

Exzellenz in der zielgruppenoptimierten Kundenkommunikation



Dipl.-Ing. José Manuel de la Rosa Govantes

Vortrag bei der GSE-Tagung Hamburg, 27./28.09.23

<e>



Über uns



Eintritt 2004

Dipl.-Ing. José de la Rosa
(Fusion mit VisualAix®)



T

Kernthema bei Gründung

Verbindung von Forschung und
Praxis zur Textverständlichkeit
und Texttechnologie



Kooperationen

Mit der Wissenschaft und
Forschung

Mit Integrationspartnern

Mit Anbietern von CCM- und
OMS-Lösungen



Gründung 2000

als Spin-off der RWTH Aachen
(Prof. Dr. Christian Stetter und
Kay Heiligenhaus M.A.)



Interdisziplinäres Team

aus ca. 43 Informatikern,
Sprachwissenschaftlern,
Psychologen und Consultants
Enge Kooperation mit der RWTH
Aachen, Prof. Dr. Thomas Niehr

T₂

Zweites Thema (seit 2005):

Digitalisierungs- und
Recherchesysteme für
Bibliotheken, Archive und Museen
Marktführer D-A-CH

Methoden und Technik



Spezifikation, Redaktion,
Kollaboration, Dokumentation,
Recherche, Analyse, automatisierte
Verarbeitung, Migration



Schriftgut-Analysen,
Sprachrichtlinien,
Textoptimierung,
Sprachschulungen,
Redaktion



Forschung

Empirische Studien,
Korpuslinguistik,
Linguistische Informatik,
Lehre und Ausbildung





Intrinsische Eigenschaften von Text

Beispiel aus der Praxis



„Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er wegen eines während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.“

Quelle: „Zur Verständlichkeit von Geschäftstexten am Beispiel Allgemeiner Versicherungsbedingungen“, Magisterarbeit Björn Eßer, RWTH Aachen, 2012; §1 Nr. 1 AHB, zit. nach Nitschke (2002), S.122.



Eins

Das systemische Problem

Zahlen und Situation eines Großkunden (Juli 2017)



Redaktion:
202 aktiv
816 gesamt



Sachbearbeitung:
> 13.000



Textbestand:
71.048 Vorlagen
69.310 Bausteine
18.134 Variablen
26.000 Aktivitäten / Monat



? Textverantwortung:
Vorstand,
Marktmanagement,
Produktgeber,
Fachabteilung, ...



Außendienst:
> 100.000



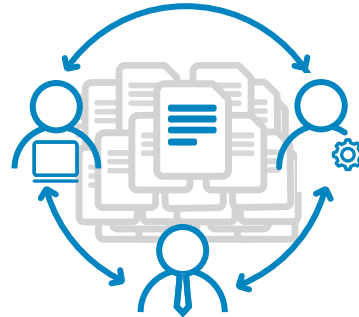
Endkunden





Schriftliche Kommunikation

Einige Dimensionen



Organisation

Kompetenzgemeinschaften
(fachlich, juristisch, ...)

Wissensmanagement,
Kollaboration, Redaktion,
Qualitätssicherung



Text

Interne Eigenschaften:
Wort-, Satz- u. Textebene

Text-, Korpus-,
strukturelle Linguistik
Verständlichkeits-/
Lesbarkeitsforschung

Sprachprüfung,
Empirische Studien

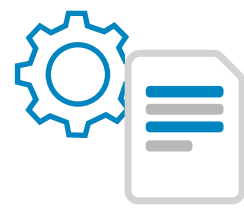


Empfänger

Zielgruppe, Motivation, ...

Kognitionswissenschaft,
Neurolinguistik, ...
Sprechakttheorie

Unterstützung
Mentales Modell



Technik

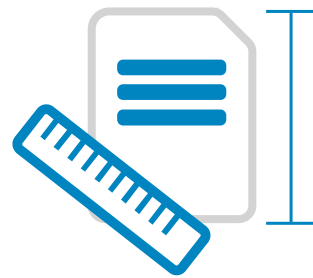
Paradigmen beeinflussen
Sprachqualität



Omni-Channel

Brief, E-Mail, SMS, App,
Internet, Chat, ...

Responsive Documents,
Variantentexte





Sprache

Linguistische Analyse

Hauptaussage

konditionaler
Nebensatz

kausale
Bestimmung

Linksattribut

Relativsatz

Rückbezug

Verweis auf
rechtliche Basis

Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er wegen eines während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.

„Der Versicherer gewährt dem Versicherungsnehmer Versicherungsschutz für den Fall, dass er wegen eines während der Wirksamkeit der Versicherung eingetretenen Schadenereignisses, das den Tod, die Verletzung oder Gesundheitsschädigung von Menschen (Personenschaden) oder die Beschädigung oder Vernichtung von Sachen (Sachschaden) zur Folge hatte, für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen privatrechtlichen Inhalts von einem Dritten auf Schadenersatz in Anspruch genommen wird.“



Sprache

Überarbeitung

Der Versicherer gewährt dem Versicherungsnehmer Deckung für Haftpflichtansprüche, die gegen ihn geltend gemacht werden.

