

T2-Tutorial: Terminologiekontrolle und Maschinelles Lektorat

Symposion des DTT e.V.
Heidelberg, 15. – 17. April 2010

Dr. Christoph Rösener
IAI / Universität des Saarlandes
c.roesener@iai.uni-sb.de

- Fachrichtung 4.6 - Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen
- Lehrstuhl Maschinelle Übersetzung (Prof. Haller)
- Abteilung Sprachdatenverarbeitung
- ca. 500-600 Studenten

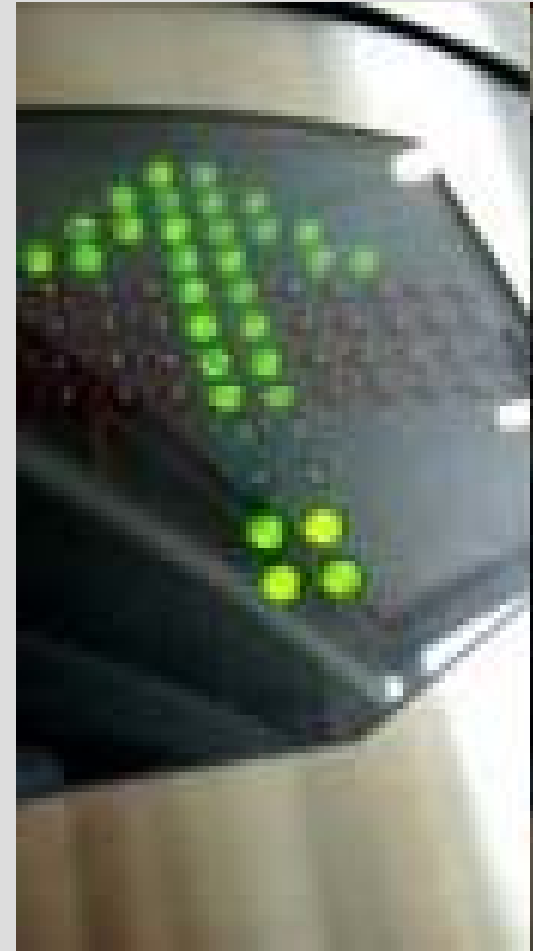
- Studiengang Bachelor VSLT
- Studiengang Master Translationswissenschaft

- <http://www.fr46.uni-saarland.de>

- Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e. V. an der Universität des Saarlandes
- seit 1985 maschinelle Sprachverarbeitung
- Forschungsprojekte (EU, national, regional)
- Produkte
 - Duden Korrektor
 - CLAT
 - Autindex
 - Dienstleistungen
 - Indexierung
 - Termextraktion und –evaluierung
 - Beratung etc.
- <http://www.iai-sb.de>

- Terminologie
 - Stellenwert, Anforderungen
- Terminologieprüfungswerkzeuge
 - Anforderungen, Funktionsweise
 - Konfigurierbarkeit
- Erschließung neuer Terminologie
- Workflow-Szenarien
- Demonstration des CLAT-Systems
 - CLAT-Server
 - MS-Word CLAT-In
 - Konfigurationswerkzeug UMMT

- Inkonsistente Schreibung
 - Bindestrich
 - *Axial-Kugellager* – *Axialkugellager*
 - Varianten
 - *Fotosensor* - *Photosensor*
 - Zahlen vs. Ziffern
 - *Viertürer* - *4-Türer*
 - Mischformen
 - *Unter-Menu* – *Untermenü*
- Verständlichkeit/Lesbarkeit:
 - **tolerierbar**
- Übersetzung:
 - **u. U. problematisch**



- Inkonsistente Bildung auf Ausdrucksebene
 - Variation in der Ableitung
Abkühlvorgang - Abkühlungsvorgang
 - Variation zwischen Ableitung und Zusammensetzung
Dampfstrahler – Dampfstrahlgerät
 - Variation zwischen Zusammensetzung und Wortgruppe
Nullpunktgleich - Abgleich des Nullpunktes
 - variierende Wortgruppensyntax
Säule hinten - hintere Säule
- Verständlichkeit/Lesbarkeit:
 - **grenzwertig**
- Übersetzung:
 - **problematisch**



- Inkonsistente Bildung auf Begriffsebene

- Zwischenstück – Füllstück

Lokalität

Funktion

- Unterlegblech – Unterlegplatte

Material

Form

- Sicherungsglasche - Verriegelungsglasche

Funktion1

Funktion2

- Verständlichkeit/Lesbarkeit:

