

acrolinx IQ Intelligente Wiederverwendung bei SAP

Ein Pilotprojekt

Christine Thielen, SAP AG

SAP Language Services/Language Technology

15. Oktober 2010

1. Sprachprüfungen bei SAP

- Ressourcen und Werkzeuge
- Nutzungsstatistiken
- Warum Intelligente Wiederverwendung?

2. Projektplanung

- Projektphasen
- Pilotbenutzer / Textsammlungen

3. Bibliotheken für die Wiederverwendung

- Generierung und Validierung
- Erfahrungen

4. Intelligente Wiederverwendung

- Erste Ergebnisse (Testphase 1)
- Ausblick (Testphase 2 und 3)

1. Sprachprüfungen bei SAP

- Ressourcen und Werkzeuge
- Nutzungsstatistiken
- Warum Intelligente Wiederverwendung?

2. Projektplanung

- Projektphasen
- Pilotbenutzer / Textsammlungen

3. Bibliotheken für die Wiederverwendung

- Generierung und Validierung
- Erfahrungen

4. Intelligente Wiederverwendung

- Erste Ergebnisse (Testphase 1)
- Ausblick (Testphase 2 und 3)



■ SAP verfügt über

- ausgereifte interne Leitfäden
- firmeneigenes Terminologie Management System
- globale Dokumentationserstellung auf English
- viele unterschiedliche Umgebungen für Autoren und Übersetzer
- eigene Computerlinguisten

■ acrocheck/acrolinx IQ

- Prototyp (FLAG) 2000 getestet
- acrocheck in produktivem Betrieb seit 2002
- Diverse Plug-Ins verfügbar (MS Word, Arbortext Epic Editor, Batch Checker)
- Entwicklungsumgebung um firmeneigene Regeln zu entwickeln

