

**ORTEC**



**XML-Tage**

**ConWare**

**25.-26. September 2007**



### **ORTEC GmbH - Berlin**

Gründung 1997 als Spin-off der Technischen Universität Berlin

Schwerpunkte:

- Publishing Management
- Prozesssteuerung
- Content und Knowledge Management
- Dokumentenverwaltung

### **ORTEC Service GmbH - Berlin und Köln**

Gründung 2004

Schwerpunkte:

- Professional Service
- Prozessberatung
- IT Consulting
- Projektmanagement
- Training

→Über 30 Experten von ORTEC und weitere Partnern zu Ihrer professionellen Unterstützung!

# ORTEC

Professionelle Partner





Bibliographisches Institut & F.A.  
Brockhaus AG  
Bildungsverlag EINS  
edition text & kritik  
Ernst Klett Verlag GmbH  
Huber Verlag  
IDG Communications Media AG  
IDW Verlag GmbH  
Langenscheidt KG  
LexisNexis Deutschland  
Richard Boorberg Verlag  
Springer Medizin Verlag GmbH  
Stollfuß Verlag  
Thieme Verlag  
Verlag C.H.Beck oHG  
Verlag CW Haarfeld

Verlag Harri Deutsch - Electronic Publishing  
Wolters Kluwer  
BMJ Publishing Group  
Editions du Juris-Classeur  
Elsevier  
Linde Verlag Wien GmbH  
Manz Verlag GmbH  
McGraw-Hill  
Oxford University Press  
WEKA Verlag AG, Zürich  
Wiley Publishing

AOK, Techniker Krankenkasse, Barmer  
Bundesrechtsanwaltskammer



### Content and Publishing Management

- Wolters Kluwer Germany



### www.Persoline.de, Information system for different health insurances

- AOK, Barmer, Techniker Krankenkasse, Ford Krankenkasse
- CW Haarfeld, Germany



### Online Services and Publishing Management

- RS Schulz Verlag, Germany
- MANZ, Austria



### CMS für das kritische Lexikon der Gegenwartsliteratur

- Edition Text und Kritik München





## Integration aller Produktionsprozesse

- ☛ Zentrale Workflow-Steuerung
- ☛ Anbindung von Redaktion und Produktion
- ☛ Personalisierung und Zugriffsmanagement
- ☛ Umfassende Qualitätssicherung
- ☛ Mitarbeiter-Qualifikation anstatt Abhängigkeit von Dienstleistern



## Integration aller Produktionsprozesse Medien- und produktneutrale Datenhaltung

- Verlagsübergreifendes XML-Datenmodell
- Unabhängige Produktkonfigurationen
- Automatisierte Medien-Ausgaben
  - Print (Buch, Zeitschrift, Loseblatt),
  - Offline (CD-ROM)
  - Online (Portal)



Integration aller Produktionsprozesse

Medien- und produktneutrale Datenhaltung

Mehrfachverwendung der Inhalte

- ☛ Flexibilität über bausteinbasierte Produktdefinitionen
- ☛ Automatische Aktualisierung der Inhalte
- ☛ Kunden- oder zielgruppenspezifische Produktausprägungen
- ☛ Datenqualifizierung

ORTEC

ConWare Suite<sup>®</sup>





## ConAgent

- Content- und Metadatenverwaltung
- Strukturbildung
- Produktzusammenstellung



## ConBuilder

- Produktausgaben
- Auflagengenerierung

## ConGuard

- Konfiguration von IO-Klassen, Linkhandlern, XSL-Skripten
- Massenkonzertierung



## ConCovery

- Browser gestütztes CD-ROM Modul
- standalone Präsentation von Inhalten
- Maskeneditor & Indexgenerator für Suchen



## ConPortal

- Onlinemodul zur Darstellung von ConWare-Inhalten im WWW
- Volltextsuche
- Datenbankneutral

## Print Content Management

- EBL Generierung
- Änderungserkennung
- Ansteuerung von Satzsystemen



## ConTax

- Taxonomieverwaltung
- Contentzuordnung



## ConSaurus

- Thesaurusverwaltung
- Wortnetz mit Verknüpfungen
- Verwaltung von Relationen

## Customer Profile Management

- produktbezogene Benutzerverwaltung
- Authentifizierungssystem für die Integration in Webseiten