- **nicht tolerierbar**

- Übersetzung:

- **sehr problematisch**



- Rückfragen der Zielgruppe (Leser)
 - Beispiel: 5 Minuten Rückfragen pro Tag und Werkstatt bundesweit **2.082.500,-** **€/Jahr**

- Rückfragen der Übersetzer
 - Beispiel: 5 Minuten Rückfragen pro Seite bei 10 Autoren: **100 Stunden/Monat**



ks113764 www.fotosearch.de

- Beeinträchtigung der Übersetzbarkeit
 - Einsatz von Translation Memory
 - Keine Durchführung der Sprachqualitätssicherung vor der Übersetzung führt zu Übersetzungsmehrkosten in Höhe eines **sechstelligen Euro-Betrags**



Gesteigerte Rechtssicherheit:

Technische Dokumentation unterliegt z.B.:

Produkthaftungsgesetz („IKEA-Klausel“)

US-Tread Act

DIN EN 292 (Sicherheit von Maschinen)

DIN EN ISO 9001

EU 5 Gesetzgebung



- kein Wildwuchs
- begriffsorientiert (konzeptbasiert)
- benennungsautonom (gleiche Datenfelder)
- Definiertes Datenmodell (DIN-konform)
 - *ISO 12620 - Computer applications in terminology - Data Categories*
- Mehrsprachig vs. einsprachig
 - Begriff – Sprache - Benennung
- in sich schlüssige Konzepte

=> Gute Qualität braucht Vorgaben!

Gute Terminologiequalität

- braucht Vorgaben
 - Aufbau einer Terminologie und Entwicklung von Termbildungs-Richtlinien

- entsteht beim Arbeiten
 - Sensibilisierung und Schulung der Redakteure

- muss überprüft werden
 - Einsatz eines Prüfwerkzeugs (Controlled Language Checkers)

