      </text>
    </revision>
    <revision>
      ...
    </revision>
  </page>
</mediawiki>
```


Behebungsauern nach Mängelklasse

Mängelklasse	Behebungsdauer	%
Erweiterung	220,7	49,2
Allgemein	203,3	71,5
Speziell	175,2	64,8
Stil	164,4	78,1
Verifizierbarkeit	155,8	67,3
Zeit	154,7	65,0
Struktur	137,2	75,5
Unerw. Inhalt	122,1	73,5
Vereinigung	111,9	87,4
Neutralität	107,2	86,4
Technik	97,5	74,6
Verschiedenes	96,3	92,6
Σ 146,8		69,7

Markierungen nach Mängelklassen