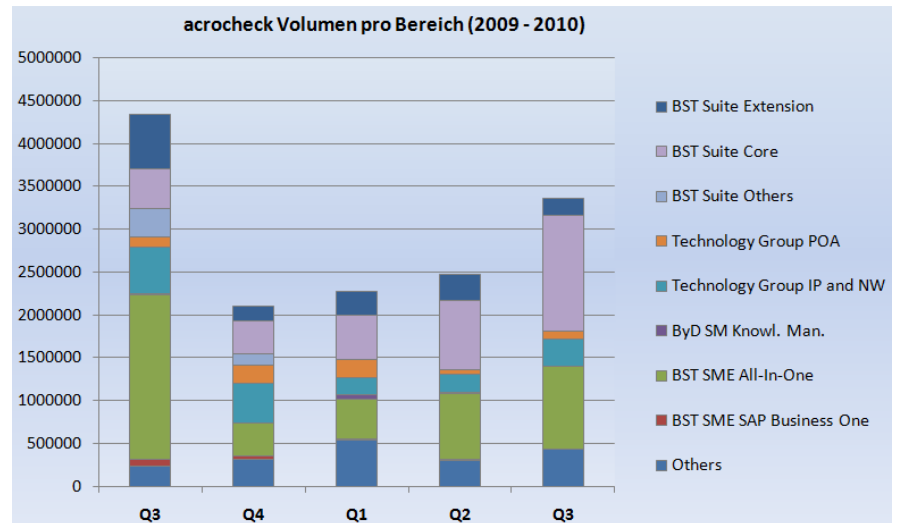
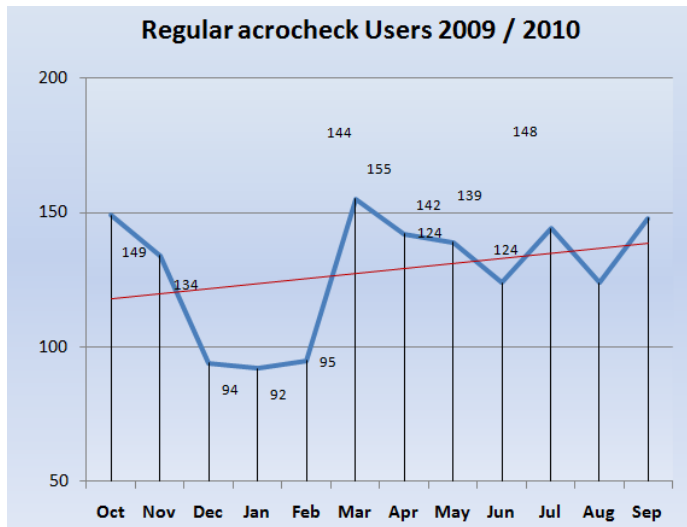
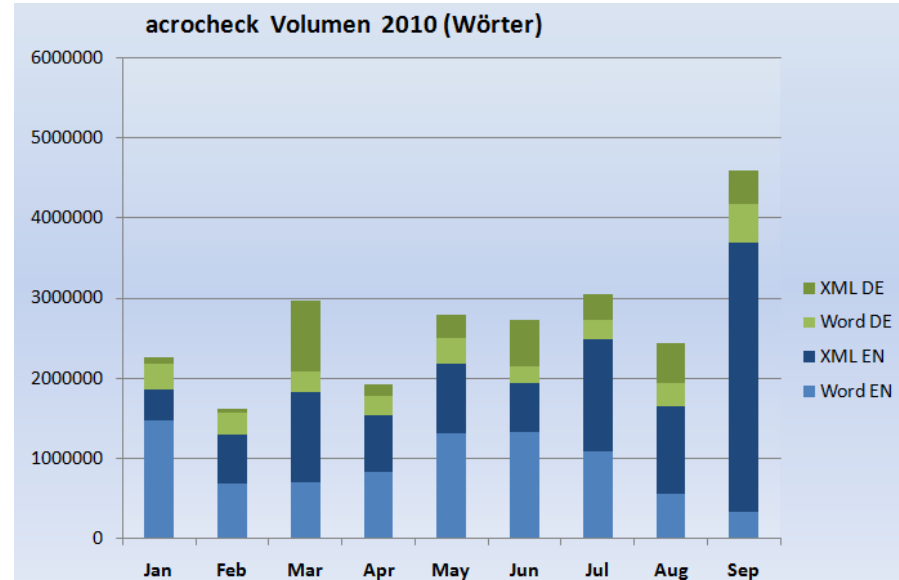
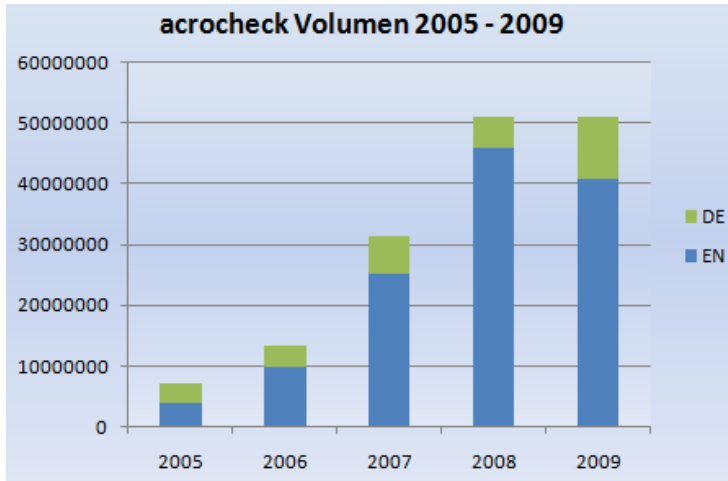
**THE BEST-RUN
BUSINESSES
RUN SAP™** 

acrolinx
INTELLIGENT TECHNOLOGY FOR TEXT QUALITY



Sprachprüfungen bei SAP

Nutzungsstatistiken



Erwartung: Mithilfe der Intelligenten Wiederverwendung wird bei der Erstellung und Übersetzung von technischer Dokumentation die Qualität gesteigert und die Kosten für die Übersetzung gesenkt.

