

data2check

DTP meets XML

Donnerstag, 24. November 2016

Zur Person:

Manuel Montero Pineda

- > Dipl.-Wirtschaftsinformatiker (FH), M.A.
- > XML-Entwickler und -Berater in den Bereichen XML Schema, OOXML, XSLT, XSL-FO u.v.m.
- > Veröffentlichungen u.a. „Professionelle XML-Verarbeitung mit Word“, „XSL-FO in der Praxis“, die „XSL-FO Referenz“ und „Schematron“; erschienen beim dpunkt-Verlag
- > Geschäftsführer der data2type GmbH

Was kann ich mit data2check machen?

Wozu?

Prüfung von Word-, InDesign- und EPUB-Dokumenten.
Später noch geplant: PDF.

Was kann ich prüfen?

Word und InDesign lassen sich auf ihren Aufbau und auf Regeleinhaltung prüfen.

Was bekomme ich zurück?

Word/InDesign mit Korrekturen/Fehlermeldungen, XML, Prüfberichte

Was kann ich mit data2check machen?

Wieviel kostet das alles?

Nichts. Der Dienst auf unseren Servern ist kostenlos.

Wer konfiguriert?

Konfiguration ist für jeden selbst möglich über eine Oberfläche. Programmierkenntnisse sind nicht notwendig.

Wo läuft die Prüfung ab?

- > Auf unseren Servern
- > Zugänglich über eine Homepage (<http://www.data2check.de/>)

Was sind denkbare Szenarien?

1. Verlag macht QS für seine InDesign-Dokumente

Die meisten Verlage haben Vorgaben für InDesign, die aber nicht geprüft werden. Es gibt lediglich eine Sichtprüfung auf die Print-PDFs.

Mit data2check ist es möglich, dass alle Satzbetriebe einheitlich arbeiten. Möglich wird dadurch z. B. der einheitliche Export nach EPUB oder die Sicherung in XML für die mögliche zukünftige Verwendung.

Was sind denkbare Szenarien?

- 2. Verlage oder Satzbetriebe nutzen den Dienst für die Copyediting-Unterstützung.**
- 3. Verlag lässt Autoren nach seiner Struktur in Word erfassen und Autoren prüfen jederzeit ihre Dokumente selbst.**
- 4. Self-Publisher lassen die Autoren in Word erfassen und wandeln die Dokumente über XML in PDF/EPUB um.**
- 5. In der Technischen Dokumentation werden viele Dokumentationen in Word erfasst. Diese lassen sich so qualitätssichern. Eine spätere mögliche Migration in ein Redaktionssystem wird so erleichtert.**

Was sind denkbare Szenarien?

Wir hoffen auf Input von Ihnen!!!

<http://www.data2check.de/>

Regelprüfung – Strukturprüfung

Regeln sind:

Strukturprüfung

- > Prüfung auf unbekannte Formatvorlagen.
- > Falsche Verwendung der Formatvorlagen, z. B. Beginn des Dokuments mit Überschriften-Ebene 2 anstatt Ebene 1.
- > Prüfung auf die richtige Verwendung von Zeichen-Formatvorlagen, z.B. kann es verboten sein, eine Zeichen-Formatvorlage in einer Überschrift zu verwenden.
- > u.v.m.

Regelprüfung – Strukturprüfung

Konfigurations-Info	Regeln für den Dokumentaufbau	Suche-/Ersetze-Regeln	Komponenten
Erster Absatz des Dokuments	muss aus dieser Liste sein: <input type="text"/>		
	1_KapitelUeberschrift <input type="text"/>		
	1_Ueberschrift2 <input type="text"/>		<input type="button" value="Hinzufügen"/>
Kommentar zu Beschränkungen des ersten Absatzes	<input type="text" value="Kommentar"/>		
Absatz-Formatvorlage 1_KapitelUeberschrift			
Funktion	Überschrift Ebene 1 <input type="text"/>		

Regelprüfung – Word-Komponenten

Word-Komponenten können wahlweise geprüft werden

- > Verwendung von Bildern einschränken.
- > Textboxen verbieten.
- > Kommentare verbieten.
- > WordArts unterbinden.
- > Diagramme verbieten.
- > ...

Regelprüfung – Word-Komponenten

Erlaubte
Komponenten:

Bilder

Bilder sind nicht erlaubt.

Bearbeitung

- Änderungsnachverfolgung ist erlaubt.
- Kommentare im Dokument sind nicht erlaubt.