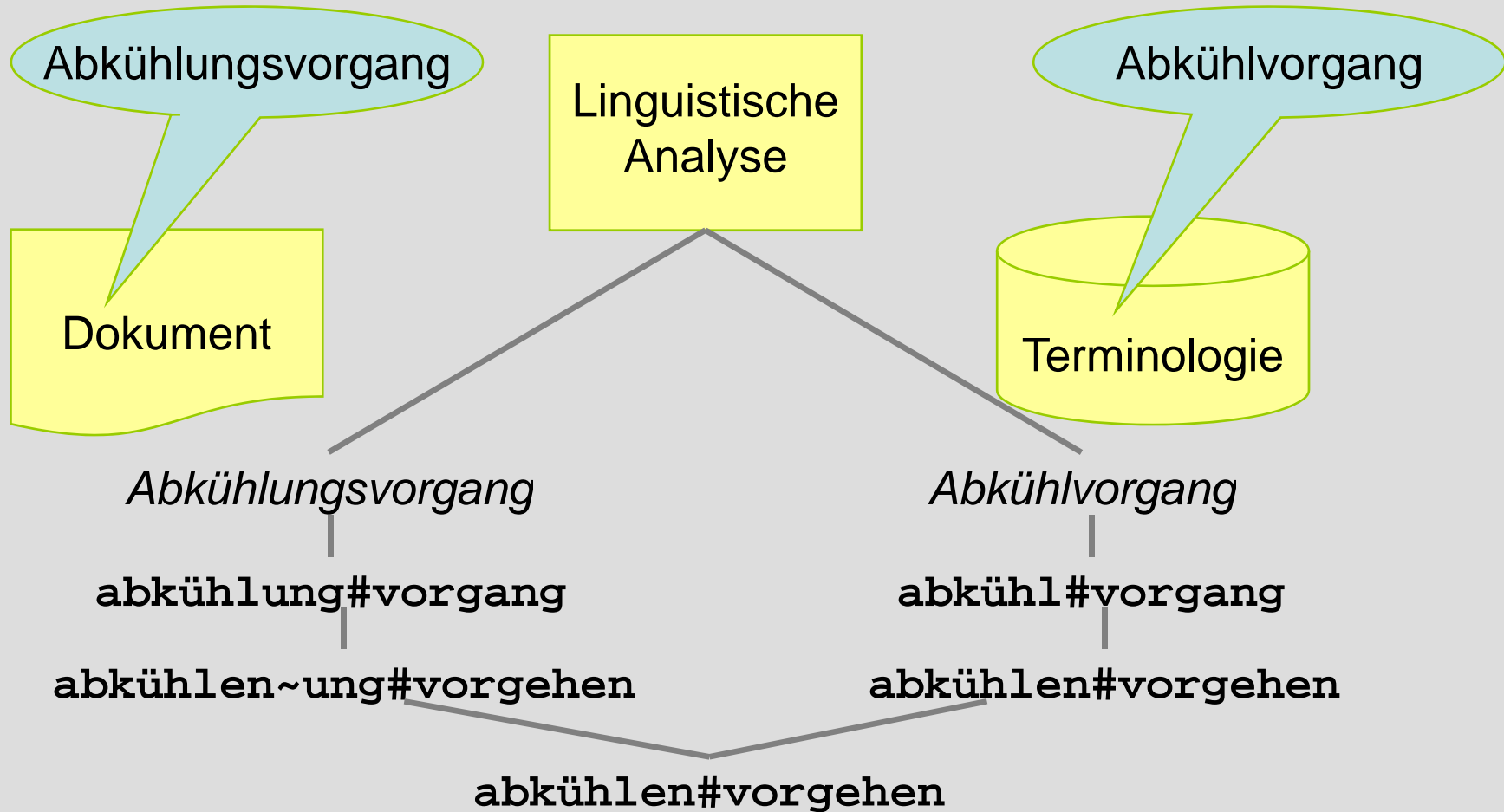


Allgemeine Anforderungen an eine Prüfsoftware



- Sinnvolle Autorenunterstützung
 - Performantes System
 - Keine Fehlalarme
 - Hoher Recall
 - Keine Fehlermeldung ohne Verbesserungsvorschlag
- Zentrale Datenhaltung
 - Proprietäre vs. Standarddatenbank
 - Export aus (mehrsprachiger) Terminologie-Datenbank
 - Synchronisierung in regelmäßigen Intervallen
 - automatisiert
 - manuell
- Konfigurierbarkeit
 - Projektspezifische Datenbestände
 - Prüfoptionen

- Basis: Linguistic Engine
 - morphologische Analyse
 - Regelmodule
- Linguistische Analyse des Terminologiebestands
- Linguistische Analyse der im Dokument vorkommenden Wörter und Sätze
- Abgleich beider Analysen



- Mögliche, nicht erlaubte, Varianten zu **Abkühlvorgang** werden automatisch erkannt
 - *Abkühl-Vorgang*
 - *Abkühlungsvorgang*
 - *Vorgang des Abkühlens*
 - *Vorgang Abkühlen*

- Allgemeine Synonymklasse: *Anzeige – Display*
- Mögliche, nicht erlaubte, Varianten von **Displayhelligkeit** werden automatisch erkannt:
 - *Display-Helligkeit*
 - *Helligkeit des Displays*
 - *Helligkeit Display*
- PLUS
 - *Anzeigehelligkeit*
 - *Anzeige-Helligkeit*
 - *Helligkeit der Anzeige*
 - *Helligkeit Anzeige*

- Linguistische Mächtigkeit
 - **Pro:** keine Notwendigkeit, alle Varianten zu hinterlegen
 - **Kontra:** manchmal zu weitgehend
 - *Starten des Fahrzeugs – Autostart*
 - *Position des Motors – Stellmotor*
- Einschränkungsmöglichkeiten durch den Benutzer

