Das Austauschformat XLIFF im Übersetzungsprozess

Mächtiges Mittel gegen Konvertierungswirrwarr

Michael Schneider, Christian Taube und Thomas Wedde

XLIFF ist ein offener und lizenzfreier Standard mit großem Potenzial für Übersetzungsprojekte. Er stellt den Text, dessen Übersetzung und wichtige Informationen zu Terminologie und Übersetzungsvorschlägen systemunabhängig dar.

XLIFF steht für XML Localisation Interchange File Format. XLIFF ist aber mehr als ein Dateiformat, es ist ein bei der OASIS (Organization for the Advancement of Structured Information Standards, www.oasis-open.org) verabschiedeter Standard zum Austausch von Dokumentendaten. Dieser Standard vereinfacht die Zusammenarbeit zwischen allen an Lokalisierungs- und Übersetzungsprozessen Beteiligten. Ein Beispiel: Für eine einfache Übersetzung schicken die Dokumentationsabteilung eines Maschinenherstellers und der Übersetzungsdienstleister Verzeichnisbäume und zahllose Dokumente inklusive Layoutdaten, Bildern, PDF-Dateien als Referenz hin und her – womöglich in verschiedenen Versionen. Hierbei können beliebige Dokumentformate beteiligt sein, die zum Erstellen und Produzieren der Dokumentation verwendet werden: Word-Dateien, InDesign- oder FrameMaker-Dokumente, reine Grafikformate, Auszüge aus Datenbanken etc.

Konvertierungen beim Übersetzungsprozess

Der Übersetzungsdienstleister setzt ein Translation-Memory-System (TM-System) ein, das vor allem und am besten den reinen Text verarbeitet. Bilder und Layoutdaten sind eigentlich nur störender Ballast, der die maximale Wiederverwendung schon vorhandener Übersetzungen behindert. Der Übersetzungsdienstleister muss zunächst die Kundendokumente so konvertieren, dass sie vom verwendeten TM-System verarbeitet werden können. Dabei wird der Text vom restlichen Dokument, sprich von Formaten und Struktur, getrennt. Nach der Übersetzung wird erneut konvertiert: Die fertigen Inhalte müssen ja in das ursprüngliche Dokument wieder eingebaut, also in das Layout eingefügt werden. Da bereits mehrfach konvertiert wurde, muss das Layout häufig von Hand repariert und – vor allem – aufwändige Qualitätssicherung betrieben werden, um sicherzustellen, dass die Konvertierungen und Layout-Korrekturen keine Fehler ergeben haben.

Schlimmer ist ein anderer Effekt: Die Übersetzer verwenden ebenfalls TM-Systeme, diesmal aber als komfortables Werkzeug zur Erstellung der Übersetzung. Moderne TM-Systeme haben speziell für den Übersetzer hochspezialisierte Oberflächen, um ihn bei der Arbeit maximal zu unterstützen. Automatisierte Übersetzungsvorschläge, Terminologieunterstützung und mehrsprachige Rechtschreibkontrollen sind nur ein paar der Funktionen, auf die ein Übersetzer heute nicht mehr verzichten sollte. Das bedeutet, dass der Übersetzer seine zu übersetzenden Texte in einer Form erhalten muss, die sein TM-System verarbeiten kann. Die auf dem Markt befindlichen TM-Systeme sind nicht kompatibel zueinander. Wird ein Übersetzer eingesetzt, der nicht das TM-System des Übersetzungsdienstleisters verwendet, müssen die Daten konvertiert werden. Diese Konvertierungen sind fehleranfällig und teils mit dramatischen Konvertierungsverlusten verbunden. Leider führt das dazu, dass ein Übersetzer in erster Linie wegen der von ihm verwendeten Werkzeuge ausgewählt wird und weniger aufgrund seiner fachlichen Qualifikation. Eine mehr als unglückliche Situation ...

Das Allheilmittel XLIFF soll die Lösung sein.

Wie XLIFF entstand

XLIFF gibt es in der dritten Überarbeitung (Version 1.2 ist zurzeit im sog. Public Review und steht damit kurz vor der Verabschiedung) als ein XML-basier-

tes Datenaustauschformat. Die Historie reicht zurück in das Jahr 2000: Vertreter der Firmen IBM, Novell, Berlitz Global-NET und Oracle gründeten eine Arbeitsgruppe, die Austauschformate im Übersetzungsprozess vereinheitlichen sollte. Daraus entstand XLIFF 1.0, das aber zunächst nicht veröffentlicht werden konnte, da sich die Beteiligten nicht über Urheberrechtsfragen einigen konnten. Statt den Streit gerichtlich zu regeln, traten die Hersteller im Jahre 2001 der OASIS bei und übergaben den „Working Draft" der OASIS. Der Working Draft 2a wurde von der OASIS am 15. September 2002, der aktuelle Standard XLIFF 1.1 schließlich am 31. Oktober 2003 veröffentlicht.