Objekte

- Textboxen/Textfelder sind nicht erlaubt. Wenn das Dokument Textboxen enthält, so werden diese als fehlerhaft markiert.
- Formen (Einfügen -> Illustrationen -> Formen) sind nicht erlaubt. Wenn das Dokument Formen enthält, so werden diese als fehlerhaft markiert.
- Diagramme (Einfügen -> Illustrationen -> Diagramme) sind nicht erlaubt. Wenn das Dokument Diagramme enthält, so werden diese als fehlerhaft markiert.

Regelprüfung – Suche-/Ersetze-Regeln

Suche-/Ersetze-Regeln

- > Komplette Regex-Unterstützung.
- > Wahlweise eingeschaltete Änderungsverfolgung.
- > Mit Kommentaren.
- > Austausch ohne Rückmeldung ist möglich.
- > Beliebig viele Regeln können unterstützt werden.
- > ...

Regelprüfung – Suche-/Ersetze-Regeln

Konfigurations-Info Regeln für den Dokumentaufbau Suche-/Ersetze-Regeln Komponenten XML-Transformation

Suche nach
datatotype

Regulärer Ausdruck
 Ersetzung durchführen

Ersetze durch
data2type

als Änderung markieren
 mit Kommentar

Kommentar bei Regelanwendung
Kommentar...

Neue Regel hinzufügen

XML-Umwandlung

XML-Transformation

- > Neben einer generischen XML-Datei kann eine Transformation konfiguriert werden.
- > Diese Erlaubt das Umwandeln, Entpacken, etc. von Word-Formatvorlagen bzw. InDesign-Formate in XML-Tags.

Regelprüfung – XML-Transformation

Wurzelement

Aktion ▼

Neuer Elementname

Absatz-Formatvorlage **1_Kapitelueberschrift**

Aktion ▼

Neuer Elementname

Abschnitt mit Container-Element umgeben

Absatz-Formatvorlage **1_Ueberschrift2**

Aktion ▼

Neuer Elementname

Abschnitt mit Container-Element umgeben

Umgebendes Element

Wie läuft die Prüfung ab?

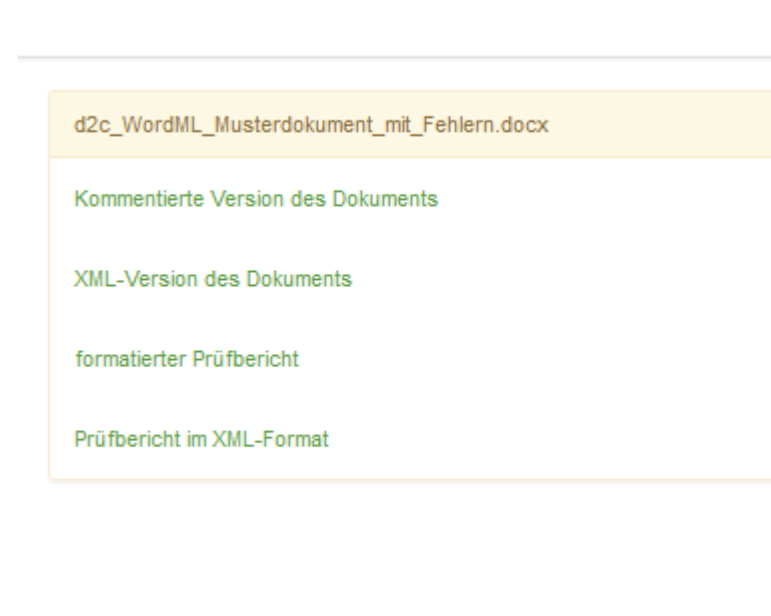
Der eingeloggte Benutzer lädt eine Datei hoch und wählt seine Prüfkongfiguration aus.

The screenshot shows the 'data2check' web application interface. At the top, there is a navigation bar with the logo 'data<2>check' and three menu items: 'Home', 'Konfiguration', and 'Dokumente'. Below the navigation bar, the main heading is 'Dokument prüfen'. Underneath the heading, there is a descriptive text: 'Sie haben hier die Möglichkeit Ihr Dokument als Word-, InDesign- und EPUB-Dateien auf ihre Korrektheit zu prüfen.' The main content area is divided into two sections. The first section, titled 'Datei hochladen:', contains a file selection button labeled 'Datei auswählen »', a text input field with the placeholder 'Keine Datei ausgewählt', a 'Prüfkongfiguration' label, and a dropdown menu currently showing 'IDML-EPUB - InDesign-Dokume'. The second section is titled 'Prüfungshistorie'.

