

IAI: Sprachkontrolle, Informationssuche, technische Dokumentation und Content- Management

Dr. Paul Schmidt

Institut für Angewandte Informationsforschung (IAI)

Saarbrücken

paul@iai.uni-sb.de

<http://www.iai.uni-sb.de>

- Institut an der Universität des Saarlandes
- Seit 1985
- Träger: GFAI e.V. (Gesellschaft zur Förderung der Angewandten Informationsforschung)
- Ca. 15 Mitarbeiter
- Auftrag: Technologietransfer im Bereich Sprachtechnologie / Computerlinguistik

IAI-Tätigkeitsfelder

- Computerlinguistische (angewandte) Forschung (34 Projekte):
 - Maschinelle Übersetzung
 - eLearning Fremdsprachen
 - Inhaltserschließung / Informationsextraktion / Informationssuche auf der Basis von Terminologieverarbeitung

IAI-Tätigkeitsfelder

- Kommerziell relevante Produkte, drei Themen:
 - Sprachkontrollwerkzeuge
 - Inhaltserschließung (Indexierung) / Informationssuche (Retrieval)
 - Dienstleistungen im Bereich Terminologieverarbeitung

1. Thema: ‚Sprachkontrollwerkzeuge‘:

- Worum geht es bei Sprachkontrolle?
- Wieso ist das wichtig für die Industrie?
- Was sind die Grundprobleme?

Sprachkontrollwerkzeuge



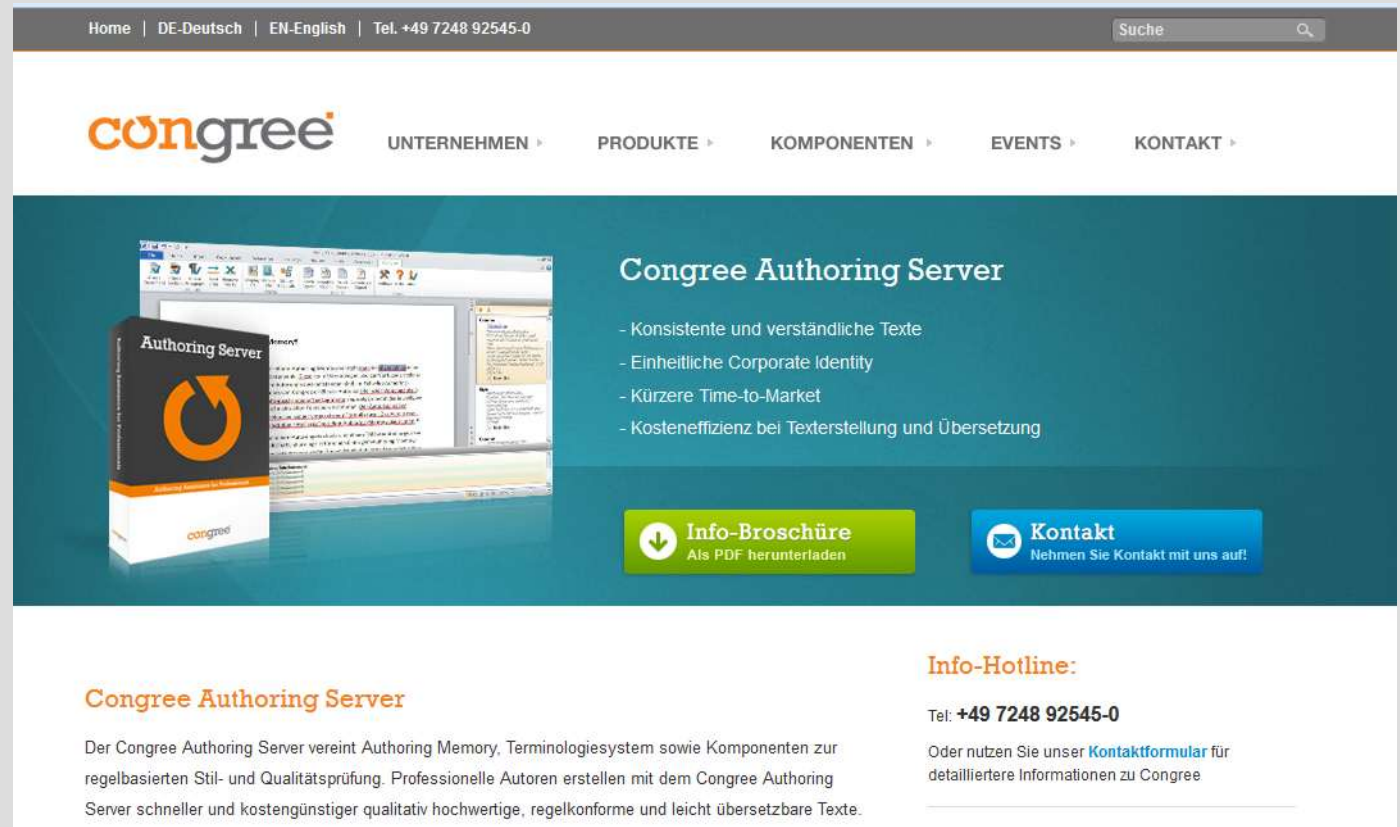
- CLAT (Controlled Language Authoring Tool):
Softwareunterstützung für Technische Redakteure