- Prüfung von Standardformulierungen
 - Bisher relativ umständlich und nicht umfassend in Stilregeln implementiert, z.B. Standardformulierung „For more information about [...], see [...]“ benötigt 25 Trigger um alle bekannten Variationen dieses Satzes abzudecken
 - Vorhandene Listen könnten in Bibliotheken importiert werden (ab acrolinx IQ Version ???, in 1.5. nur über einen Workaround oder manuell)
- Unterstützung von Nicht-Muttersprachlern
 - Bei SAP wird ab 2011 auf Englisch als alleinige Entwicklungssprache umgestellt und somit werden immer mehr Texte / Dokumentation von Nicht-Muttersprachlern geschrieben
- Effizientere Nutzung von Translation Memories → Niedrigere Übersetzungskosten

1. Einleitung

- Sprachprüfungen bei SAP
- Warum Intelligente Wiederverwendung?

2. Projektplanung

- Projektphasen
- Pilotbenutzer / Textsammlungen

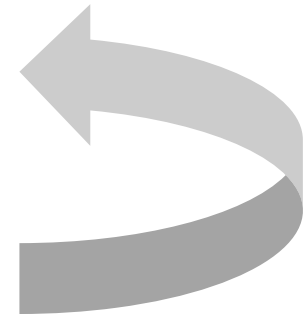
3. Bibliotheken für die Wiederverwendung

- Generierung und Validierung
- Erfahrungen

4. Intelligente Wiederverwendung

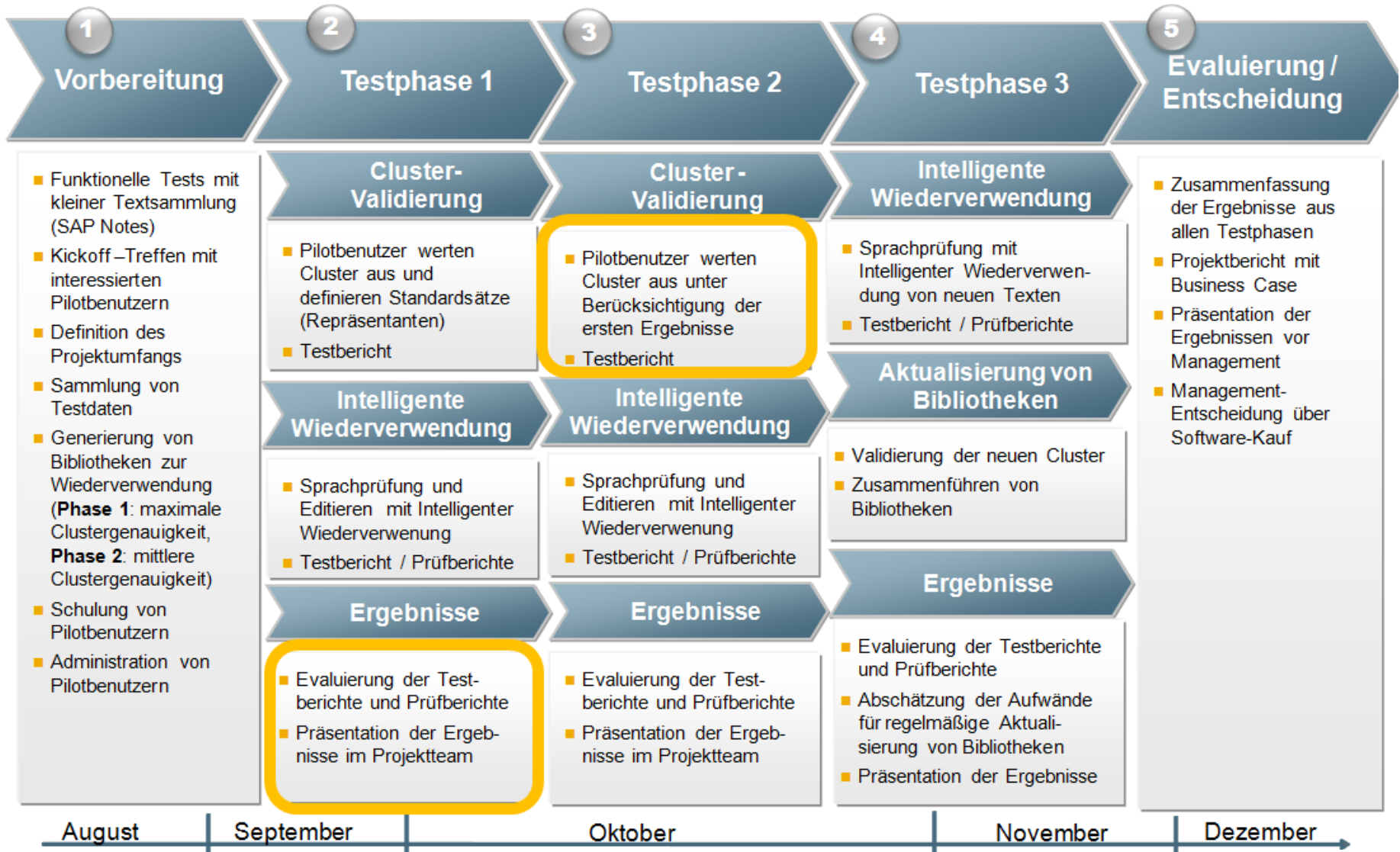
- Erste Ergebnisse (Testphase 1)
- Ausblick (Testphase 2 und 3)

