

Spezifikation „Layout-getreue Darstellung von OOXML-Dokumenten in Open Source Office Applikationen“

Version 6, 12. Dezember 2011

Open Source Business Alliance (OSBA) Working Group
„Office Interoperability“

Inhaltsverzeichnis

Einleitung	2
Zielsetzung.....	2
Organisation mittels Working Group der OSBA.....	2
Zeitplan (Änderungen vorbehalten).....	2
Rahmenbedingungen	3
Begriffsverständnis.....	3
Testdokumente.....	3
Code-Basis.....	3
Lizenz.....	4
Technische Anforderungen	4
Use Case 1: Formatierung von Rahmen und Bildern in .docx.....	4
Ziele.....	4
Testdokumente.....	4
Use Case 2: Formatierung von Tabellen in .docx.....	4
Ziele.....	4
Bemerkungen.....	5
Testdokumente.....	5
Use Case 3: Formatierung von Aufzählungen in .docx.....	5
Ziele.....	5
Bemerkungen.....	5
Testdokumente.....	6
Use Case 4: Formatierung von Kommentaren in .docx und .xlsx.....	6
Ziel.....	6
Bemerkungen.....	6
Testdokumente.....	6
Use Case 5: Einbettung von Schriften in OOXML und ODF.....	6
Ziele.....	6
Bemerkungen.....	6
Testdokumente.....	7
Ausschreibungsverfahren	7
Angaben zum Anbieter.....	7
Angaben zum Vorgehen.....	7
Sprache.....	7
Termin.....	7
Auswahlverfahren.....	7

Einleitung

Das vorliegende Dokument stellt die Spezifikation dar, aufgrund derer Softwarefirmen eine Offerte für die technische Umsetzung der Anforderungen erarbeiten können. Die Spezifikation basiert auf den Ergebnissen des Workshops vom 10. und 11. Oktober 2011 in Zürich, an dem öffentliche Institutionen und LibreOffice/OpenOffice-Entwicklern teilnahmen und in gemeinsamen Vorgehen die Anforderungen festlegten.

Zielsetzung

Das Ziel der folgenden Anforderungen ist es, die Layout-getreue Darstellung von OOXML-Dokumenten in LibreOffice/OpenOffice zu verbessern. Die Vorgabe dabei ist nicht, eine perfekte Integration der OOXML-Spezifikation in LibreOffice/OpenOffice zu erreichen. Vielmehr soll das pragmatische Vorgehen ermöglichen, mittels gezielten Korrekturen beim OOXML-Filter und ggf. weiteren Software-Komponenten in LibreOffice/OpenOffice eine möglichst ähnliche Visualisierung von OOXML-Dokumenten wie in Microsoft Office 2007/2010 zu erreichen und damit den Anwendernutzen zu verbessern.

Organisation mittels Working Group der OSBA

Dieses Projekt wird durch die Working Group „Office Interoperability“ der Open Source Business Alliance (OSBA) geleitet. Die Working Group steht allen Mitgliedern der OSBA offen. Interessierte können sich beim Sprecher der Working Group melden.

Die Mitglieder der Working Group stellen gemeinsam die Finanzierung der Umsetzungsarbeiten sicher. Sie bestimmen einen Sprecher, der die zentrale Anlaufstelle für Kommunikation und Koordination des Projekts darstellt. Der Sprecher der Working Group veröffentlicht die Spezifikation mittels Medienmitteilung, nimmt die Offerten der Softwarefirmen entgegen, begleitet das Auswahlverfahren des Auftragnehmers, erarbeitet den Mustervertrag, koordiniert die Finanzierungsvereinbarungen und Vertragsunterzeichnungen, überwacht die Umsetzungsarbeiten und begleitet die Abnahme der Softwareentwicklung. Der Sprecher der Working Group unterzeichnet jedoch keine Verträge im Namen der Working Group. Aufträge werden direkt von Working Group Mitgliedern und weiteren Geldgebern an den Auftragnehmer vergeben.

Dazu werden die Umsetzungsarbeiten in geeignete Bereiche aufgeteilt und durch die Mitglieder der Working Group sowie weiteren Geldgebern direkt als Auftrag an den ausgewählten Auftragnehmer vergeben. Die Umsetzungsarbeiten werden aufgrund ihrer überschaubaren Grösse und zur Bewältigung des Koordinationsaufwandes an einen einzigen Auftragnehmer vergeben. Durch einen Mustervertrag oder einen harmonisierten individuellen Vertrag, den alle geldgebenden Institutionen mit dem Auftragnehmer abschliessen, ist sichergestellt, dass die Leistungen aufeinander abgestimmt sind und letztlich ein funktionierendes Endprodukt geliefert wird.