CLAT-Kunden

- Automobilindustrie: Daimler, Volkswagen, Porsche, Audi, BMW, Renault
- Technologieunternehmen: Siemens, Sun Microsystems, Heidelberger Druckmaschinen, MAQUET AG
- Stadt Wien

- Seit 1. Oktober 2010: Gemeinsame Firma mit der Firma ‚across‘

www.congree.com



The screenshot shows the Congree website homepage. At the top, there is a navigation bar with links for 'Home', 'DE-Deutsch', 'EN-English', and 'Tel. +49 7248 92545-0'. A search bar is located on the right. Below the navigation bar is the Congree logo and a menu with 'UNTERNEHMEN', 'PRODUKTE', 'KOMPONENTEN', 'EVENTS', and 'KONTAKT'. The main content area features a large image of the 'Congree Authoring Server' software box and a screenshot of the software interface. To the right of the image, the text 'Congree Authoring Server' is displayed, followed by a list of features: 'Konsistente und verständliche Texte', 'Einheitliche Corporate Identity', 'Kürzere Time-to-Market', and 'Kosteneffizienz bei Texterstellung und Übersetzung'. Below this list are two buttons: 'Info-Broschüre Als PDF herunterladen' and 'Kontakt Nehmen Sie Kontakt mit uns auf!'. At the bottom of the page, there is an 'Info-Hotline:' section with the phone number '+49 7248 92545-0' and a link to a 'Kontaktformular' for more information.

„Sprachkontrollwerkzeuge“:

- Duden-Korrektor:
Allgemeine Rechtschreib-
und Grammatikkontrolle
- Elektronisches
Hauptprodukt des Duden-
Verlags



„Sprachkontrollwerkzeuge“:

- Technische Dokumentation: Es geht um Fachkommunikation
 - Firmenintern: Entwicklungsabteilung mit Vertrieb, Ingenieure untereinander, Ingenieur mit Werksleitung, ... Vertriebler untereinander ... Ingenieur mit Reparaturbetrieb
 - Firma – Kunden
 - Auf verschiedenen Ebenen:
 - Fachlich: Experten (z.B. BMW – Vertragswerkstätten)
 - Globalisierung: Mehrsprachige (Fach-)Kommunikation

„Sprachkontrollwerkzeuge“:

- Textsorten:
 - Handbücher
 - Reparaturanleitungen
 - Hilfen
 - Werbung
 - Memos
 - Protokolle
 -

„Sprachkontrollwerkzeuge“:

- Anforderungen technische Dokumentation:
 - Verständlichkeit der Kommunikation
 - ...
- Wodurch wird Verständlichkeit erzielt?
 - Sätze / Texte haben keine Fehler
 - Sätze / Texte sind nicht zu kompliziert
 - Sätze / Texte haben nur eine Interpretation, sind nicht mehrdeutig
 -

Behinderung der Verständlichkeit:

- *„Es gab eine Undichtheit am Kraftstoff-entlüftungswellrohr von rechter Tankkammer zu Tankeinfüllstutzen infolge Knickbeschädigung anlässlich der Tankmontage.“*
- *"Kopieren Sie die Dateien von Ordner (A) in Ordner (B). Er enthält die Werkzeug- und Ersatzteilkataloge."*
- *"Den Schlauch oben abklemmen."*