Die Arbeitsgruppe wollte den Datenfluss im gesamten Übersetzungsprozess unterstützen. Redaktionssysteme, Layout-Programme und klassische Textverarbeitungssoftware sollten ihre Inhalte als XLIFF-Datenstrom bereitstellen, entweder direkt oder über spezielle Filter. Die XLIFF-Struktur sollte so beschaffen sein, dass die XML-Daten unkonvertiert von TM-Systemen und Workflow-Systemen aufgenommen und weiterverarbeitet werden konnten. Insbesondere wurde die Schnittstelle zum (Einzel-)Übersetzer berücksichtigt.

Was XLIFF kann

Für diese Ziele lassen sich in XLIFF u. a. folgende Inhalte abbilden:

- Ursprünglicher Inhalt der Dokumente: Der gesamte Inhalt der Dokumente ist darstellbar. Das ist in erster Linie der vollständige Text in einer sinnvollen Reihenfolge. Der Text wird vorab in seine kleinsten unabhängig übersetzbaren Einheiten zerlegt, die sogenannten Segmente. Ein Segment entspricht meist einem Satz in der Quellsprache. Es kann sich auch um eine Kapitel- oder Spaltenüberschrift, eine Tabellenzelle oder einen ganzen Absatz handeln. Letztlich muss das Redaktionssystem bzw. der XLIFF-Filter anhand der Dokumentenstruktur und konfigurierbarer Segmentierungsregeln entscheiden, was ein Segment darstellt. Aus Sicht der TM-Systeme ist eine feine Granularität vorzuziehen; dabei gilt: Ein Satz ist immer ein Segment, ein Segment ist nicht zwangsläufig ein Satz.
- Kontextinformationen: Für das Verständnis und die Übersetzung ist es notwendig, Informationen wie das Fachgebiet, die verantwortliche Abteilung und Verweise auf externes Material zur Verfügung zu haben. Die Dekontextualisierung der Übersetzungsschnipsel wird zunehmend als Problem erkannt. XLIFF bietet an dieser Stelle Wertvolles: Zu jedem Produktionsschritt werden Kontextinformationen generiert. Rückfragen eines Übersetzers beispielsweise werden direkt in den XLIFF-Daten transportiert, ohne dass ein gesonderter Kommunikationsweg nötig ist.
- Übersetzungsvorschläge aus dem Translation-Memory-System oder aus maschineller Übersetzung: XLIFF ist hier besonders reichhaltig ausgestattet. Dem Übersetzer werden nicht nur mehrere Vorschläge geliefert, diese Vorschläge werden zudem mit Meta-Informationen verknüpft, sodass zu jedem Vorschlag angezeigt wird, aus welchem TM-System, aus welcher Quelle, in welchem Kontext und mit welcher Treffsicherheit diese Übersetzung ausgewählt wurde. Der Übersetzer erkennt, ob das Segment maschinell übersetzt wurde, und kann besser beurteilen, ob und welcher Vorschlag verwendbar ist.
- Informationen zu Qualität und Quelle der Übersetzungsvorschläge: Der Übersetzer erhält bei mehreren Übersetzungsvorschlägen Entscheidungshilfen. Übersetzte Segmente, die aus „ähnlichen" Übersetzungsprojekten stammen, deren Kontext womöglich bekannt und deren Terminologie abgesichert ist, erhalten den Vorzug vor maschinell übersetzten Texten.
- Terminologie: Informationen aus Terminologiedatenbanken wie Bedeutung, Übersetzung und Verwendung eines Terms werden direkt dem Segment mitgegeben, in dem ein Term aus der Datenbank verwendet wird. Über Filtereinstellungen lässt sich regeln, ob fachgebietsübergreifend gesucht wird oder textartspezifisch alle Synonyme inklusive einer Verwendungsempfehlung dargestellt werden.
- Die Übersetzung selbst.
- Verwaltungsinformationen wie Status, Termine, Zielsprachen etc. für die Workflow-Steuerung.
- Daten aus der Produktionsplanung und -steuerung: ERP, PDM-Systeme, Warenwirtschaftssysteme.
- Daten zur Unterstützung der Qualitätssicherung: abgelehnte oder verbesserte Übersetzungen, vorherige Übersetzungen und Bewertungen.
- Segmentabhängige Layoutinformationen wie Fettdruck einzelner Wörter, wissenschaftliche und mathematische Sonderzeichen oder Referenzierungen wie Marken für die automatische Indexerstellung.