Haftpflichtansprüche unterliegen gesetzlichen Haftpflichtbestimmungen privatrechtlichen Inhalts.

Sie bestehen dann, wenn der Versicherungsnehmer für Personen- oder Sachschäden verantwortlich gemacht werden kann.

Dieser Schaden muss während der Versicherungszeit entstanden sein, damit er unter die gewährte Deckung fällt.

Hauptaussage

**Verweis auf
rechtliche Basis**

**Kausale
Bestimmung**

Bedingung

„Der Versicherer gewährt dem Versicherungsnehmer während der Wirksamkeit der Versicherung einen Versicherungsschutz für den Schaden, der während der Versicherungsdauer durch einen eingetretenen Schadenereignis, eine Verletzung oder Gesundheitsschaden (Personenschaden) oder die Vernichtung von Sachen (Sachschaden) für diese Folge aufgrund gesetzlicher Haftpflichtbestimmungen entstanden ist.“



Sprache

Thematische Kohärenz zur
Unterstützung mentales Modell
(Thema/Rhema Paradigma,
anaphorische Bezüge,
Rekurrenzen)

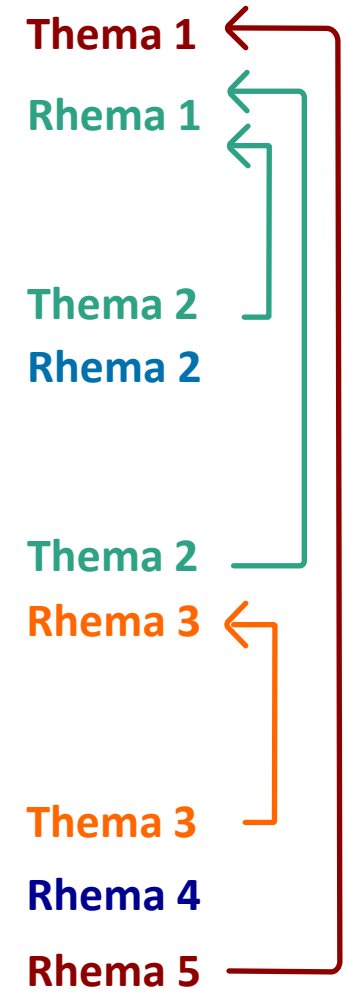
„Der Versicherer gewährt a
Versicherungsschutz für de
während der Wirksamkeit d
eingetretenen Schadenere
Verletzung oder Gesundheit
(Personenschaden) oder die
Vernichtung von Sachen (S
für diese Folge aufgrund ge
Haftpflichtbestimmungen r

Der *Versicherer* gewährt dem Versicherungsnehmer *Deckung*
für *Haftpflichtansprüche*, die gegen ihn geltend
gemacht werden.

Haftpflichtansprüche unterliegen gesetzlichen
Haftpflichtbestimmungen privatrechtlichen Inhalts.

Sie bestehen dann, wenn der Versicherungsnehmer
für *Personen- oder Sachschäden* verantwortlich
gemacht werden kann.

Dieser Schaden muss während der Versicherungszeit
entstanden sein, damit er unter die gewährte *Deckung* fällt.





Ein Extrem: Leichte Sprache

Punktuell sinnvoll
(nicht negativ konnotiert,
Nutzen für gewisse Zielgruppen)

<http://www.leichtesprache.org>
<http://www.leichte-sprache.de>



Diese Geschichte geht um eine neue Sprache. Die Sprache heißt Leichte Sprache. Sie ist für jeden leicht zu verstehen. Sie hat ganz leichte Regeln. Man schreibt nur in Hauptsätzen. Man wiederholt Wörter immer wieder. Man drückt sich nur in deutschen Wörtern aus. Wörter aus mehreren Wörtern werden mit Binde-Strichen getrennt. Es heißt zum Beispiel Sommer-Urlaub und Fabrik-Schornstein. Man verwendet auch möglichst den Wem-Fall. Der Wes-Fall ist zu schwierig.

Kluge Leute haben sich die Leichte Sprache für Menschen mit Lern-Schwierigkeiten ausgedacht. Aber auch andere Menschen können Leichte Sprache verwenden. Es gibt jetzt Bücher in Leichter Sprache.

Rec
Kind
Zum
Kind
Dann

(Martin Doerry, DER SPIEGEL 29 / 2016)

Kognitionswissenschaftliche Bedenken

„Das Ideal ist aber keine Aneinanderreihung einfacher Sätzchen, denn diese erfordern wieder Ressourcen, um die inhaltliche Beziehung zwischen ihnen zu erfassen.“

(Ballstaedt, 2019, S. 220)

Keine Lösung z.B. für Versicherungsunternehmen

- repräsentative Funktion
- zielorientierte Kommunikation
- auch durch Ästhetik und Verbindlichkeit der Sprache