Zeitplan (Änderungen vorbehalten)

- 12. Dezember 2011** Veröffentlichung der Spezifikation und der Testdokumente (Zugang auf Anfrage)
- 31. Januar 2012** Eingabefrist von Offerten durch Softwarefirmen
- 29. Februar 2012** Abschluss des Auswahlverfahrens, Auftragnehmer ist festgelegt
- 31. März 2012** Alle Verträge sind unterzeichnet, Umsetzungstätigkeiten beginnen
- 31. Mai 2012** Alle Anforderungen sind umgesetzt, getestet und abgenommen
- 30. Juni 2012** Quellcode ist unter einer Open Source Lizenz veröffentlicht und im LibreOffice Projekt integriert, das Projekt ist abgeschlossen

Rahmenbedingungen

Die Rahmenbedingungen regeln übergeordnete Vorgaben zur Umsetzung der Anforderungen.

Begriffsverständnis

- Wenn nachfolgend von Microsoft Word, Microsoft Excel oder Microsoft PowerPoint die Rede ist, sind damit die Versionen 2007 und 2010 gemeint.
- OOXML wird in diesem Dokument als Oberbegriff für die mit Microsoft Office 2007 eingeführten neuen XML-basierten Dokumentenformate .docx, .xlsx und .pptx verwendet.
- Die Bezeichnung LibreOffice/OpenOffice steht repräsentativ für alle auf OpenOffice.org basierenden Open Source Office Applikationen.
- Mit „darstellungsgetreu“ ist eine pixelgenauen Positionierung bei einer Auflösung von 70 DPI gemeint.
- Mit „valide“ ist eine Validierung des XML für ODF 1.0/1.1 gemäss den „strict“ Anforderungen und für ODF 1.2 gemäss der „conformant“ Anforderungen gemeint.

Testdokumente

Die Testdokumente der Use Cases dienen dem Auftragnehmer beim Testen der Umsetzungsarbeiten als Referenzdateien und gelten als Abnahmekriterium des jeweiligen Use Cases. Dies sind typischerweise .docx-Dateien mit bestimmten Layout-Elementen sowie zugehörige, aus Microsoft Office exportierte PDF-Dateien, welche die Zieldarstellung repräsentieren. Wird somit ein Testdokument mit der verbesserten OOXML-Darstellung in der Open Source Office-Applikation geöffnet und als PDF exportiert, sollte dieses weitgehend identisch mit dem Referenz-PDF übereinstimmen.

Die Erfüllung der Anforderungen wird mittels Pixelvergleich zwischen den Testdokumenten und dem PDF-Export des verbesserten OOXML-Filters bestimmt. Der Pixelvergleich ist durch den Auftragnehmer mittels Officeshots¹ oder einer vergleichbaren Verifikationsmethode umzusetzen.

Darüber hinaus muss LibreOffice/OpenOffice in der Lage sein, das Testdokument als valides ODF zu speichern und verlustfrei zu laden. Falls hierzu ODF XML Erweiterungen notwendig sind, müssen diese dem OASIS ODF TC vorgeschlagen werden, Validität gilt dann inklusive dieser vorgeschlagenen ODF-Erweiterungen. Die Validität muss mit dem Validator des Apache ODF Toolkits² oder einem gleichwertigen Validator geprüft werden.

Die Testdokumente sind in den entsprechenden Ordnern im gemeinsamen Verzeichnis auf Filespots³ abgelegt. Interessierte Auftragnehmer erhalten Zugang auf das Verzeichnis auf Anfrage an den Sprecher der Working Group.

Code-Basis

Die zu entwickelnde Software soll auf der Code-Basis des aktuellen OOXML-Filters in LibreOffice aufbauen. Dieser Filter stellt die heute beste Open Source Implementierung des OOXML-Standards dar.

Lizenz

Der Auftragnehmer verpflichtet sich, die entwickelte OpenOffice/LibreOffice-Software im Rahmen dieses Projekts unter der Apache License Version 2.0 zu veröffentlichen. Diese Lizenz erlaubt die breitest mögliche Weiterverwendung der Entwicklungsleistungen.

¹ <http://officeshots.org>

² <http://incubator.apache.org/odf/odftoolkit/conformance/ODFValidator.html> <http://odf-validator.rhcloud.com>

³ <https://www.filespots.com>

Technische Anforderungen

Nachfolgend sind die Detailspezifikationen der so genannten Use Cases aufgeführt. Diese Use Cases stellen eine Bündelung von thematisch ähnlich gelagerte Anforderungen dar. Sie enthalten je eine inhaltliche Beschreibung sowie mehrere Testdokumente.