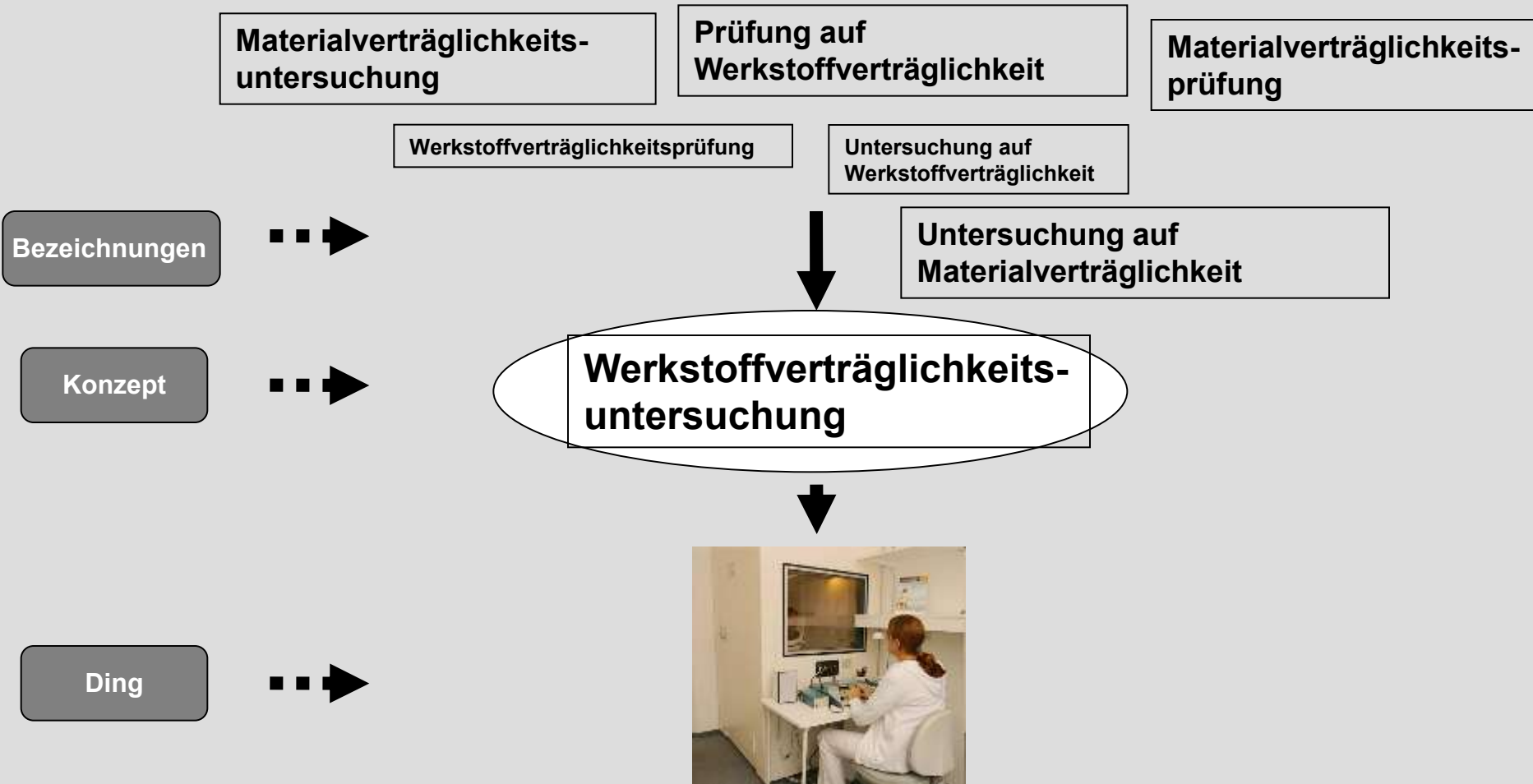
Kontrollierte Sprache:

- Ausschluss aller sprachlichen Elemente, die die Verständlichkeit behindern
 - Qualitativ einwandfreier Text
 - Nur bestimmte sprachliche Konstruktionen verwenden (z.B. ‚kein Passiv‘)
 - Nicht zu viele Bedeutungseinheiten in einen Satz, keine zu komplexen Konstruktionen
 - Keine Mehrdeutigkeiten !!!

Kontrollierte Sprache:

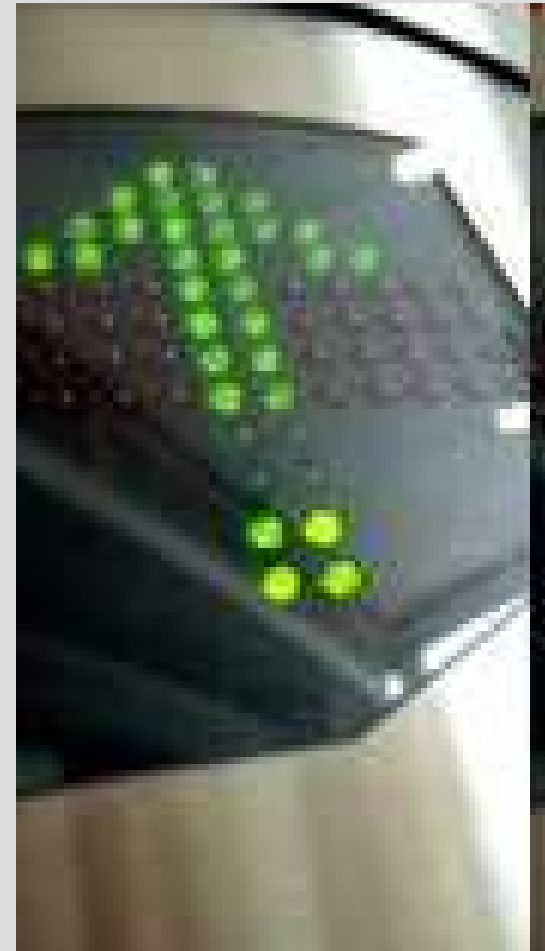
- Mehrdeutigkeiten müssen v.a. im Fachvokabular vermeiden werden
- Praktisch wichtigster Punkt bei der Sprachkontrolle: Eindeutige Terminologie: Für jeden Begriff nur eine Bezeichnung und für jede Bezeichnung nur ein Begriff
- Das ist das anerkannte Prinzip der modernen Terminologearbeit

Problem: Terminologie



Terminologie-Inkonsistenzen:

- Inkonsistente Schreibung
 - Bindestrich
 - *AxialKugellager – Axialkugellager*
 - Varianten
 - *Fotosensor - Photosensor*
 - Zahlen vs. Ziffern
 - *Viertürer - 4-Türer*
 - Mischformen
 - *UnterMenu – Untermenü*



Terminologie-Inkonsistenzen:

- 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



Terminologie-Inkonsistenzen:

- Inkonsistente Bildung auf Begriffsebene
 - Zwischenstück – Füllstück
 - Lokalität Funktion
 - Unterlegblech – Unterlegplatte
 - Material Form
 - Sicherungsglasche - Verriegelungsglasche
 - Funktion1 Funktion2



Negative Effekte

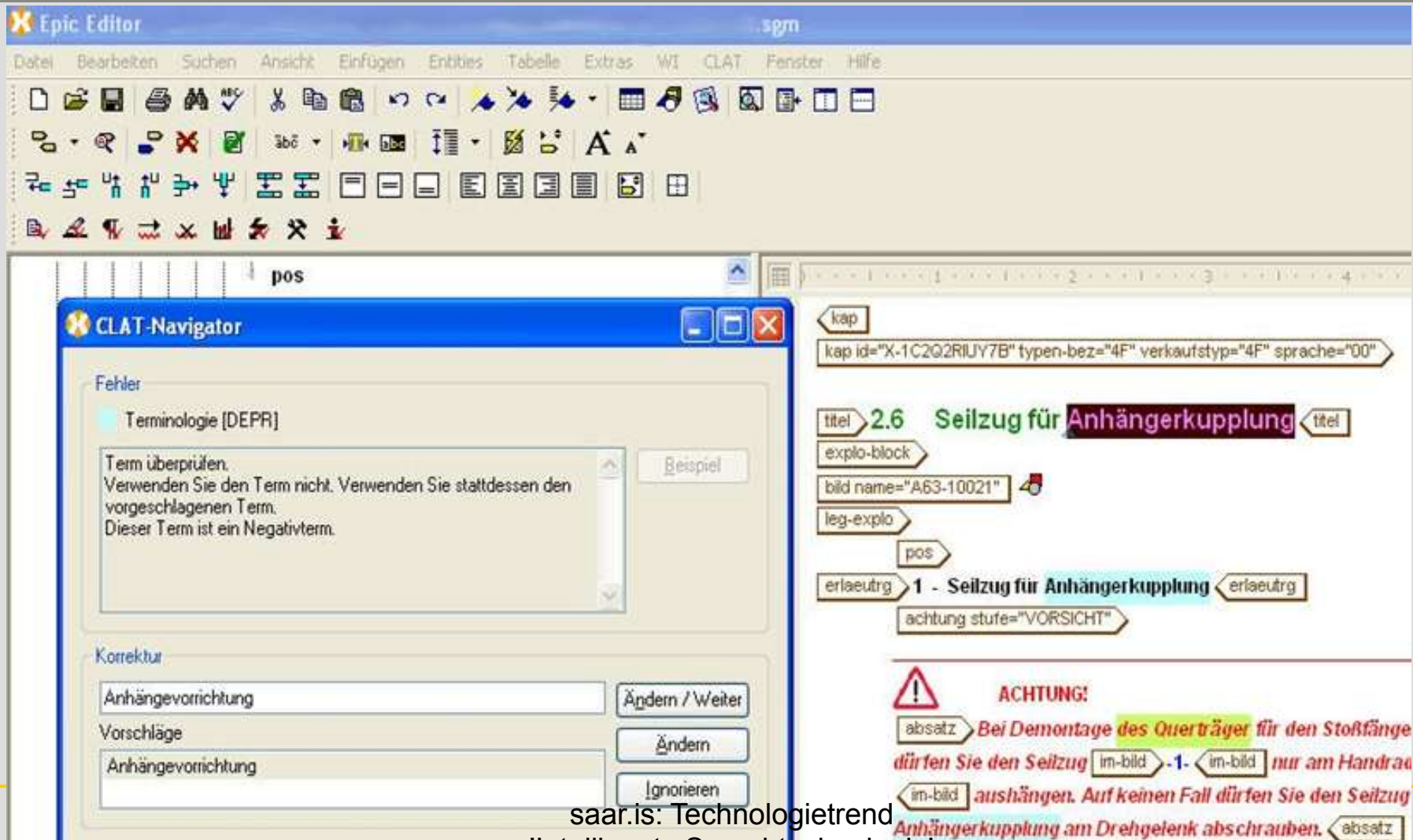
- Beeinträchtigung der Verständlichkeit
 - Rückfragen der Zielgruppe (Leser)
 - Beispiel: 5 Minuten Rückfragen pro Tag und Werkstatt bundesweit