Lesbarkeitsindexe

Messen von Längen, Vergleichen von Listen; Oberflächenstrukturen

S = Anzahl Sätze, W = Anzahl Wörter, WE = einsilbige Wörter, $W3$ = mind. 3 Silben,
 SL = mittlere Satzlänge, WL = mittlere Wortlänge, $W6$ = mind. 6 Buchstaben

Amstad-Formel („Reading Ease“)

$$RE = 180 - SL - 58,5 * WL$$

Lix („Lesbarkeitsindex“)

$$Lix = \frac{W}{S} + \left(\frac{W6}{W} \right) * 100$$

SMOG („Simple Measure of Gobbledygook“)

(Variante für Deutsch GSMOG vorhanden)

$$SMOG = 1,043 * \sqrt{30 * \left(\frac{W3}{S} \right)} + 3,1291$$

Erste neue Wiener Sachtext-Formel

$$WSTF = 0,1935 * W3\% + 0,1672 * SL + 0,1297 * W6\% - 0,0327 * WE\% - 0,875$$

Hohenheimer Verständlichkeitsindex

Formeln 1-4, zusätzlich eingerechnet:

- Mittlere Satzlänge und Satzteilänge in Wörtern
- Anteil der Satzteile mit mehr als zwölf Wörtern
- Anteil der Sätze mit mehr als zwanzig Wörtern
- (sehr kleine) korpusbasierte Ansätze, 1.000 häufigsten

Wortformen TIGER Korpus

- ...

(Exakte Formel nicht offengelegt)

- Zahlreiche weitere auf Oberflächenstrukturen
- Interessante Ansätze wie Type-Token-Ratio
- Abstraktheits-Suffix-Verfahren nach Günther und Groeben (1978) *(Suchen nach -heit, -ie, -ik, -ion, -ismus, -ität, -keit, -nz, -tur, -ung)*

Betrachten wir einfache Beispiele

Indexe können Kohärenzen nicht messen; andere (positive) Ergebnisse evozieren zumindest Diskussionsbedarf...

Nachts ist es kälter als draußen.

(Rhetorischer Alogismus, Stilmittel z.B. Darstellung unlogischer Sachverhalte)

Die Frau küsst der Mann.

(Änderung der Satzhypothese während des Lesens)

**Der Ober brachte dem Gast den Teller.
Er war schmutzig.**

(Referenzproblem)

Nicht dieser Mann war der Täter.

(Verständnis von Negationen braucht Zeit)

Ich sah Kraniche auf dem Flug nach New York.

(Syntaktisch mehrdeutig)

Quelle 2-4: Vorlesung Textverständlichkeit, Prof. Dr. Eva Jakobs (RWTH Aachen)

Quelle 5: Vorlesung Künstliche Intelligenz, Prof. Dr. Katharina Morik (Uni Dortmund)

Die Welt ist alles, was der Fall ist.

**Was der Fall ist, die Tatsache, ist
das Bestehen von Sachverhalten.**

**Das logische Bild der Tatsachen ist
der Gedanke.**

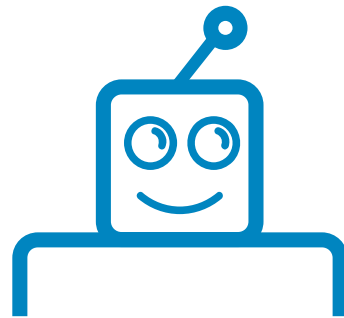
Der Gedanke ist der sinnvolle Satz.

**Der Satz ist eine Wahrheitsfunktion
der Elementarsätze.**

**(Der Elementarsatz ist eine
Wahrheitsfunktion seiner selbst.)**

(Wittgenstein, 1999)

*Die Vielfalt der Sprache und ihrer Stilmittel ist unendlich und
kann nicht (sinnvoll oder vollständig) „nachprogrammiert“
werden.*