1. Sätze sammeln → Satzbanken
2. Sätze clustern → Wiederverwendungsbibliotheken
3. Cluster bearbeiten → Wiederverwendungsbibliotheken
4. Wiederverwendungsbibliotheken für die Prüfung mit Intelligenter Wiederverwendung aktivieren
5. Prüfung mit Intelligenter Wiederverwendung
6. Cluster bearbeiten aufgrund der Ergebnisse der Prüfung
7. Neue Sätze sammeln → erweiterte Satzbank
8. Sätze clustern → Aktualisierte Wiederverwendungsbibliothek



acrolinx Intelligent Reuse at SAP

Project Plan





- 3 Pilotbenutzer aus dem Bereich der SAP-Lösungen für den Mittelstand (SAP ByDesign, SAP Best Practices)
- Profil der Pilotbenutzer: Autoren, S&G Owner (Experten für Standards und Richtlinien für das Schreiben bei SAP)
- Größe und Art der Textsammlungen:

	Texttyp (EN)	Umfang (Sätze)
By Design	Notes (Hinweise)	11015
By Design	How To, Quick Guide, Essentials	11630
Best Practices	Configuration Guide, Business Process Documentation	33239

1. Einleitung
 - Sprachprüfungen bei SAP
 - Warum Intelligente Wiederverwendung?
2. Projektplanung
 - Projektphasen
 - Pilotbenutzer / Textsammlungen
3. **Bibliotheken für die Wiederverwendung**
 - Überblick / Allgemeines Feedback
 - Erfahrungen im Einzelnen
4. Intelligente Wiederverwendung
 - Erste Ergebnisse (Testphase 1)
 - Ausblick (Testphase 2 und 3)

Es wurden 3 Bibliotheken für die Wiederverwendung aus den Textsammlungen erstellt (acrolinx IQ Version 1.5), einmal mit maximaler Cluster-Genauigkeit (Testphase 1), einmal mit mittlerer Cluster-Genauigkeit (Testphase 2)

Zusammenstellen von Satzbank-Clustern							Validierung (Phase 1)			
Bibliothek	Text-sammlung	Satz-bank	Maximale Cluster-Genauigkeit		Mittlere Cluster-Genauigkeit		Maximale Cluster-Genauigkeit			Dauer
	Sätze	Sätze	Cluster	Sätze	Cluster	Sätze	Cluster	Sätze	Davon „an“	
New Notes	11015	4322	161	472	222	659	170	569	122	6h
Best Practices	33239	12383	555	1436	919	2634	496	1285	120	?
ByDesign	11630	1302	17	35	61	131	34	80	26	5 h

Einstellungen:

Cluster Settings

Minimum word count

Minimum cluster size

Cluster strictness

Initial cluster status

acrolinx-Support bei SAP

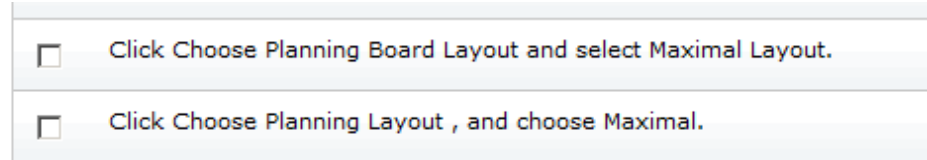
- Generierung der Bibliotheken (mithilfe des Batch Checkers) und Verwaltung im acrolinx Dashboard ohne größere Probleme
- Löschen einer Bibliothek ist umständlich (nicht im Dashboard möglich), Verzeichnisse unter *data* müssen explizit gelöscht werden
- Noch unklar, ob/wie wir die Generierung der Satzbanken beeinflussen können
- Fragen, die später aufgetaucht sind:
 - Welchen Einfluss haben Formatierungen (getaggte Format wie HTML) auf die Zusammenstellung der Cluster und wie können wir das beeinflussen?
 - Welchen Einfluss haben Segmentierungen (Satzendezeichen als Bestandteil ein Clustersatzes)?

Pilotbenutzer

- Benutzerberechtigungen: Zugriff im Dashboard auf die Komponente *Reuse* und Berechtigung, die Sprachserver-Konfiguration (EN) neu zu laden
- Nach kurzer Schulung ist die Bearbeitung der Cluster im acrolinx Dashboard ohne größere Probleme möglich
- Benutzer können Cluster-Repräsentanten ändern, neue Sätze zu Clustern hinzufügen, Cluster splitten und neue Cluster erstellen
- Etwas gewöhnungsbedürftig: explizites Refresh vor dem Laden der Sprachserver-Konfiguration



- ByDesign Dokumentation: relativ geringe Anzahl an Clustern obwohl gute Kandidaten in der Satzbank waren
- Bei mittlerer Clustergenauigkeit tauchen einige dieser Sätze dann in den Clustern auf



- Best Practices: sehr hohe Anzahl an Clustern, die nicht gebraucht wurden
- Hat die Formatierung (Tags für Screen Elements, Navigation Path, ...) Einfluss auf die Cluster-Zusammenstellung?
- Metadaten tauchen in Clustern auf, e.g.

```
<tr ct="POMNI" lsddata="{1:'Hide in This Release'}" id="GAKG.KTPBrowserView.Suppress:hideInThisRelease" td="ltr" class="urMnuRowOff">  
<tr ct="POMNI" lsddata="{1:'Hide in All Releases'}" id="GAKG.KTPBrowserView.Suppress:hideInAllReleases" td="ltr" class="urMnuRowOff">
```

→ Anpassung der Segmentierung im Batch Checker?

- Hauptänderungen: Zuweisung des Cluster-Repräsentanten

Active	ID	Representative	Matches
<input type="checkbox"/> on <input type="checkbox"/> off	10	Hide in This Release	0
		<input type="checkbox"/> Hide in All Releases	
		<input checked="" type="checkbox"/> Hide in This Release	

- Möglichkeit des Imports von vorhandenen Standardformulierungen wird vermisst, sie wurden zum Teil manuell eingegeben

Bibliotheken für die Wiederverwendung

Erfahrungen: Testphase 1



- Relativ zeitaufwändig bei einer großen Menge von Clustern
- Können bestimmte Textstellen (Menüpfade, Bildelemente, ...) schon für die Generierung der Cluster ausgeschlossen werden?
Hintergrund: Oberflächenelemente werden zum Teil automatisch in die Dokumentation eingefügt, der Autor kann/sollte daran nichts ändern
- Möglichkeit der Verwendung von Textvariablen wird stark vermisst

Beispiel:

Go to the [work center name] work center.