## DocumentStore

- Dokumentenverwaltung
- Ressourcenholder für andere ConWare-Clients
- Versionierung von Dokumenten

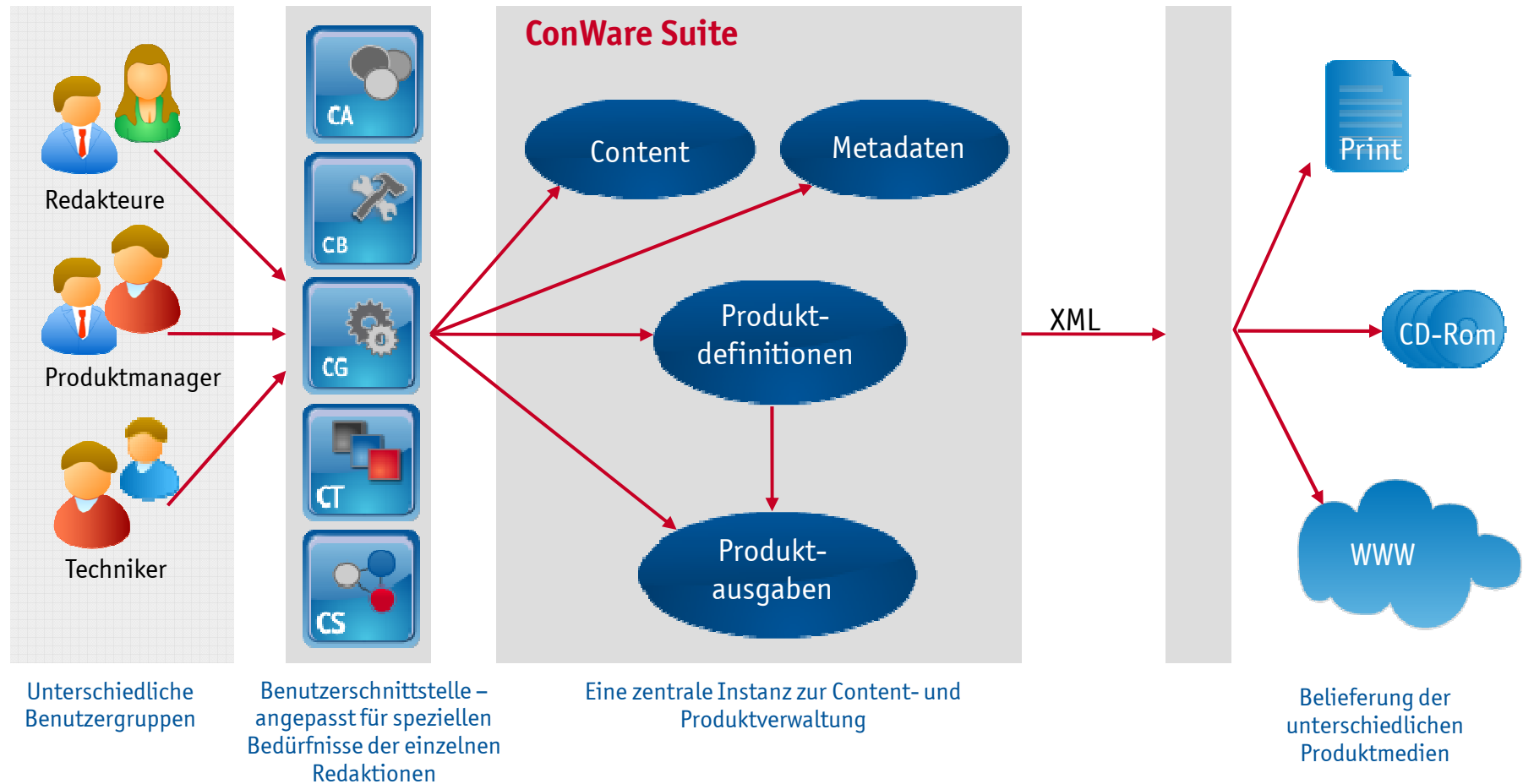