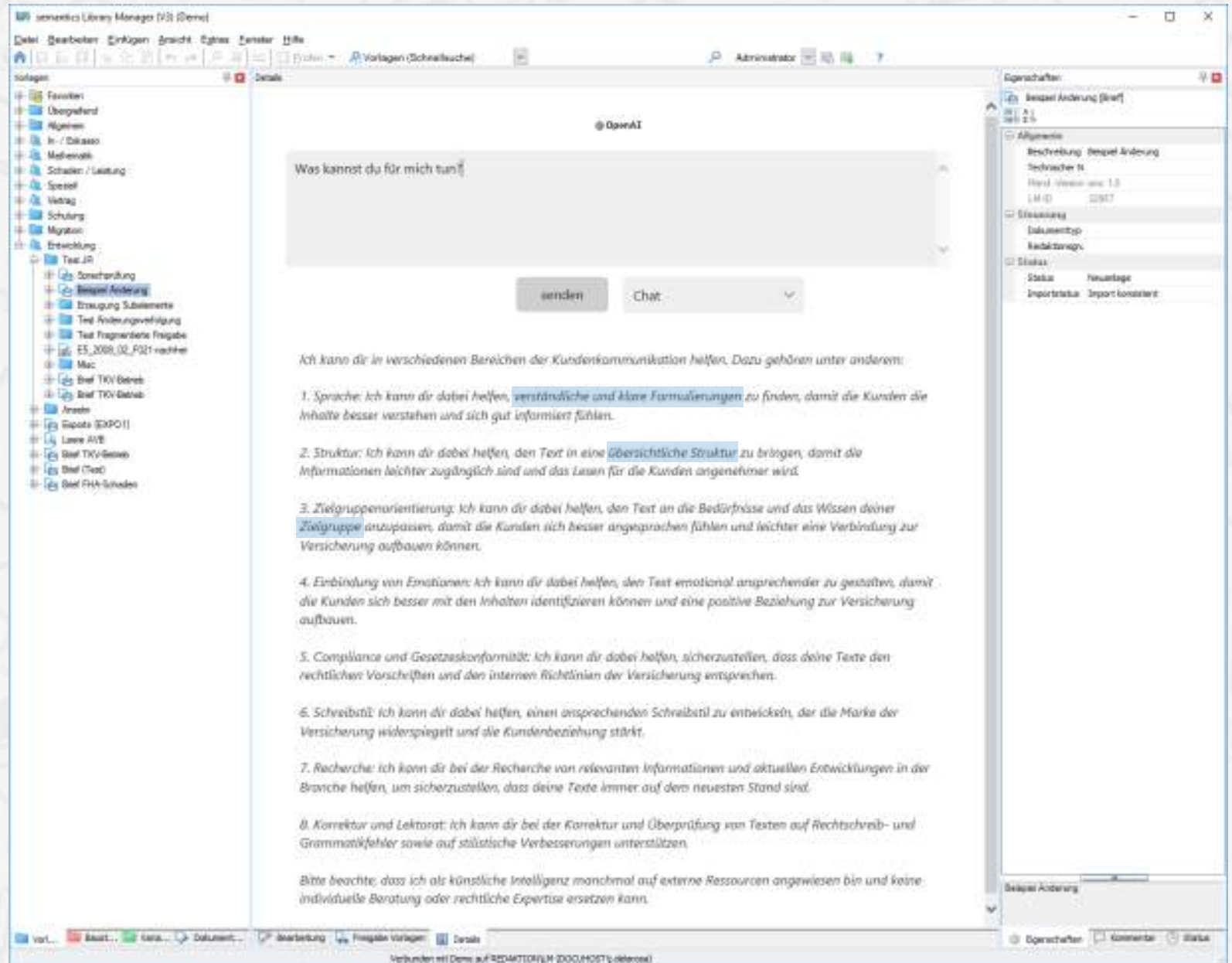




Hallo ChatGPT,

was kannst du für mich tun?

Integration beliebiger KI in
Library Manager





J



...not bad!

J Was bedeutet Textverständlichkeit?

Textverständlichkeit bezieht sich auf die Fähigkeit eines Lesers, von Lesern leicht verstanden und interpretiert zu werden. Es bezieht sich auf die Klarheit, Konsistenz und Struktur eines Textes, die es den Lesern ermöglichen, den Inhalt effektiv aufzunehmen und zu verstehen. Texte mit hoher Verständlichkeit sind gut strukturiert, verwenden klare und präzise Sprache, vermeiden komplizierte Satzkonstruktionen und Fachjargon und passen sich dem Leserniveau an. Ein gut verständlicher Text erleichtert es den Lesern, die Informationen zu verarbeiten und die beabsichtigte Botschaft zu entschlüsseln.

ChatGPT liegt nicht ganz falsch... aber...

Verständlichkeit ist eine Relation zwischen Text und lesender Person.

Unser Ziel: **Angemessenheit** für Zielgruppe; in Wortwahl, Umfang, Struktur, Diktion, Illokution, ...

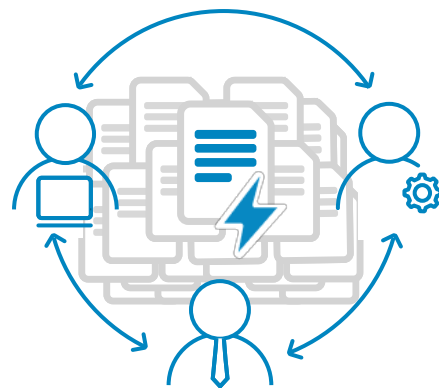
Versicherungen verkaufen ein Versprechen.