Use Case 1: Formatierung von Rahmen und Bildern in .docx

Ziele

Mit Microsoft Word formatierte *Rahmen* von Tabellen und Paragraphen müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können:

1. Die Rahmendicke entspricht dem Originaldokument
2. Der Rahmenstil entspricht dem Originaldokument
3. Die Position des Rahmens entspricht dem Originaldokument

In Microsoft Word-Dokumente eingebettete *Bilder* müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können:

1. LibreOffice/OpenOffice unterstützt den Import, die Einbettung und den Export aller Bildformate, welche auch von Microsoft Office unterstützt werden.
2. Enthält ein OOXML-Dokument ein unbekanntes Bildformat, wird anstelle des Bildes ein Platzhalter mit den Originaldimensionen angezeigt, um das Dokumentenlayout des Originals zu bewahren.

Testdokumente

Siehe Dateien im entsprechenden Ordner auf Filespots.

Use Case 2: Formatierung von Tabellen in .docx

Ziele

1. Mit Microsoft Word formatierte *Tabellen* müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können.
2. In Microsoft Word erstellte *verschachtelte Tabellen* müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können
3. In Microsoft Word definierte vertikale und horizontale *Abstände* in und um Tabellen müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können.

Bemerkungen

Mit Office 2007 hat Microsoft neue Vorlagen für Tabellendesigns eingeführt. Damit ist es möglich, Tabellen mit einem Mausklick aus einer Vielzahl von Designs zu formatieren. Eine Spezialität dabei ist die unterschiedliche Formatierung gerader/ungerader Spalten und Zeilen. Diese Designanweisungen für

Tabellen werden im OOXML-Format nicht „hardcodiert“ gespeichert, weshalb der LibreOffice/OpenOffice-Filter nicht weiss, wie die Tabellenelemente (Rahmen, Hintergründe und Schriften) formatiert werden sollen. LibreOffice/OpenOffice muss beigebracht werden, wie die in einer .docx-Datei eingebetteten Designklassen für Tabellen in akkurate LibreOffice/OpenOffice-Formatierungsanweisungen umgesetzt werden müssen. Als pragmatisches Vorgehen ist eine „Hardcodierung“ der Formatierungen in LibreOffice/OpenOffice denkbar.

Verschachtelte Tabellen werden in LibreOffice/OpenOffice mit einem Rahmen emuliert. Das Problem dabei ist, dass der Rahmen nicht auf die nächste Seite umbrechen kann. Tabellen tun dies aber problemlos, weshalb es zu Darstellungs- und Handlingproblemen kommt.

Testdokumente

Siehe Dateien im entsprechenden Ordner auf Filespots.

Use Case 3: Formatierung von Aufzählungen in .docx

Ziele

Mit Microsoft Word formatierte Listen (nummeriert oder unnummeriert) müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können:

1. Die Dimensionen (bspw. Einrücktiefe der Elemente) entsprechen den Massen und der Darstellung des Originaldokuments.
2. Die Bulletpoint-Symbole entsprechen dem Originaldokument.
3. Der Nummerierungsstil entspricht dem Originaldokument.
4. Die Schachtelung entspricht dem Originaldokument.
5. Die Dokumentenstruktur wird beim Import aus dem anschliessenden Export in das OOXML-Format nicht verändert.

Bemerkungen

Probleme mit der Darstellung des Bulletpoints hängen oftmals mit nicht vorhandenen Schriftarten zusammen. Als pragmatisches Vorgehen ist eine Einbettung der Bulletpoint-Symbole in LibreOffice/OpenOffice denkbar.

Die Dokumentenstruktur soll nicht verändert werden, um nicht ungewollt Probleme mit sich darauf stützenden automatischen Inhalten (bspw. Inhaltsverzeichnissen oder dergleichen) sowie mit VBA-Makros zu produzieren.

Testdokumente

Siehe Dateien im entsprechenden Ordner auf Filespots.

Use Case 4: Formatierung von Kommentaren in .docx und .xlsx

Ziel

In Microsoft Word und Microsoft Excel angebrachte Kommentare müssen, wenn im OOXML-Format abgespeichert, darstellungsgetreu in LibreOffice/OpenOffice importiert und auch wieder nach OOXML exportiert werden können.

Bemerkungen

In LibreOffice könnte bereits ein Bugfix eingeflossen sein, der das Problem mit der korrekten Positionierung von Kommentaren in Calc behebt. Ist dies der Fall, wäre eine Portierung des Patches verlangt.

Testdokumente

Siehe Dateien im entsprechenden Ordner auf Filespots.