Terminologieprüfung für Termstatus admitted

Institut für Angewa

Informationsforsch

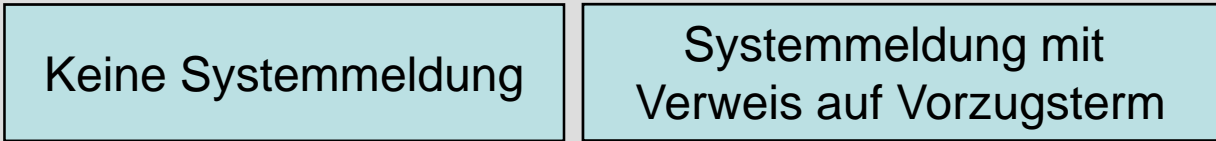
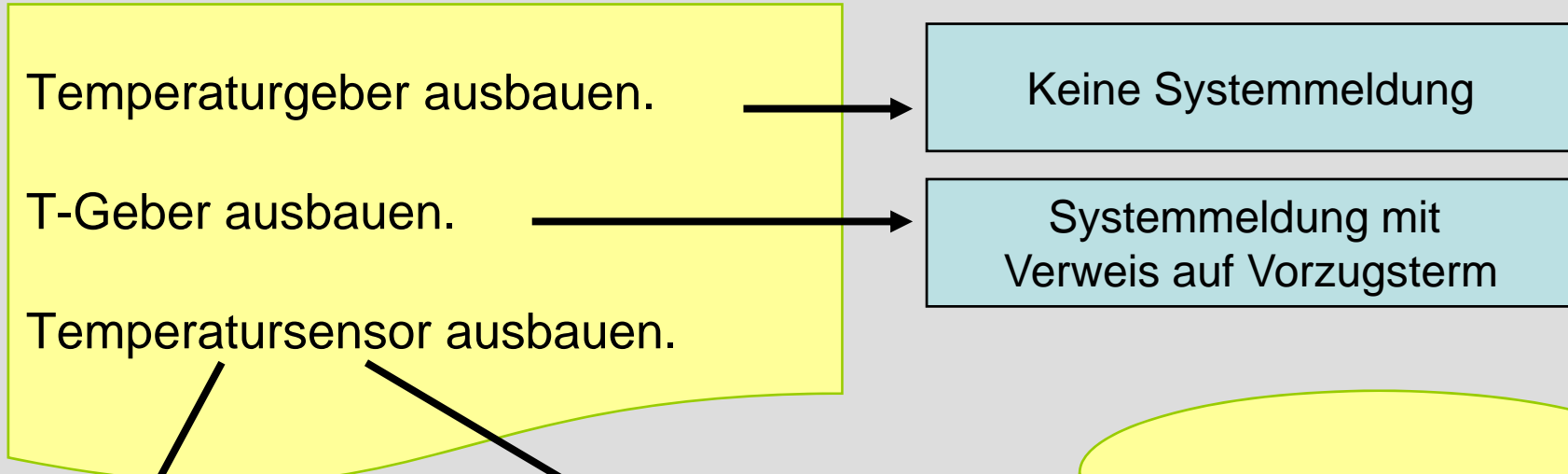


- Preferred (Vorzugsterm)
- Deprecated (Negativterm)
- Admitted (erlaubte Variante)
 - Häufigkeit von Status admitted korreliert mit dem Bearbeitungszustand der Datenbank
- Datenbankeintrag:
 - Konzept-ID 123; Temperaturegeber; preferred
 - Konzept-ID 123; Temperatursensor; admitted
 - Konzept-ID 123; T-Geber; deprecated

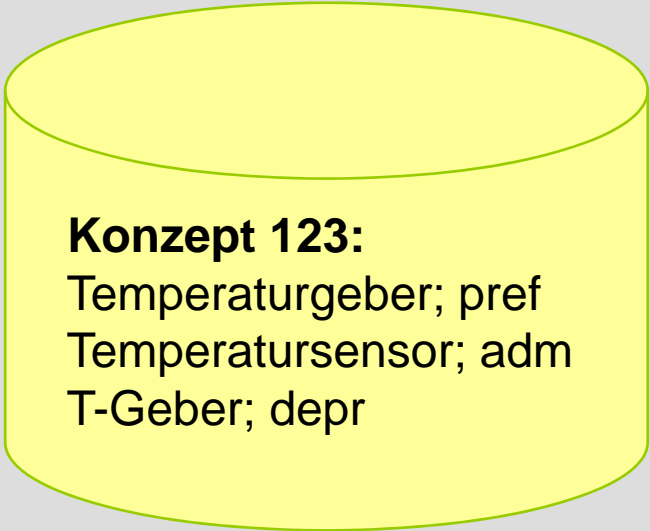
Terminologieprüfung für Termstatus admitted



Institut für Angewandte
Informationsforschung



ODER



- In unterschiedlichen Fachgebieten oder Dokumentarten ist ein und derselbe Term einmal Vorzugsterm und einmal Negativterm
- Datenbankeinträge:
 - Konzept-ID 0815; Frontscheibe; preferred
 - Konzept-ID 0815; Scheibe; deprecated
 - Konzept-ID 4711; Scheibe; preferred
 - Konzept-ID 4711; Bremsscheibe; admitted

Scheibe ausbauen.