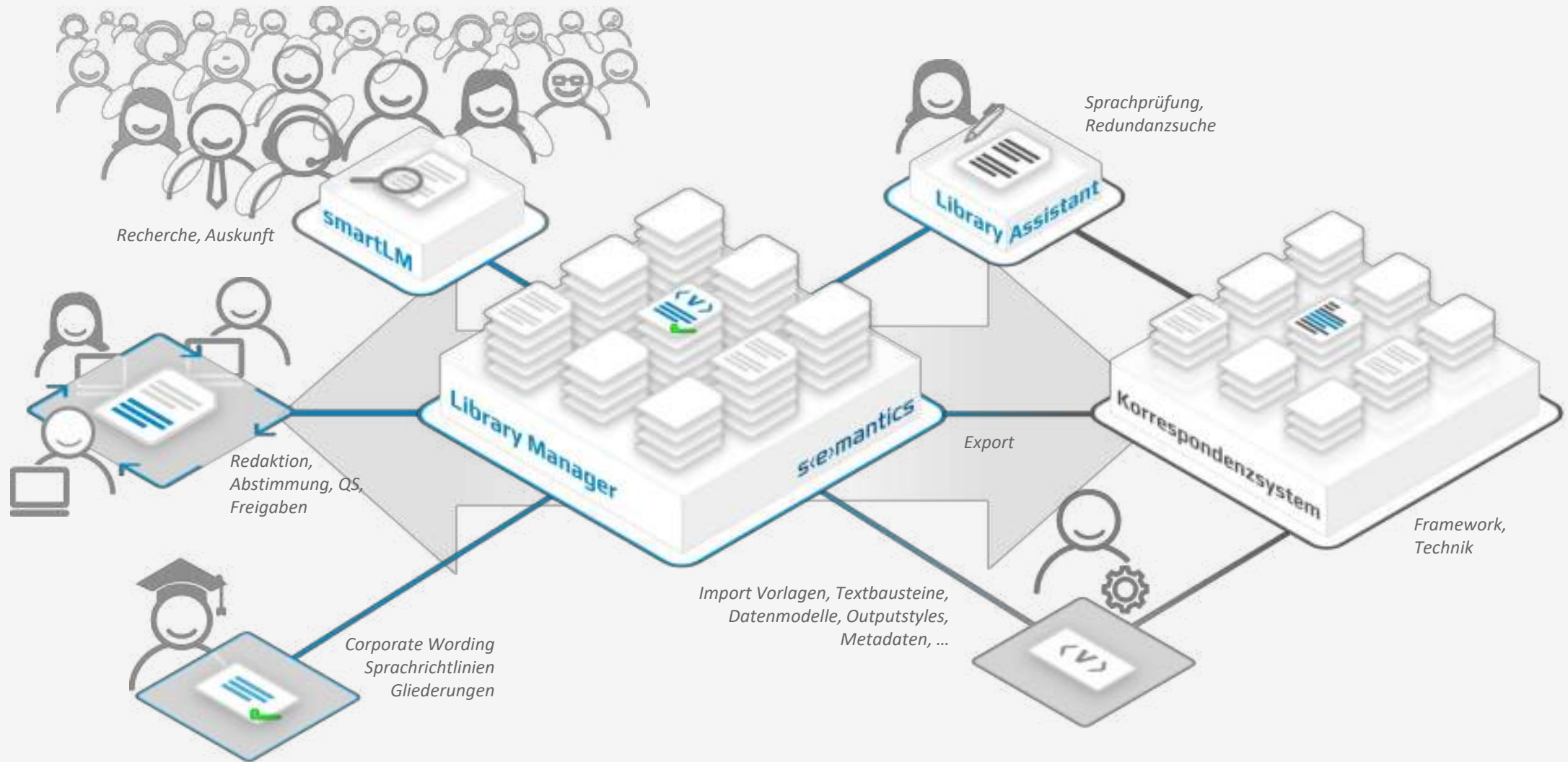
J Bitte schreibe eine weitere Variante, die insbesondere junge Menschen ansprechen könnte.

 Natürlich! Hier ist eine Variante, die speziell auf junge Menschen abzielt:

Der Versicherer sorgt dafür, dass du geschützt bist. Wenn während der Zeit, in der du versichert bist, etwas Schlimmes passiert, wie zum Beispiel ein Unfall, bei dem jemand verletzt wird oder Dinge kaputtgehen, dann übernimmt der Versicherer die Kosten. Du musst dir keine Sorgen machen, dass du für solche Schäden selbst aufkommen musst. Die Versicherung ist da, um dich abzusichern und dir den Rücken freizuhalten, falls du von jemand anderem wegen des Schadens verklagt wirst. Du kannst also unbesorgt sein und dich auf deine Aktivitäten konzentrieren, während die Versicherung sich um die finanzielle Seite kümmert.

...wow! 





Corporate Library schließt die Lücke zwischen Fachbereich und CCM / OMS.

Library Manager

Medienbruchfreie Workflows
von Spezifikation über
Redaktion bis Produktion



Kollaboration

Workflows ohne Medienbrüche,
Staging, Import / Export,
Statistik, Fertigstellungsgrade,
Plausi-Prüfung, Testabdeckung,
formatierte Vorschau