2.082.500,- €/Jahr



- Das IAI-Kontrollierte Sprache-Tool prüft beim Schreiben, ob die Richtlinien eingehalten werden
 - Korrektheit: Grammatik und Rechtschreibung
 - Bestimmte sprachliche Konstruktionen, die schwer verständlich sind
 - **Standardterminologie!!!!**

IAI-Tool: CLAT



The screenshot shows the Epic Editor interface with a document titled 'pos'. A 'CLAT-Navigator' window is open, displaying a terminology error. The error message reads: 'Terminologie [DEPR] Term überprüfen. Verwenden Sie den Term nicht. Verwenden Sie stattdessen den vorgeschlagenen Term. Dieser Term ist ein Negativterm.' The suggested correction is 'Anhängerkupplung'. The main document content includes technical specifications and a warning section:

kap id="X-1C2Q2RIUJY7B" typen-bez="4F" verkaufstyp="4F" sprache="00"

titel **2.6 Seilzug für Anhängerkupplung**

explo-block

bild name="A63-10021"

leg-explo

pos

erlaeutrg **1 - Seilzug für Anhängerkupplung**

achtung stufe="VORSICHT"

ACHTUNG!

absatz Bei Demontage des Querträgers für den Stoßfänge dürfen Sie den Seilzug **1** nur am Handrad aushängen. Auf keinen Fall dürfen Sie den Seilzug Anhängerkupplung am Drehgelenk abschrauben.