Systemmeldung mit Hinweis auf
widersprüchlichen Status
PLUS
Anzeige beider Konzepte

Konzept 0815:

Frontscheibe; preferred
Scheibe; deprecated

Konzept 4711:

Scheibe; preferred
Bremsscheibe; admitted

- Sinnvolle Vorschlagsgenerierung, auch für Sonderfälle
- Benutzerdefinierte Behandlung von:
 - Erkennung unterschiedlicher Variantenmuster
 - Synonymisierung
 - Interpretation von *Status admitted*
 - Behandlung von Benennungen, die Vorzugs- und Negativterm sind

- Über Konfigurationswerkzeug
- Konfiguration von verschiedenen, sprachlichen Regel- und Datenpaketen für
 - verschiedene Anwendungen
 - Dokumentarten
 - Abteilungen
- Größte Anwendungsflexibilität durch Konfigurierbarkeit aller Ressourcen

- Realisierung
 - Parallel zur Terminologieprüfung
 - unter Berücksichtigung bereits vorhandener Terminologie
 - direkte Anbindung an den Terminologie-Workflow
- Ausgabe
 - in der Grundform
 - mit Frequenz und Kontext
- Filterfunktion
 - mit anwendungsspezifischen Filtern

Formular zur Termkandidatenermittlung



Termkandidaten

Termkandidaten

- Forstwirtschaft
- Inhaltsverzeichnis
- Linienbetrieb
- Metallhydrid-Speicher
- Naturflächenverbrauch

Kontext

Konzepte, die noch einen höheren Entwicklungsbedarf

Häufigkeit: 1

Termkandidat: * Metallhydrid-Speicher **Priorität:** Normal

einen höheren Entwicklungsbedarf haben, sind die Speicherung in Metallhydrid-Speichern oder in Nan

Title list 1 *

- Domain1
- Domain2
- Domain3

Title list 2 *

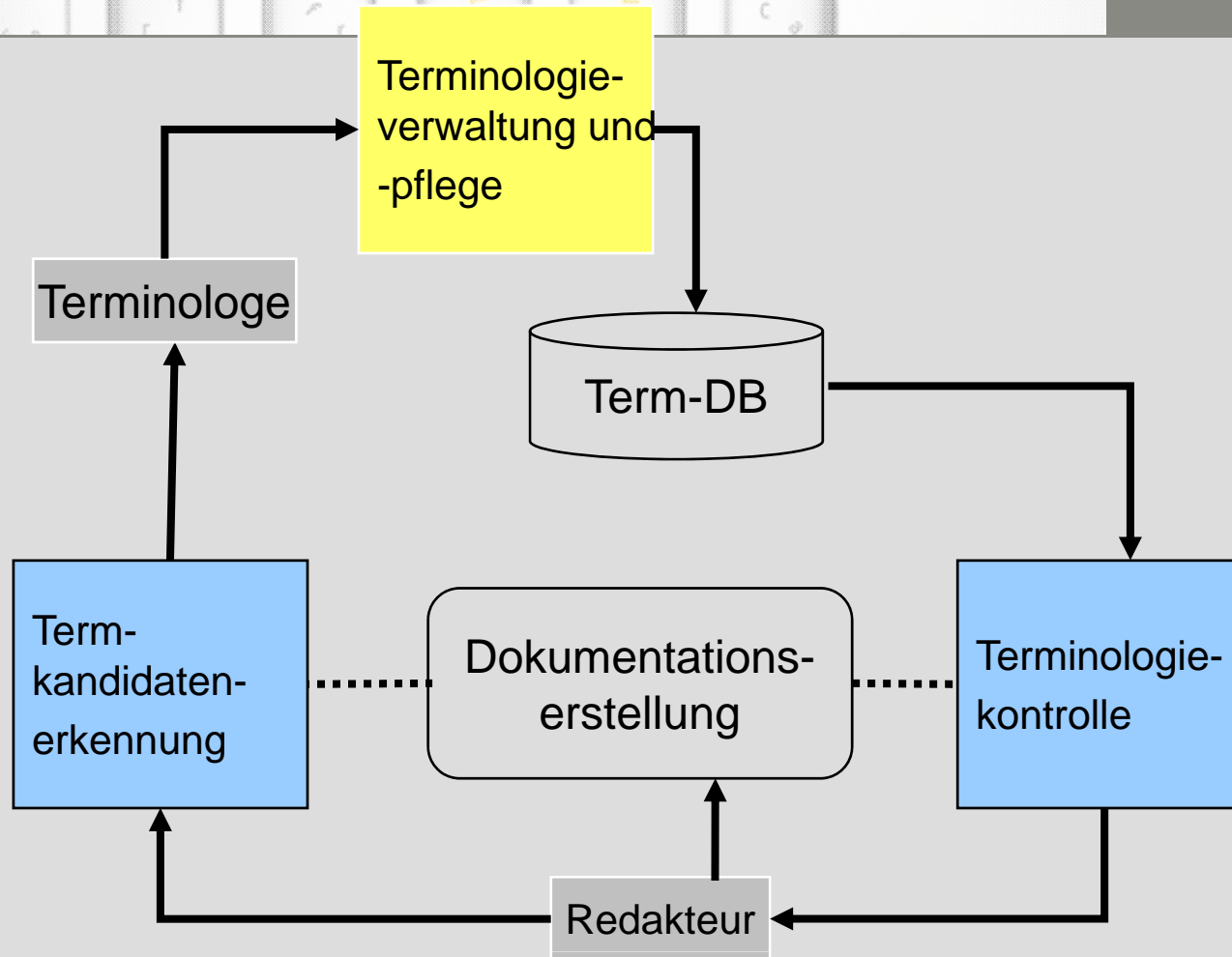
- Customer literature
- Workshop literature