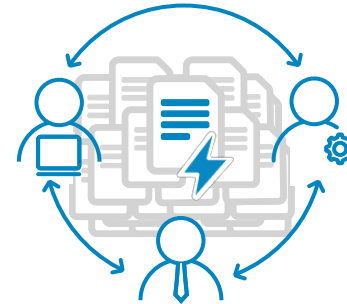
Sprachqualität

Corporate Wording,
Struktur / Gliederung,
Verständlichkeit



Dokumentation

Versionen,
Verwendungen,
Querbeziehungen,
Annotationen



Library Manager



Organisation

Rechte / Rollen, Bibliothek,
Suche, Recherche,
intelligente Massen-Updates



Redundanzsuche

N-Gramme,
Ermittlung von Abdeckungen
und Paraphrasierungen



Varianten vs. Bausteine

Unterstützung
thematische Kohärenz,
Mentales Modell,
Variantentests

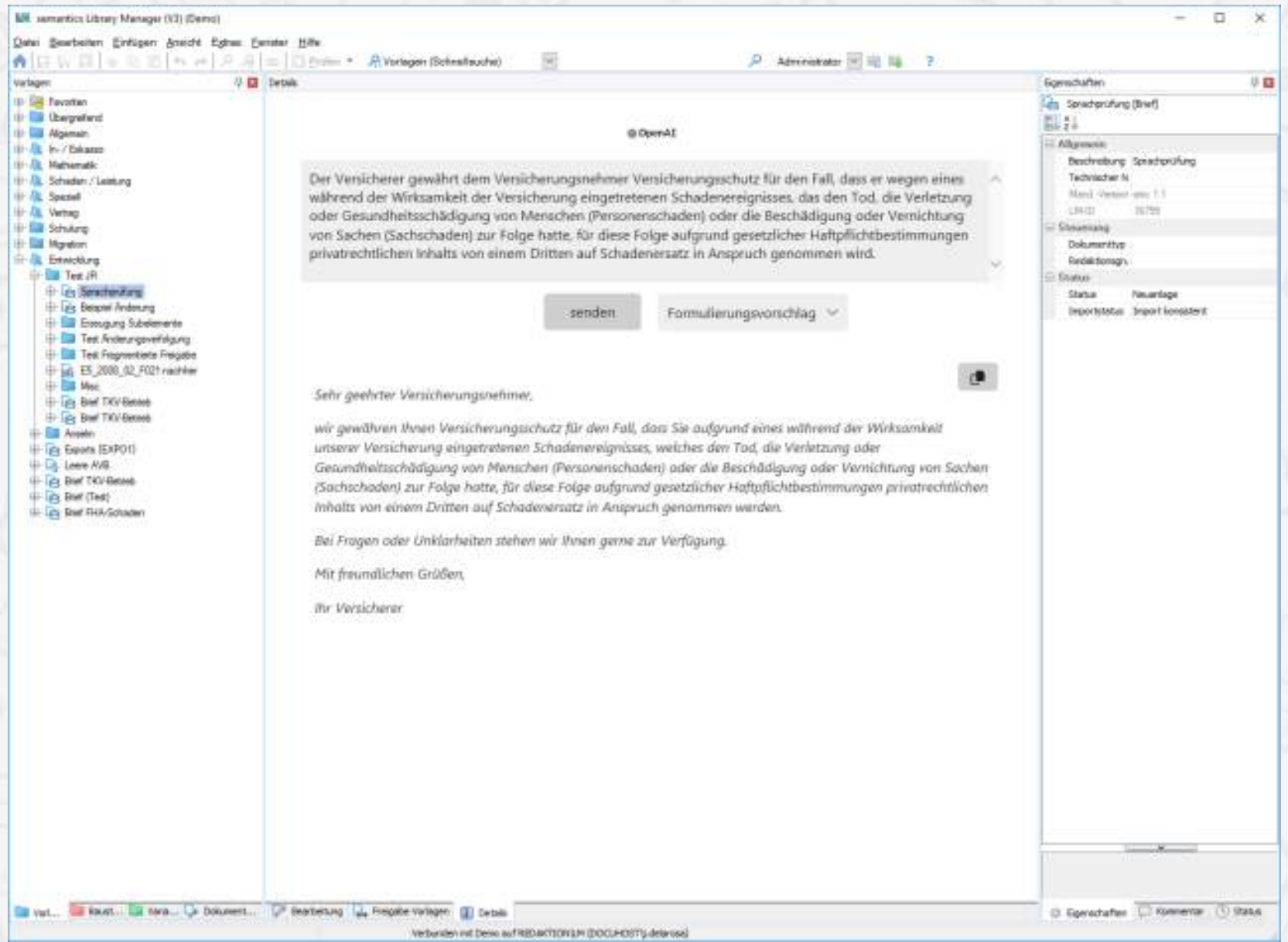


Integration KI

Aufnahme in Workflows,
Erzeugung Formate, ...