ID	Representative	Matches	La
2	Go to the 'Business Configuration' work center.	0	Ne
	<input type="checkbox"/> Go to 'Business Configuration' Work Center	0	
	<input type="checkbox"/> Go to 'Business Configuration' Work Center.	0	
	<input type="checkbox"/> Go to 'Business Configuration' work center.	0	
	<input type="checkbox"/> Go to 'Business configuration' Work Center.	37	
	<input type="checkbox"/> Go to 'Business configuration' work center.	0	
	<input type="checkbox"/> Go to Work Center 'Business Configuration'.	0	
	<input checked="" type="checkbox"/> Go to the 'Business Configuration' work center.	0	
	<input type="checkbox"/> Go to the work center 'Business Configuration'	0	
	<input type="checkbox"/> Go to the work center 'Business Configuration'.	0	
	<input type="checkbox"/> Go to the work center 'Business configuration'.	0	
	<input type="checkbox"/> Go to work center 'Business Configuration'	0	
	<input type="checkbox"/> Go to work center 'Business Configuration'.	0	
	<input type="checkbox"/> Go to work center 'Business Configuration'.	0	
	<input type="checkbox"/> go to 'Business Configuration' Work Center	0	
<input type="button" value="Edit Cluster"/> <input type="button" value="Create as New Cluster"/> <input type="button" value="Delete"/> <input type="button" value="Set Representatives"/>			
6	Go to the 'Customer Invoicing' work center.	198	30
	<input type="checkbox"/> Go to 'Customer Invoicing' Work Center	18	
	<input type="checkbox"/> Go to 'Customer Invoicing' work center	18	
	<input type="checkbox"/> Go to 'Customer Invoicing' work center.	18	
	<input type="checkbox"/> Go to Work Center 'Customer Invoicing'.	18	
	<input type="checkbox"/> Go to the 'Customer Invoicing' Work Center	18	
	<input type="checkbox"/> Go to the 'Customer Invoicing' Work Center.	18	
	<input checked="" type="checkbox"/> Go to the 'Customer Invoicing' work center.	18	
	<input type="checkbox"/> Go to the Work Center 'Customer Invoicing'.	18	
	<input type="checkbox"/> Go to the work Center 'Customer Invoicing'.	18	
	<input type="checkbox"/> Go to the work center 'Customer Invoicing'.	18	
	<input type="checkbox"/> Go to work center 'Customer Invoicing'.	18	
<input type="button" value="Edit Cluster"/> <input type="button" value="Create as New Cluster"/> <input type="button" value="Delete"/> <input type="button" value="Set Representatives"/>			

1. Einleitung

- Sprachprüfungen bei SAP
- Warum Intelligente Wiederverwendung?

2. Projektplanung

- Projektphasen
- Pilotbenutzer / Textsammlungen

3. Bibliotheken für die Wiederverwendung

- Überblick / Allgemeines Feedback
- Erfahrungen im Einzelnen

4. Intelligente Wiederverwendung

- Erste Ergebnisse (Testphase 1)
- Ausblick (Testphase 2 und 3)

Intelligente Wiederverwendung

Erste Erfahrungen (Testphase 1)



- Generell wenig falsche Alarme
- Schwer zu erkennen, ob es „fehlende Alarme“ gibt
- Müssen alle Sätze mit/ohne Satzendezeichen in der Bibliothek aufgeführt werden?

When creating or updating your demand forecast you typically do the following:

- Calculate or update the key figures
- Review the key figure **Actuals**
- Review the key figure **Final Act**
- **If your plan is multilevel, review**
- Review the key figure **Statistic**
- Review the key figure **Final For**

Prerequisites

Edit your demand plan setup. For more information, see [Edit Demand Plan](#).

- Warum werden hier mehrere Reuse-Sätze aktiviert?

- Wie können Symbole und Formatierungen bei automatischer Ersetzung korrekt übernommen werden?

Intelligente Wiederverwendung

Nächste Schritte (Testphasen 2 und 3)



Testphase 2:

- Anpassung der Segmentierung vor Erstellen der Cluster?
- Validierung der Cluster, die mit mittlerer Cluster-Genauigkeit erzeugt wurden → neue Bibliotheken
- Sprachprüfung mit Intelligenter Wiederverwendung auf Basis der neuen Bibliotheken

Testphase 3:

- Sprachprüfung mit Intelligenter Wiederverwendung von neuen Texten → Generierung neuer Satzbanken, Aktualisierung/Zusammenführung von bestehenden Bibliotheken



Danke!

Fragen und Antworten