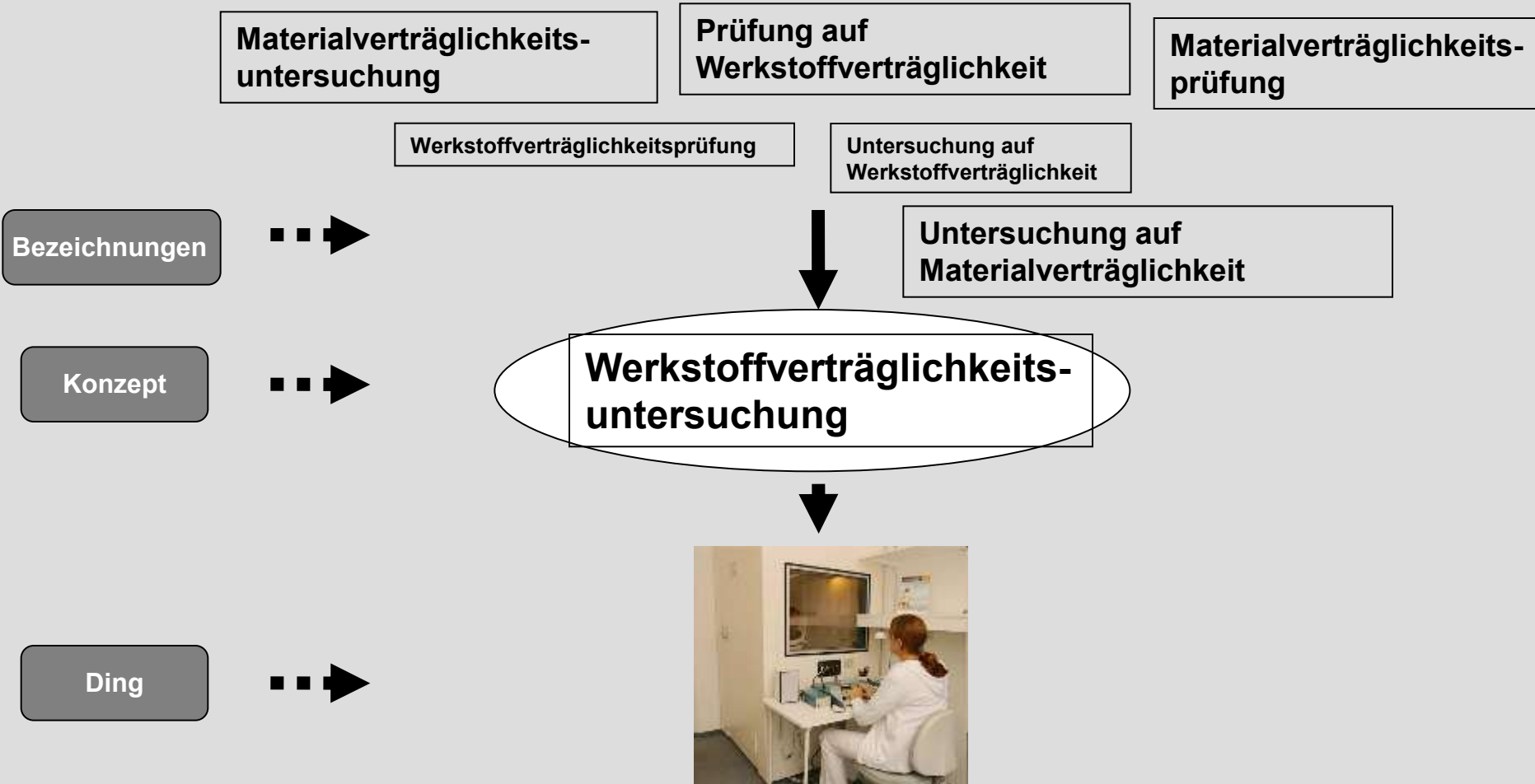
„Vorteile guter Terminologiewerkarbeit“:

- Bessere Verständlichkeit
- Reibungslose Kommunikation
- Besserer Wissenstransfer
- Weniger terminologische Brüche beim Wiederverwenden und ‚Rekombinieren‘ von Textbausteinen
- Weniger Rechtsstreit durch mehr Eindeutigkeit
- Corporate Language / corporate identity
- Wiedererkennungswert
- Sinkende Übersetzungskosten durch Einsatz von automatischer Übersetzung