Vollständige Einbettung

Prompt Engineering

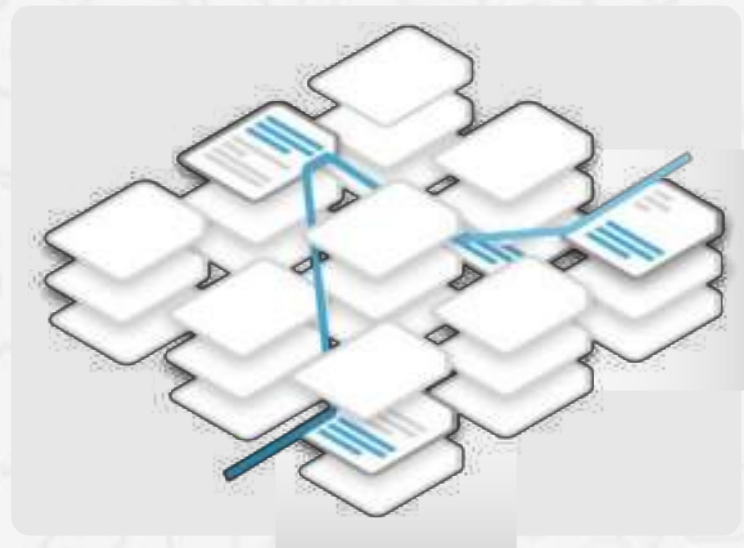




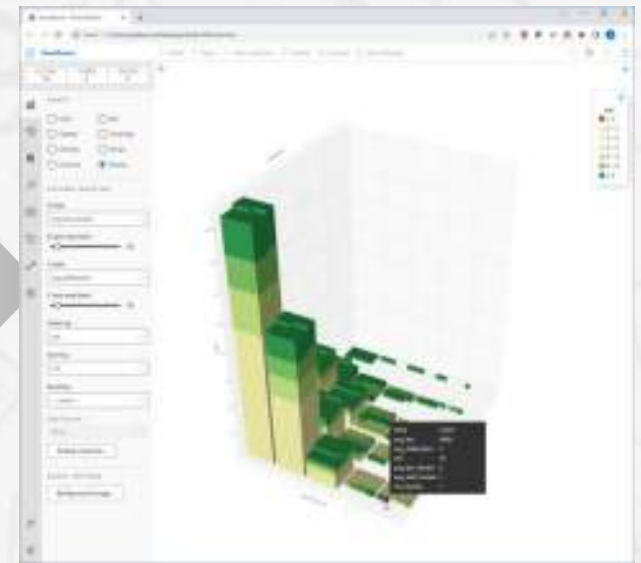
Verarbeitung KI

Maschinelle Sprachprüfung

Anwendung in der Breite
*Redundanzen anhand
lemmatisierter N-Gramm-Suche,
automatische Konsolidierung*



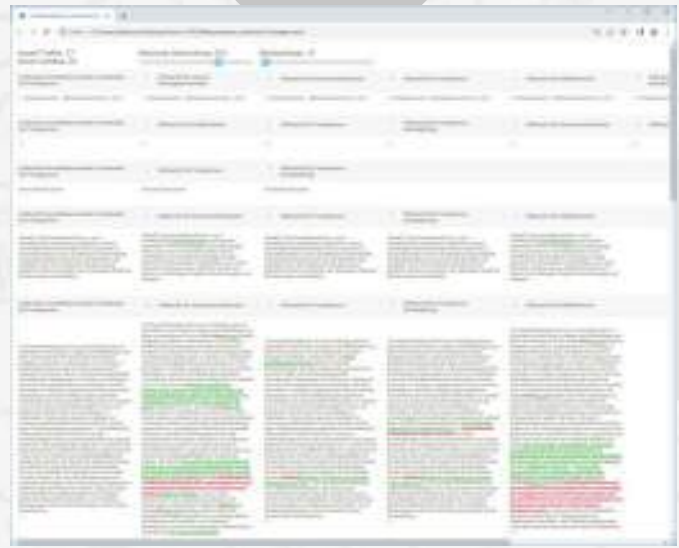
Redundanzsuche



Statistik über Ähnlichkeiten



Sprachprüfung



Detaillierte Aufstellung, Steuerung Konsolidierung



Ergebnis Abdeckung



Weitere Anwendung KI

Vielfältige Möglichkeiten

Hier:

Erläuterung Unterschiede zwischen Versionen

Auffällige Qualitätsunterschiede je Aufruf
(Ergebnis zweiter Aufruf schlicht falsch)

Die beiden Texte sind fast identisch und unterscheiden sich nur geringfügig in einem einzelnen HTML-Attribut.

In Text 1 wurde das Attribut ``xml:space="preserve"``` verwendet, während in Text 2 dieses Attribut fehlt.

Das Attribut ``xml:space="preserve"``` bewirkt, dass alle Leerzeichen und Zeilenumbrüche im Text beibehalten werden. Ohne dieses Attribut würde whitespace, also Leerzeichen und Zeilenumbrüche, normalerweise bei der Darstellung des Textes entfernt werden, um die Lesbarkeit zu verbessern.

Da es sich bei den beiden Texten ansonsten um den gleichen Inhalt handelt, hat das Fehlen des Attributs ``xml:space="preserve"``` in Text 2 keine Auswirkungen auf die Bedeutung des Textes. Es könnte jedoch zu geringfügigen optischen Unterschieden führen, wenn der Text ohne dieses Attribut angezeigt wird.

The screenshot shows the 'semantics Library Manager (V3) (Demo)' application. The main window displays a list of documents on the left and a comparison view on the right. The comparison view shows two versions of a document, with a table of differences below. The table has columns for 'Datum / Uhrzeit', 'Benutzer', 'Status', 'Info', 'Ticket', 'Kommentar', and 'Vergleichen'. The table shows three entries for 'Beispiel Änderung (unc 1.0)' with a status of 'geändert'.