Definition:

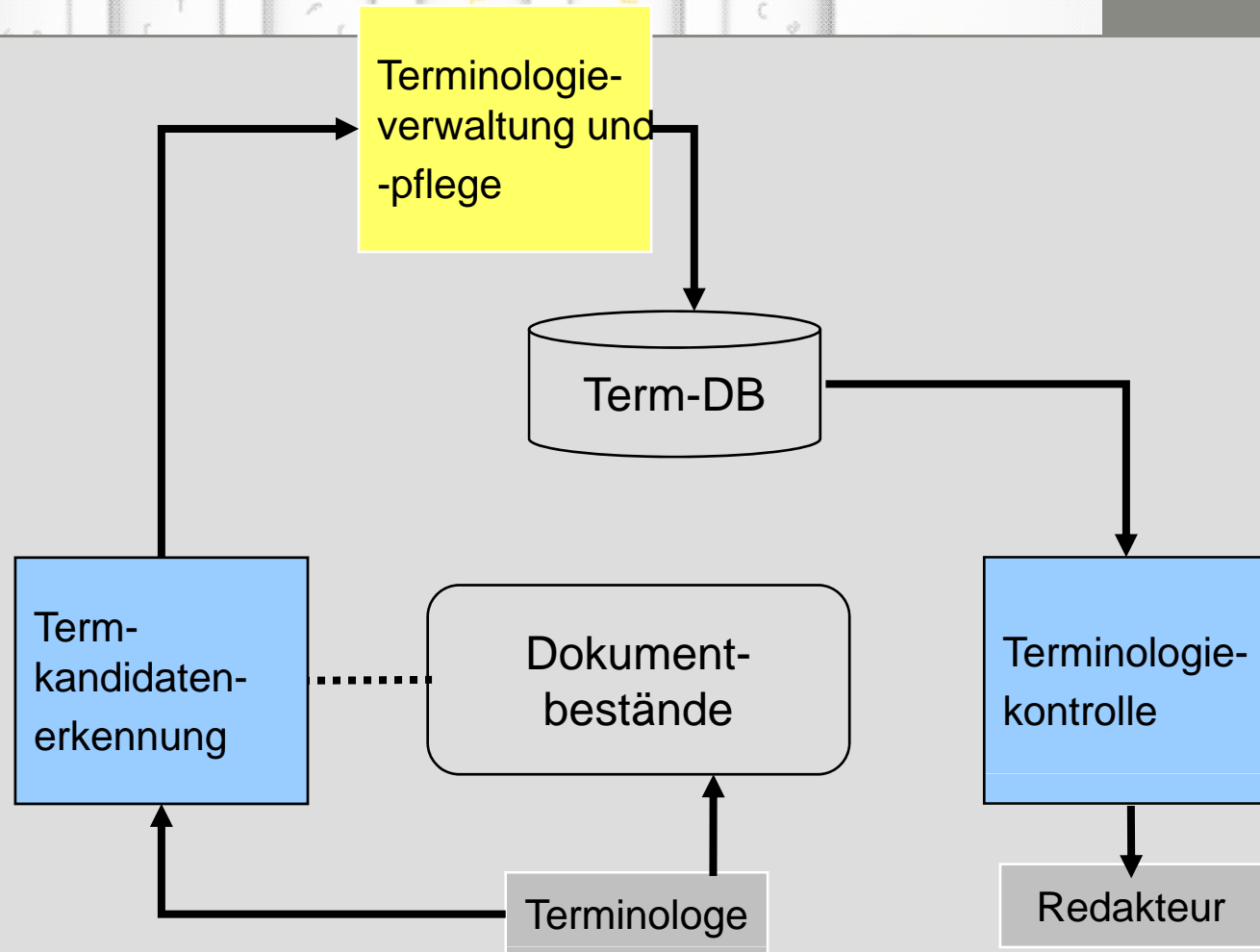
Kommentar:

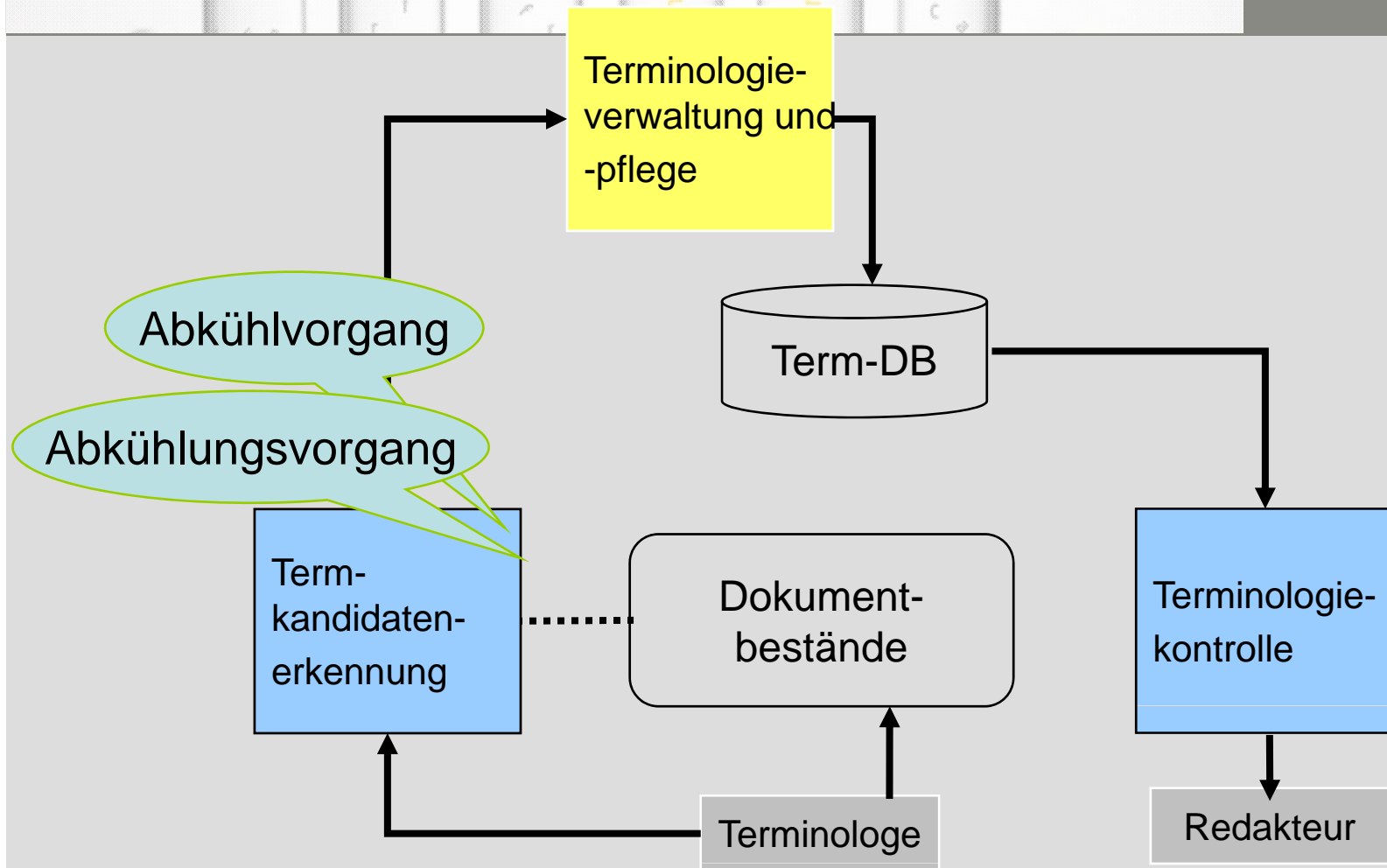
Verschicken Speichern unter Abbrechen Aktualisieren