Wie läuft die Prüfung ab?

Es kommen verschiedene Dokumente zurück.

1. Das Word-/InDesign-Dokument mit den Fehlermeldungen.
2. Eine XML-Version.
3. Ein formatierter Prüfbericht.
4. Ein technischer XML-Prüfbericht.



Wie läuft die Prüfung ab? Ergebnisse

Die kommentierte/überarbeitete Version:

The screenshot shows a document editor interface. The main content area displays XML code for a Word document. The code is as follows:

```
<?xml:version="1.0"?>
<?mso-application:progid="Word.Document">..... (1)
<w:wordDocument.
xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml"> (2)
  <w:body>..... (3)
    <w:p>..... (4)
      <w:r>..... (5)
        <w:t>Hallo Welt!</w:t>..... (6)
      </w:r>
```

Annotations in the image include:

- A blue box highlights the title "Einführung in data2type".
- A purple box highlights the XML namespace declaration: `xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml">`.
- A purple box highlights the text "Hallo Welt!" within the XML code.

On the right side, there are three error messages from data2type:

- data2check-error**: Treffer durch String-Vergleich mit "datatotype". Wurde ersetzt durch "data2type".
- data2check-**: Unbekannte Absatz-Formatvorlage "Untertitel".
- data2check-**: Unbekannte Zeichen-Formatvorlage "Link".

Wie läuft die Prüfung ab? Ergebnisse

Die XML-Datei:

```

<title>Einführung in datatotype</title>
<Subtitle>Der Grundaufbau</Subtitle>
<_2AbschnittStart>Abschnitt Start</_2AbschnittStart>
<d2tp>Das folgende Hallo Welt-Beispiel wurde manuell erzeugt und stellt die kürzeste Variante eines
  WordML-Dokumentes mit einem Absatz dar:</d2tp>
<_2CodeStart>Code Abschnitt</_2CodeStart>
<_3CodeAbschnitt><?xml version="1.0"?></_3CodeAbschnitt>
<_3CodeAbschnitt><?mso-application progid="Word.Document"?>          (1)</_3CodeAbschnitt>
- <_3CodeAbschnitt>
  <w:wordDocument xmlns:w="
  <link>http://schemas.microsoft.com/office/word/2003/wordml</link>
  "> (2)
  </_3CodeAbschnitt>
  <_3CodeAbschnitt> <w:body> (3)</_3CodeAbschnitt>
  <_3CodeAbschnitt> <w:p> (4)</_3CodeAbschnitt>
  <_3CodeAbschnitt> <w:r> (5)</_3CodeAbschnitt>
  <_3CodeAbschnitt> <w:t>Hallo Welt!</w:t> (6)</_3CodeAbschnitt>
  <_3CodeAbschnitt> </w:r></_3CodeAbschnitt>
  <_3CodeAbschnitt> </w:p></_3CodeAbschnitt>
  <_3CodeAbschnitt> </w:body></_3CodeAbschnitt>
  <_3CodeAbschnitt></w:wordDocument></_3CodeAbschnitt>
<_2CodeEnde>Code Ende</_2CodeEnde>
- <d2tp>
  Die PI (
  <_4Link>Processing Instruction</_4Link>

```

Wie läuft die Prüfung ab? Ergebnisse

Der formatierte Fehlerbericht:

data2check – Fehlerreport

- Zusammenfassung

Formatvorlagen – 6 Fehler

Absatz-Formatvorlage

- [Fataler Fehler] Unbekannte Absatz-Formatvorlage "Untertitel". (1)

Zeichen-Formatvorlage

- [Fataler Fehler] Unbekannte Zeichen-Formatvorlage "Link". (5)

Mikrotypografie – 1 Fehler

Contains-Fehler

- [Fehler] Treffer durch String-Vergleich mit "datatotype". Wurde ersetzt durch "data2type". (1)

Missbrauch von Objekten – 2 Fehler

Textfelder

- [Fehler] Textfelder sind nicht erlaubt. (1)

Bilder

- [Fehler] Bilder sind nicht erlaubt. (1)

Fragen?



Wieblinger Weg 92a

69123 Heidelberg

T: ++49-(0)6221-7391264

F: ++49-(0)6221-7391266

[E: montero@data2type.de](mailto:montero@data2type.de)