Below the table, there is a section titled 'Unterschiede @ OpenAI' which contains the following text:

Die beiden Texte sind nahezu identisch. Es gibt nur eine kleine Veränderung:

In Text 1 lautet der Satz "wir möchten Sie darüber in Kenntnis setzen", während es in Text 2 "wir möchten Sie darüber informieren" heißt.

Die Bedeutung der beiden Sätze ist jedoch gleich. Es handelt sich also um einen reinen Wortwechsel ohne inhaltliche Auswirkungen.

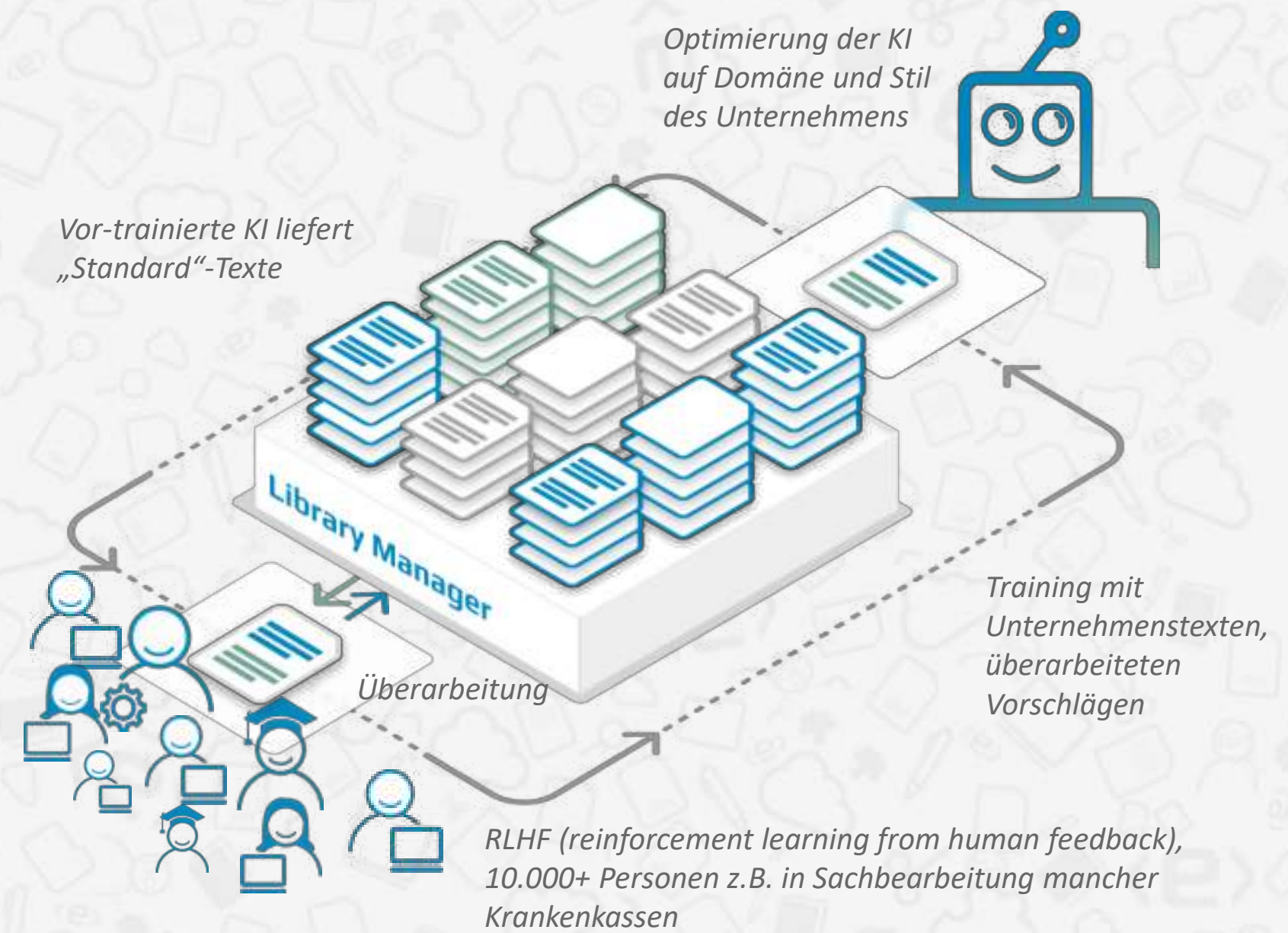
The interface also includes a sidebar with 'Eigenschaften' (Properties) and a 'Beispiel Änderung (Brief)' section.





Weiterentwicklung der KI

Integration Training und RLHF in Library Manager





Dipl.-Ing.
José Manuel de la Rosa Govantes

semantics
Kommunikationsmanagement
GmbH
Viktoriaallee 45
52066 Aachen

+49 241 89 49 89 29
j.delarosa@semantics.de
www.semantics.de



Besten Dank für Ihre Aufmerksamkeit!

