

# LanguageTool WikiCheck

Stil- und Grammatikfehler automatisch finden

Workshop



Daniel Naber  
WikiCon 2014



# Roadmap

- Einführung: LanguageTool WikiCheck
  - Ausprobieren und selber Fehler beheben
- Einführung: Fehlerregeln
  - Ausprobieren und selber Fehlerregeln schreiben
- Weitere Arten der Nutzung
- Diskussion



# Umfrage

- Wer kennt LanguageTool bereits?
- Erwartungen, Wünsche?
- Wer ist Softwareentwickler?



# Über

- Daniel Naber
  - Benutzer:dnaber
  - Softwareentwickler aus Potsdam
  - Entwickler von LanguageTool und [openthesaurus.de](http://openthesaurus.de)



# Teil 1: LanguageTool WikiCheck



# Finde die Fehler

"Vielleicht zu Ehren der Gäste aus Deutschland wurden am Montag, dem 24. Februar 1914 im Michaelteater zwei Vorstellung in Deutscher Sprache gegeben."

- 
- Ausprobieren auf [language-tool.org](https://www.language-tool.org)
  - Fehler, die eine einfache Rechtschreibprüfung nicht finden kann.



# LanguageTool

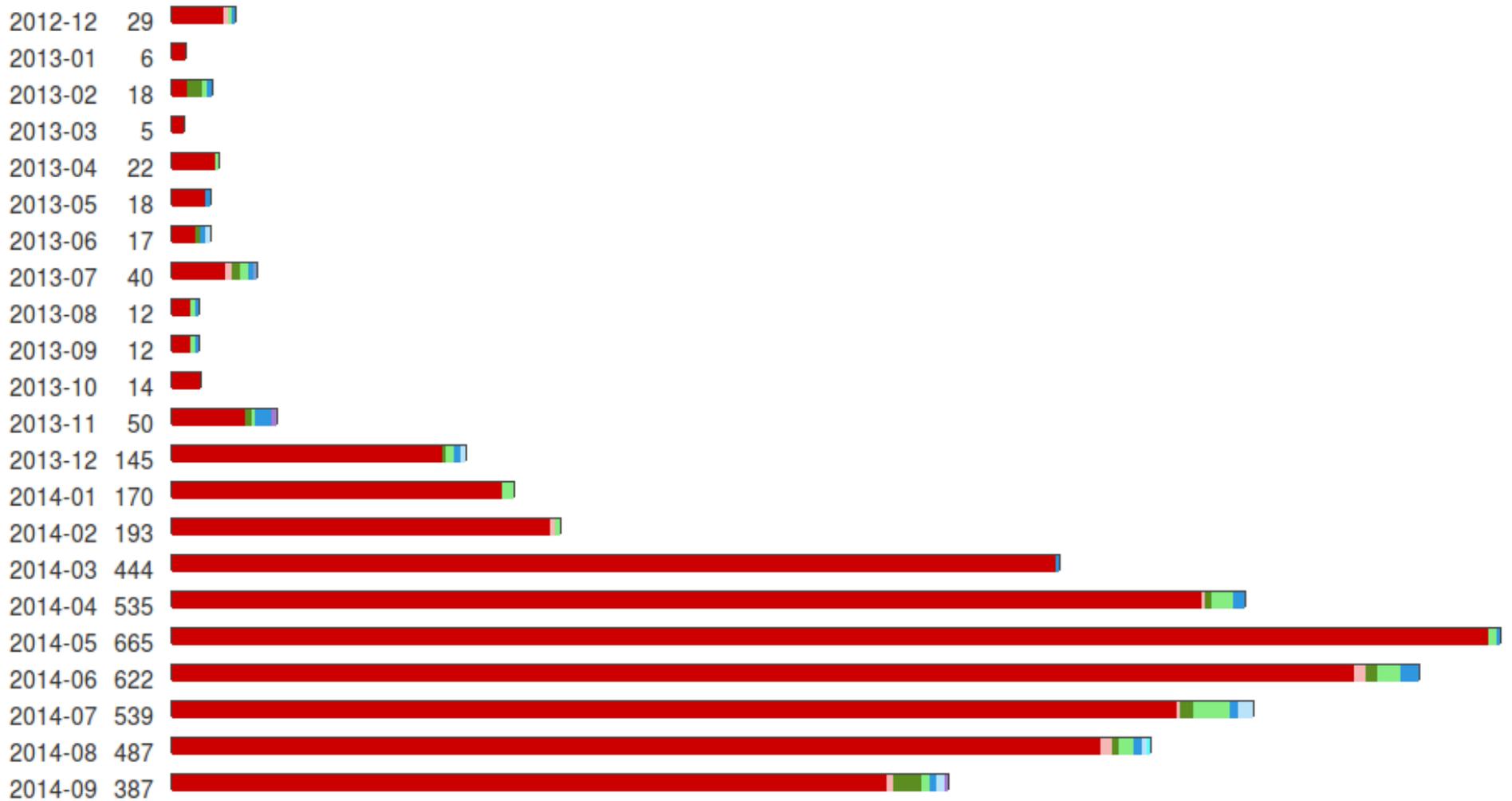
- Stil- und Grammatikprüfung
- Open Source (LGPL)
- existiert seit 2003
- 30 Sprachen (Deutsch, Englisch, Niederländisch, Spanisch, ...)
  - <https://languagetool.org/languages/>