Transparenter, lizenzfreier Standard

Bei dieser Mächtigkeit ist XLIFF ein offener Standard. Durch die Ansiedlung bei der OASIS ist sichergestellt, dass die XLIFF-Spezifikationen, ihre gesamte Dokumentation und der Prozess der Weiterentwicklung auch in Zukunft für alle Interessierten transparent, einsehbar und weiterentwickelbar sind.

Mehr noch: Die Verwendung von XLIFF ist – jetzt und in Zukunft – an keine Lizenzgebühren gebunden. Die Übergabe eines Standardisierungsprojektes an die OASIS bedingt ausdrücklich, dass alle Copyrights, Urheberrechte und Lizenzrechte unwiderruflich kostenfrei auf die OASIS übergehen und die OASIS das gesamte Material frei und öffentlich zur Verfügung zu stellen kann.

XLIFF ist ausdrücklich von Produkten und Firmen unabhängig. Das bedeutet für die Übersetzungs- und Lokalisierungsbranche: XLIFF bleibt unabhängig und offen.

An dieser Stelle ein paar Worte zur Konformität von XLIFF: XLIFF basiert auf bestehenden, standardisierten XML-Formaten wie UIML, OpenTAG oder TMX und fasst deren Anwendungsbereich zusammen. So ist es relativ einfach, bestehende Schnittstellen zu TMX auf XLIFF zu erweitern. Da XLIFF aber wesentlich breiter angelegt ist als TMX, das nur dem Translation-Memory-Austausch dienen soll, wäre ein simpler TMX-zu-XLIFF-Konverter bestenfalls ein Startpunkt. Sinnvoller ist es, die jeweilige Applikation mit echten, vollständigen XLIFF-Schnittstellen auszurüsten, die bestenfalls dynamisch in alle Richtungen Daten austauschen.

XLIFF ist selbst ein XML-Format. XML als Basistechnologie wird heute in allen Bereichen der Dokumentenbearbeitung ein-

XLIFF-Werkzeuge

- Heartsome Translation Suite bzw. XLIFF-Editor (www.heartsome.net)
- Open Language Tools XLIFF Translation Editor (http://open-language-tools.dev.java.net)
- file2xliff4j (http://file2xliff4j.sourceforge.net): ein Satz an Java-Klassen für die Konvertierung von HTML, Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Rich Text Format (RTF), OpenOffice.org Text und FrameMaker (MIF) nach XLIFF.
- Das Okapi-Framework (http://okapi.sourceforge.net) stellt einen Satz an Schnittstellen zur Verfügung, die bei der Entwicklung von Tools für die verschiedenen Bearbeitungsschritte im Übersetzungsprozess eingesetzt werden können. XLIFF als Austauschformat spielt eine wesentliche Rolle in den Filterkomponenten von Okapi. Der Autor des Frameworks, Yves Savourel, hat das Tool „Rainbow" als Referenzimplementation von Okapi ebenfalls der Open Source Community übergeben.
- Transolution, ein Projekt auf Basis der Python-Plattform (derzeit von den Autoren leider nicht aktiv verfolgt): Eine CAT-Suite, die mit XLIFF als Bearbeitungsformat arbeitet (http://transolution.python-hosting.com).

gesetzt. Es hat sich ein reichhaltiges Sortiment an Werkzeugen und Bibliotheken zur Softwareentwicklung etabliert. Jedes XML-basierte Redaktionssystem kann problemlos um eine XLIFF-Schnittstelle erweitert werden. Die meisten Layoutsysteme bieten nützliche XML-Schnittstellen. Auch hier zeigt sich, wie überaus wertvoll XML als offenes, weltweit standardisiertes Format zum Datenaustausch ist. XLIFF als eine spezialisierte Anwendung von XML verdient es, häufiger eingesetzt und unterstützt zu werden.

Kontakt:
Dr.-Ing. Wolfgang Sturz, Sprecher des Arbeitskreises Forum Open Language Tools (www.folt.de), Geschäftsführer transline Deutschland GmbH
sturz@transline-group.com

Autoren

Michael Schneider, Geschäftsführer beo Gesellschaft für Sprachen & Technologie mbH, Stuttgart
michael.schneider@beo-doc.de

Christian Taube, Senior Account Manager Matrix Communications AG, München
christian.taube@matrix-ag.com

Thomas Wedde, Senior Account Manager euroscript Deutschland GmbH, Berlin, Bonn, Augsburg
thomas.wedde@euroscript.de