Terminologie-Workflow Szenario 1

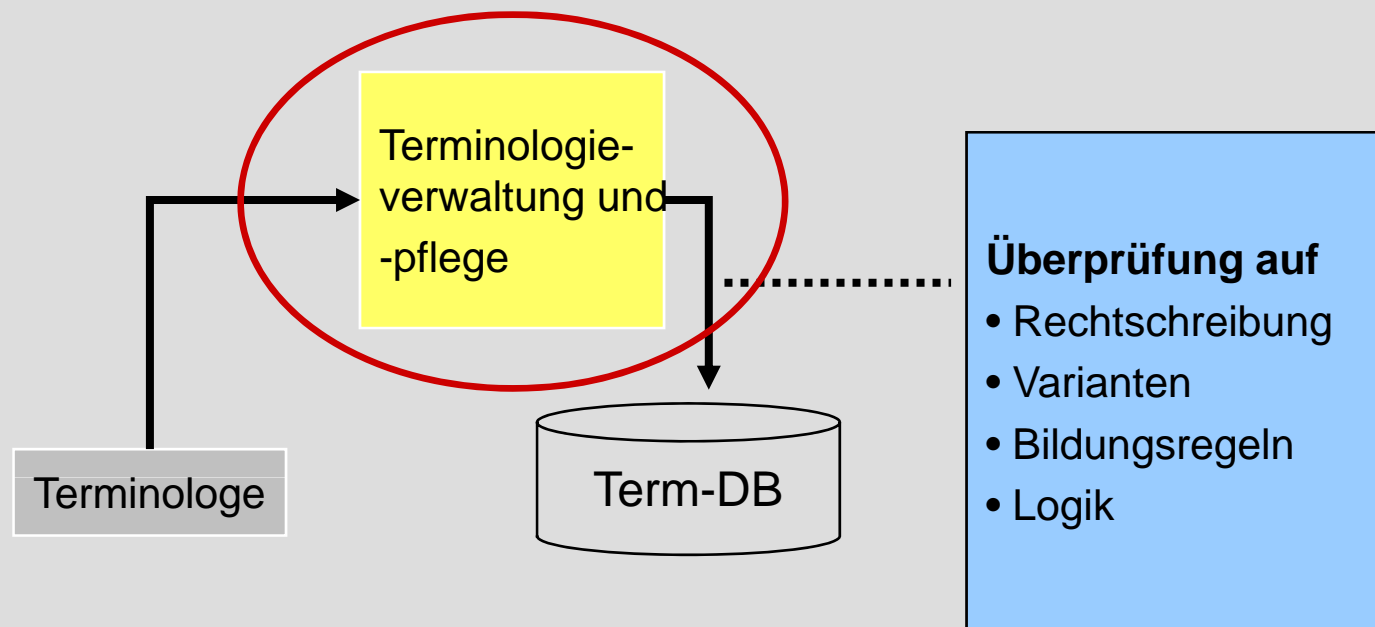


Terminologie-Workflow Szenario 2





Zusätzliche Unterstützung des Terminologie-Workflows – zukünftige Szenarien



DIN 2330 - Begriffe und Benennungen - Allgemeine Grundsätze

Demonstration des CLAT-Systems

- CLAT-Server
- MS-Word CLAT-In
- Konfigurationswerkzeug UMMT

Vielen Dank für Ihr Interesse!

Dr. Christoph Rösener

Fachrichtung 4.6 - Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen; Fakultät 4 – Philosophische Fakultät II; Sprach-, Literatur- und Kulturwissenschaften: Universität des Saarlandes

E-Mail: c.roesener@mx.uni-saarland.de

Institut der Gesellschaft zur Förderung der Angewandten Informationswissenschaft e.V. (IAI) an der Universität des Saarlandes

E-Mail: c.roesener@iai-sb.de