## Verwaltungsdatenbank

- optionale Erweiterung zur Erfassung weiterer Metadaten

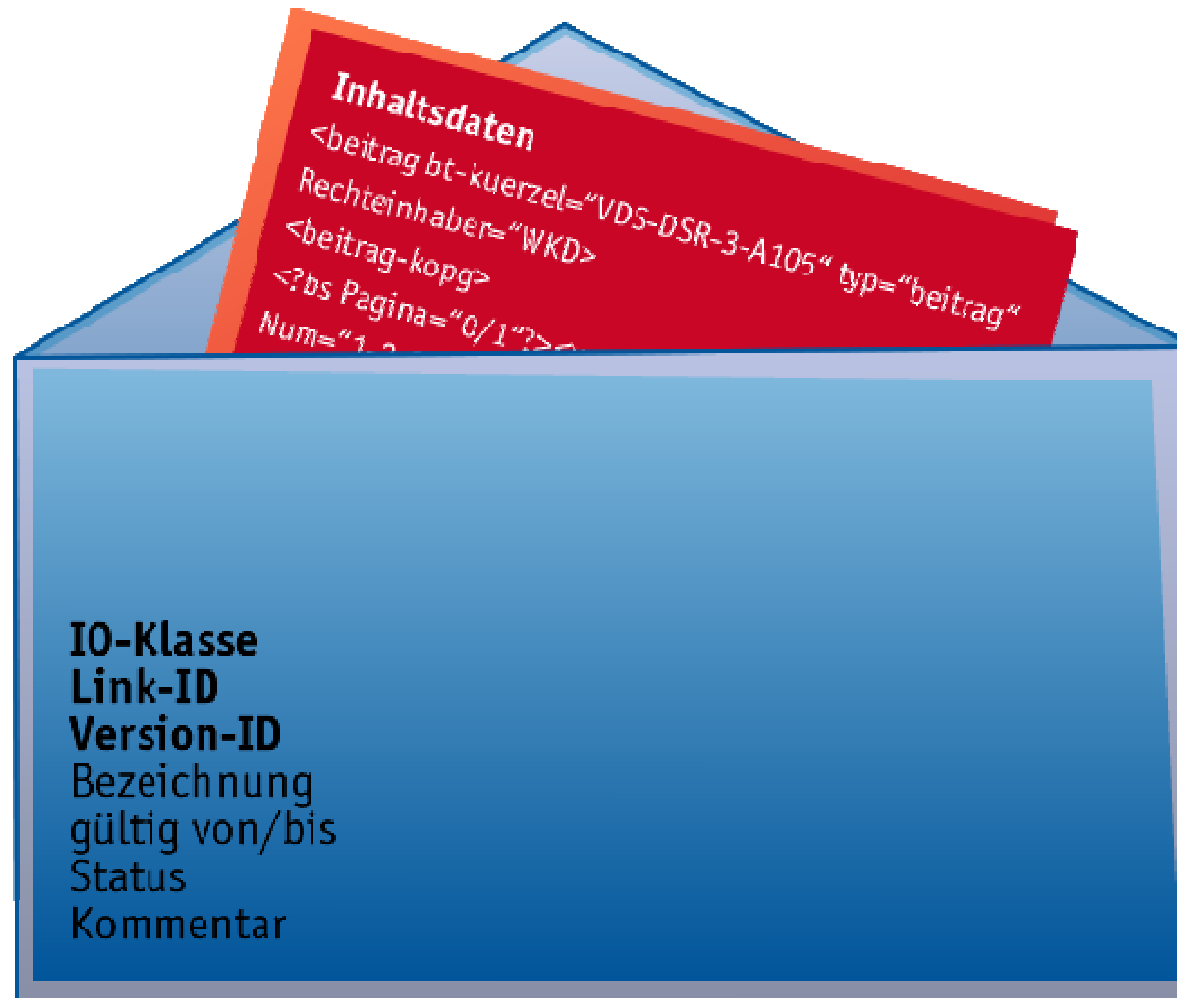