# LanguageTool WikiCheck

- Wikipedia-Integration von LanguageTool
- seit Juli 2014 auf Tool Labs
- bisher mindestens 4.000 Fehler damit behoben

# Meine Statistik (XTools)





# Fehler-Beispiele

- Ein-und Zweifamilienhäuser
- (...) 5-türiger Kompaktcombi. Vorerst nur ohne Klimaanlage lieferbar (...)
- spiegelt ... wieder
- Elmshausen weißt eine geschlossene Bebauung auf.

# Funktionsweise FeedChecker

- <http://tools.wmflabs.org/language-tool/feedMatches/list?lang=de>
- Gründe für Fehlalarme
  - Wikitext
  - Regel nicht perfekt
- Welche Fehler werden gefunden?
- Häufige vs. interessante Fehler (z.B. Datumsfehler)



# Und jetzt...

- Selber ausprobieren:
  - <https://de.wikipedia.org/wiki/Benutzer:Dnaber>
- ca. 20 Minuten



# Teil 2: Selber Regeln schreiben



# Funktionsweise Fehlersuche

- "ökonomischer Gottesdienst"
  - Rechtschreibprüfung reicht hier nicht
- Idee: Muster, also eine Folge von Wörtern/Wortarten
- Zu beachten:
  - es gibt Varianten: "ökonomischer Gottesdienst", "ökonomischen Gottesdienstes", ...
  - möglichst keine Fehlalarme



# Funktionsweise Fehlersuche

- Demo Liste der Regeln
- Demo Regel-Editor

# Praktischer Teil (ca. 30 Minuten)

- Letzte Änderungen (auch eigene) durchschauen und nach potenziellen Mustern suchen
  - (Interessant ist der Fehler, die Korrektur ist nur zweitrangig)
- Vorhandene Regeln anschauen
- Regel-Editor benutzen
- <https://de.wikipedia.org/wiki/Benutzer:Dnaber>



# Teil 3: Sonstige Nutzung & Diskussion



# Downloads

- als Desktop-Programm, LibreOffice/OpenOffice-Add-on, API (Java, HTTP)
- Wikipedia-Version: indexieren, suchen



# Diskussion

- Wie gehen wir mit den gefundenen Fehlern um (z.B. 250 Fehler "Datum passt nicht zu Wochentag"!)?
- Wie können wir mehr inhaltliche Fehler automatisch finden?
- Was tun gegen die vielen Trivialfehler wie vergessenes Leerzeichen nach Satzende?



# Fazit

- Regel-Ideen gewünscht
- RSS-Feed zu Fehlern in 'Letzte Änderungen' abonnieren
- Eintrag auf Announcement-Mailing-Liste (unten auf [languagetool.org](http://languagetool.org))