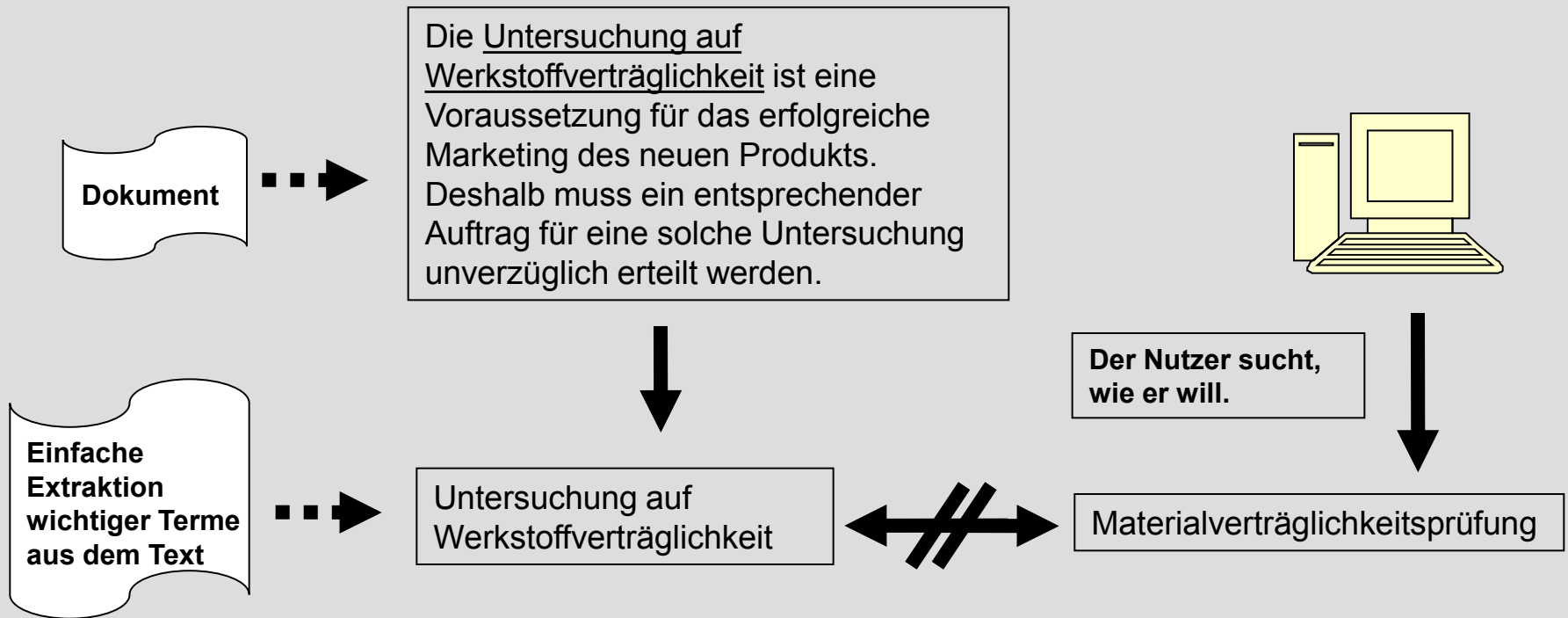
2. Thema: Informationsverarbeitung

- Information in Dokumenten will man wieder finden
- Inhaltserschließung / Informationssuche
 - AUTINDEX (Automatische Inhaltserschließung)
 - FINDauT (Informationssuche Lucene / Solr-Basis)

Problem: Terminologie

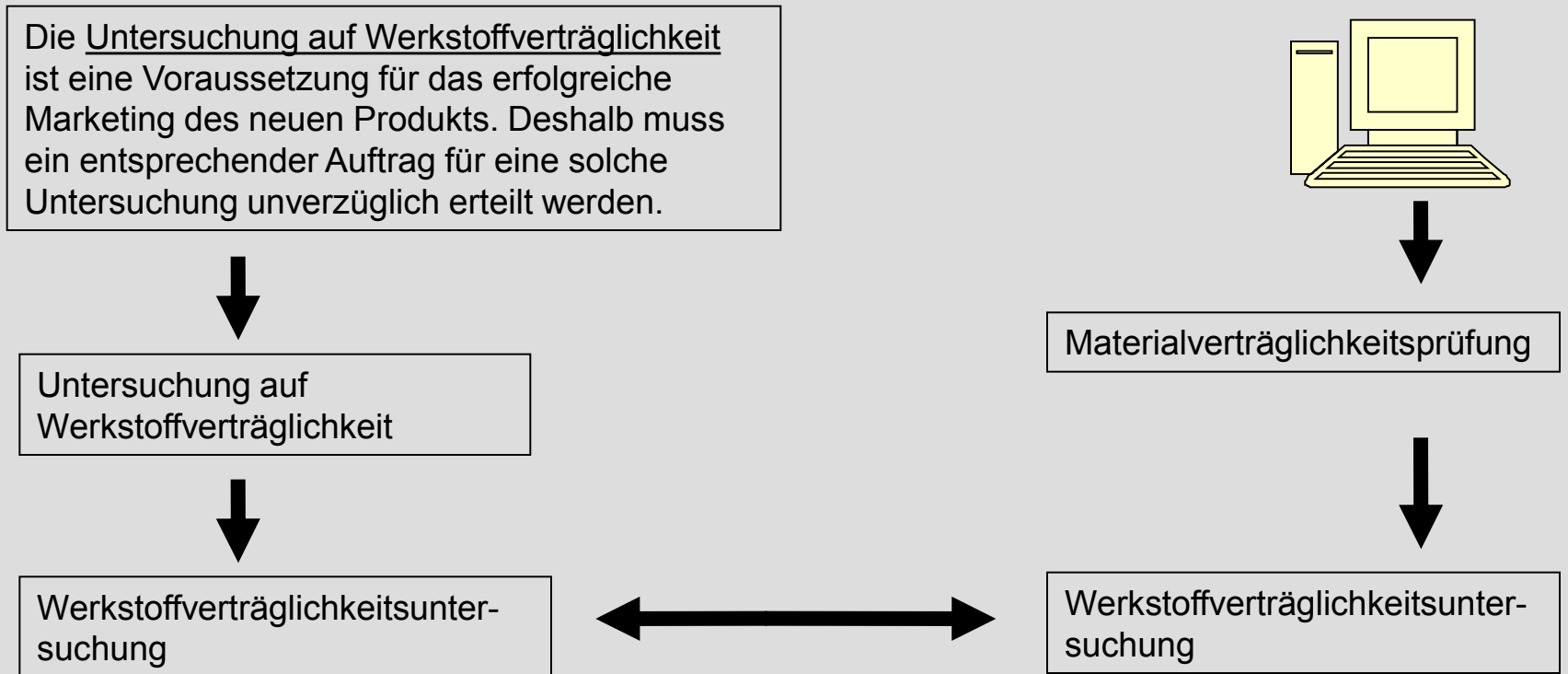


Terminologische Inkonsistenz und Informationssuche



Wie dann?