Use Case 5: Einbettung von Schriften in OOXML und ODF

Ziele

1. Beim Speichern aus LibreOffice/OpenOffice können per Option nicht zur Grundausstattung gehörige Schriften sowohl in OOXML- als auch im ODF-Dateien eingebettet werden.
2. In OOXML- und ODF-Dateien eingebettete Schriften werden beim Import des Dokuments in LibreOffice/OpenOffice geladen und korrekt dargestellt.

Bemerkungen

Beim Speichern von OOXML- und ODF-Dateien muss LibreOffice/OpenOffice Schriften intelligent einbetten; d.h. die Entwickler müssen spezifizieren, welche Schriften auf einem System standardmässig erwartet werden können. Die in OOXML-Dateien exportierten Schriften müssen von Microsoft Office gelesen und geladen werden können.

Ferner ist zu beachten, dass die Einbettung von Schriften Auswirkungen auf die Dateigrösse hat. Schriftdateien können mehrere hundert Kilobytes umfassen und so den Speicherbedarf einer einzelnen Datei massiv in die Höhe treiben. Dieser Effekt ist mittels Einbettungs-Option beim Speicherprozess abzufedern.

Die Schriften-Einbettung soll mittelfristig in den ODF-Standard einfließen, da ein solches Feature bis jetzt noch nicht Teil des Standards ist. Als Beispiel einer Umsetzung sei auf Web Fonts verwiesen.

Die Einbettung von Schriften betrifft auch rechtliche Aspekte: Schrifthersteller legen in den Schriftdateien fest, ob eine Schrift in Dokumente ganz bzw. teilweise eingebettet und in dieser Form weitergegeben werden darf oder nicht (Copyright). Heute ist dies bei vielen, aber noch nicht bei allen Schriften der Fall. Es kann auch sein, dass der Schrifthersteller nur die Einbettung eines Subsets erlaubt, also eine teilweise Einbettung ermöglicht. In diesem Fall werden nur diejenigen Schriftzeichen im Dokument gespeichert, die im Text tatsächlich verwendet werden. Wird der Text auf einem Fremdsystem mit einem Zeichen ergänzt, welches nicht im Subset enthalten ist, muss eine Ersatzschriftart verwendet werden, was wiederum Formatierungsprobleme auslöst.

Testdokumente

Siehe Dateien im entsprechenden Ordner auf Filespots.

Ausschreibungsverfahren

Angaben zum Anbieter

Die Offerte muss eine detaillierte Vorstellung des Anbieters, seiner Geschichte, seiner Geschäftsleitung, seinen Mitarbeitern und deren Kompetenzen und Referenzprojekte und -kunden enthalten. Dazu werden Angaben über die Anzahl Mitarbeitende, deren Hintergrund und Erfahrung (CVs) und die bisherigen Referenzprojekte (Kunden, Projektumfang, Tätigkeiten etc.) erwartet. Ausserdem muss auf das für dieses Projekt geforderte Knowhow bezüglich LibreOffice/OpenOffice-Entwicklung, OOXML- und ODF-Spezifikation, OOXML-Filter etc. eingegangen und dargelegt werden, wie und weshalb der Anbieter diese Kompetenzen ausweist.

Angaben zum Vorgehen

Des Weiteren muss die Offerte beschreiben, nach welchem Vorgehen die technischen Anforderungen umgesetzt werden. Dazu wird ein Lösungskonzept mit Tätigkeiten, Zeitplan und Entwicklungsmethodik erwartet. Aspekte wie Testing, Dokumentation, Sicherheit und Stabilität sollten dabei berücksichtigt werden. Ausserdem muss beschrieben werden, wie nach Abschluss der Entwicklungstätigkeiten der entstandene Software Quellcode in den Maintree des LibreOffice Open Source Projekts Eingang findet und in weiteren offiziellen Releases integriert bleibt.

Sprache

Die Offerte und die Korrespondenz dazu muss in deutscher oder englischer Sprache verfasst sein.

Termin

Die Eingabe von Offerten erfolgt als PDF per Email bis spätestens 31. Januar 2012 an den Sprecher der Working Group. Bei Fragen oder Anregungen kann der designierte Sprecher der Working Group „Office Interoperability“ kontaktiert werden:

Dr. Matthias Stürmer, Ernst & Young
matthias.stuermer@ch.ey.com
+41 58 289 61 97

Auswahlverfahren

Die geldgebenden Mitglieder der Working Group sowie weitere Geldgeber entscheiden autonom über die Auswahl des Auftragnehmers. Falls aus Sicht der Mitglieder der Working Group keine passende Offerte eingereicht wurde, kann die Ausschreibung jederzeit aufgehoben werden.