## BRAK - Client

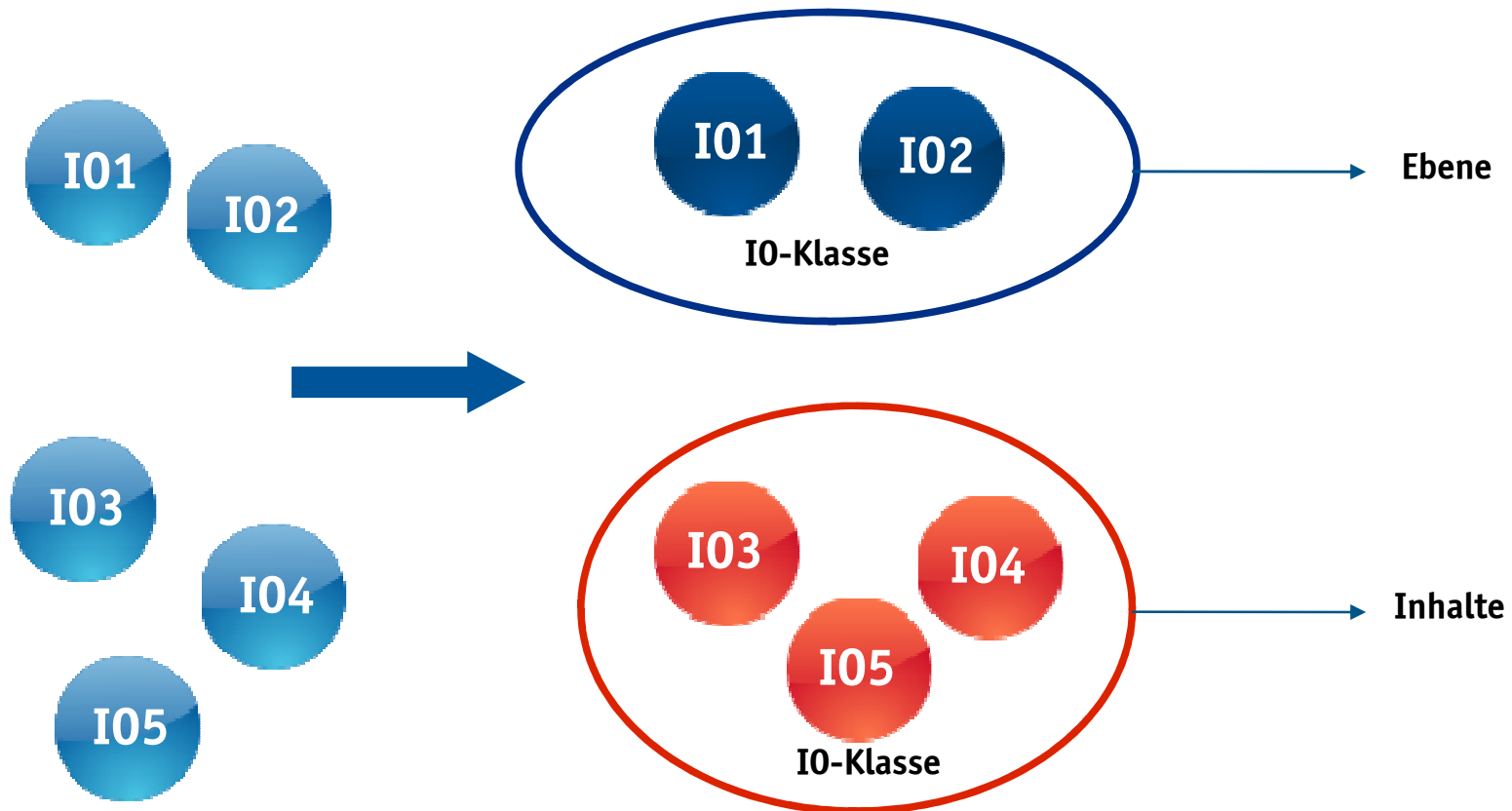
- spezielle Ausprägung des ConAgent für die Optimierung der Redaktionprozesse der BRAK - Redaktion