Standardisierung der Schlagwörter: Zurückführung von Termen auf die standardisierten Schlagwörter → Standardterminologie



- AUTINDEX: Automatische Inhaltserschließung durch Schlagwortvergabe mit Verarbeitung von Termvarianten
- FindAUT: Suche mit Termvarianten

- Bibliotheken:
 - Uni St. Gallen
 - Uni Darmstadt
 - Landesbibliothek Vorarlberg,
 - ...
- Banken-Dokumentmanagement-System:
 - Deutscher Sparkassen- und Giroverband

- Informationsanbieter:

- Juris, Saarbrücken
- Wolters Kluwer
- Datev
- ZPID (Zentrum für Psychologische Information und Dokumentation)
- GRS (Gesellschaft für Anlagen- und Reaktorsicherheit, Berlin)
- FIZ-Technik / WTI, Frankfurt
- Fa. Siemens

3. Thema: Dienstleistungen Terminologie

- Termextraktion / Terminologieaufbau
- Terminologiebereinigung / -standardisierung
- Translation Memory Bereinigung

- Terminologieaufbau
 - Terminologie fällt nicht vom Himmel, sondern muss aufgebaut werden: Aufwändiger Prozess
 - Meist ist eine große Zahl an Dokumenten vorhanden, aber keine Terminologie
 - Aus den Dokumenten kann die Terminologie extrahiert werden
 - ... dann gesäubert und standardisiert werden
 - ... mehrsprachig erzeugt werden

- Beispiel Termextraktion für Elektromobilität
 - Automatische Ermittlung von Termkandidaten aus Dokumenten
 - (Halb-)automatischer Aufbau der Terminologie
 - Techniken aus der Computerlinguistik:
 - Linguistische Eigenschaften von Termen (*Defibrillator, Doppelscheinwerfer, Wankstabilisierung, passiver Gasdruck-Stoßdämpfer, Automatikgetriebe mit fünf Fahrstufen*)
 - Statistische Verteilungen über große Korpora
 - Statistisch ermittelte Kohäsion

Brennstoffzelle

- HT-Brennstoffzelle
- NT-Brennstoffzelle
- PE-Brennstoffzelle
- SOFC-Brennstoffzelle
- Feststoffoxid-Brennstoffzelle
- wasserstoffbetriebene Brennstoffzelle
- Wasserstoff-Brennstoffzelle
- Schmelzcarbonat-Brennstoffzelle
- Klein-Brennstoffzelle
- Protonenaustauschmembran-Brennstoffzelle
- mobile PEM-Brennstoffzelle
- Carbonatschmelzen-Brennstoffzelle
- Phosphorsäure-Brennstoffzelle
- Schmelzkarbonat-Brennstoffzelle
- Protonenaustausch-Membran-Brennstoffzelle
- PEM-Brennstoffzelle

Brennstoffzelle

- Brennstoffzellen-Antriebssystem
- Brennstoffzellen-APU
- Brennstoffzellen-BHKW
- Brennstoffzellen-Blockheizkraftwerk
- Brennstoffzellen-Elektrofahrzeug
- Brennstoffzellen-Gerät
- Brennstoffzellen-Heizanlage
- Brennstoffzellen-Heizgerät
- Brennstoffzellen-Heizkraftwerk
- Brennstoffzellen-Heizsystem
- Brennstoffzellen-Hybridsystem
- Brennstoffzellen-Komponente
- Brennstoffzellen-Kraftwerk
- Brennstoffzellenabgas
- Brennstoffzellenabwärme
- Brennstoffzellenaggregat
-
- 72 weitere

Zusammenfassung

- Es gibt heutzutage hochentwickelte fertige Sprachtechnologie, die viele Probleme bei der Informationsverarbeitung, bei der Texterstellung, beim Dokumentenmanagement lösen kann
- Das gilt für die Verarbeitung in der Ausgangssprache wie für mehrsprachige Szenarien
- Kernproblem Standardterminologie
- Das IAI hat viele Techniken und Ressourcen, die praktische Probleme lösen können
- IAI ist auch an der Entwicklung von Innovationen gemeinsam mit Ihnen interessiert, die man z.B. in öffentlich geförderten Projekten entwickeln und in die Anwendung bringen kann.

Technologien können besichtigt werden