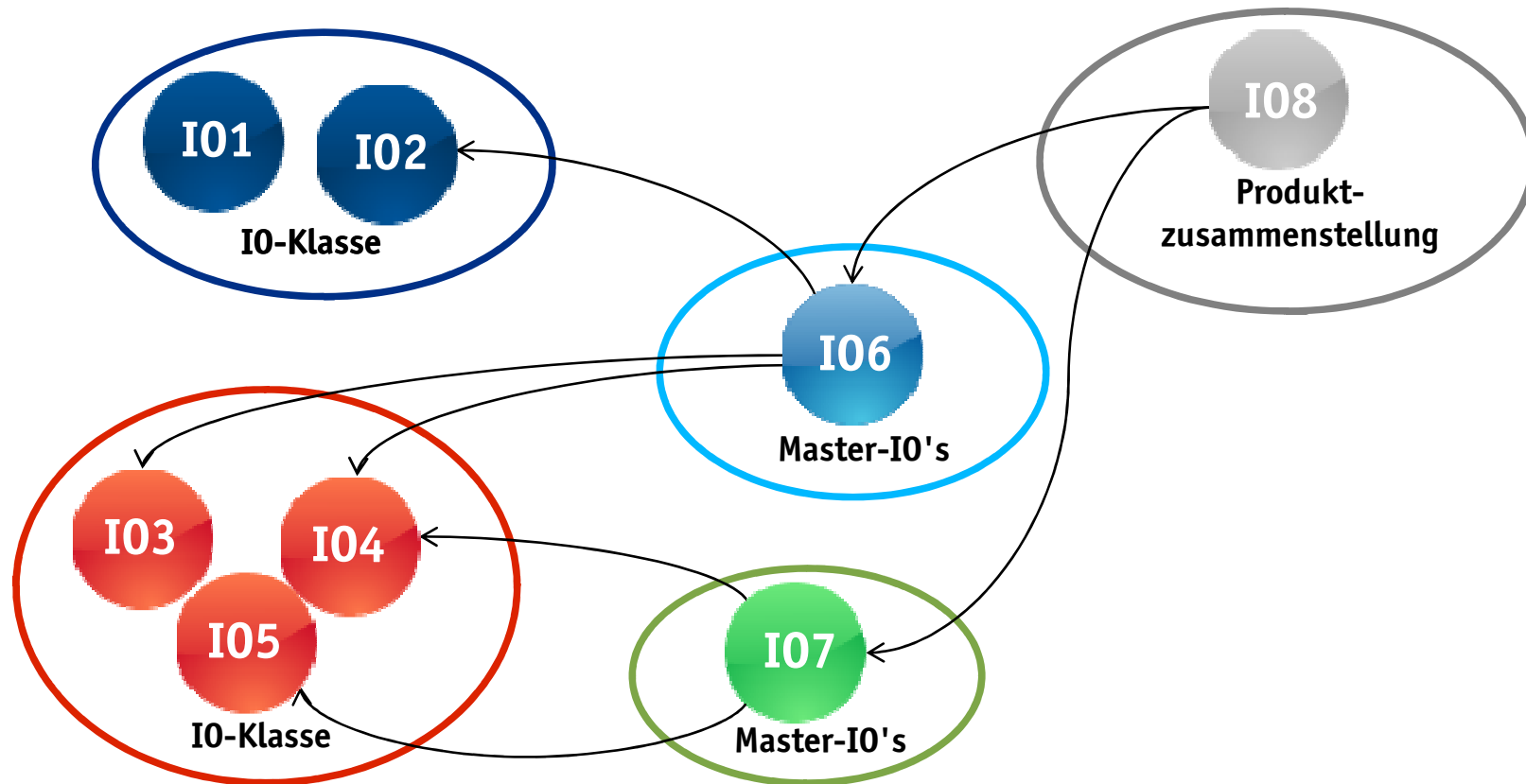




**IO-Klasse**  
**Link-ID**  
**Version-ID**  
Bezeichnung  
gültig von/bis  
Status  
Kommentar









## Wissensrepräsentation

## Wissensrepräsentation – Warum?

Durch die manuelle Erstellung von Strukturen durch „Master-IO´s“ bzw. „Produktzusammenstellungen“ wird typischerweise jeweils ein Aspekt bzw. Kontext für eine Publikation modelliert.

Um in der Zukunft eine automatisierte Gliederung bzw. Zusammenstellung des Content zu erreichen, müssen bestimmte Informationen aus dem Content extrahiert und in Wissensrepräsentationen vorgehalten werden.

Für den Anwender werden die Prozesse zunehmend inhaltlich statt bisher technisch geführt sein, womit die Benutzerfreundlichkeit stark gesteigert wird.

Somit lässt sich der Produktionsablauf weiter automatisieren und wir kommen unserem Ziel der „On-Click-Production“ etwas näher.



## Wissensrepräsentation – Wie?

Der Content bzw. ein Dokument wird in Absätze, Sätze, Worte, Werte und Phrasen zerlegt.

Durch die Worte wird eine Wortsammlung erstellt, die Werte bilden eine Wertesammlung, etc.

Zwischen Worten und Dokumenten wird eine Relation erzeugt (Indexierung).

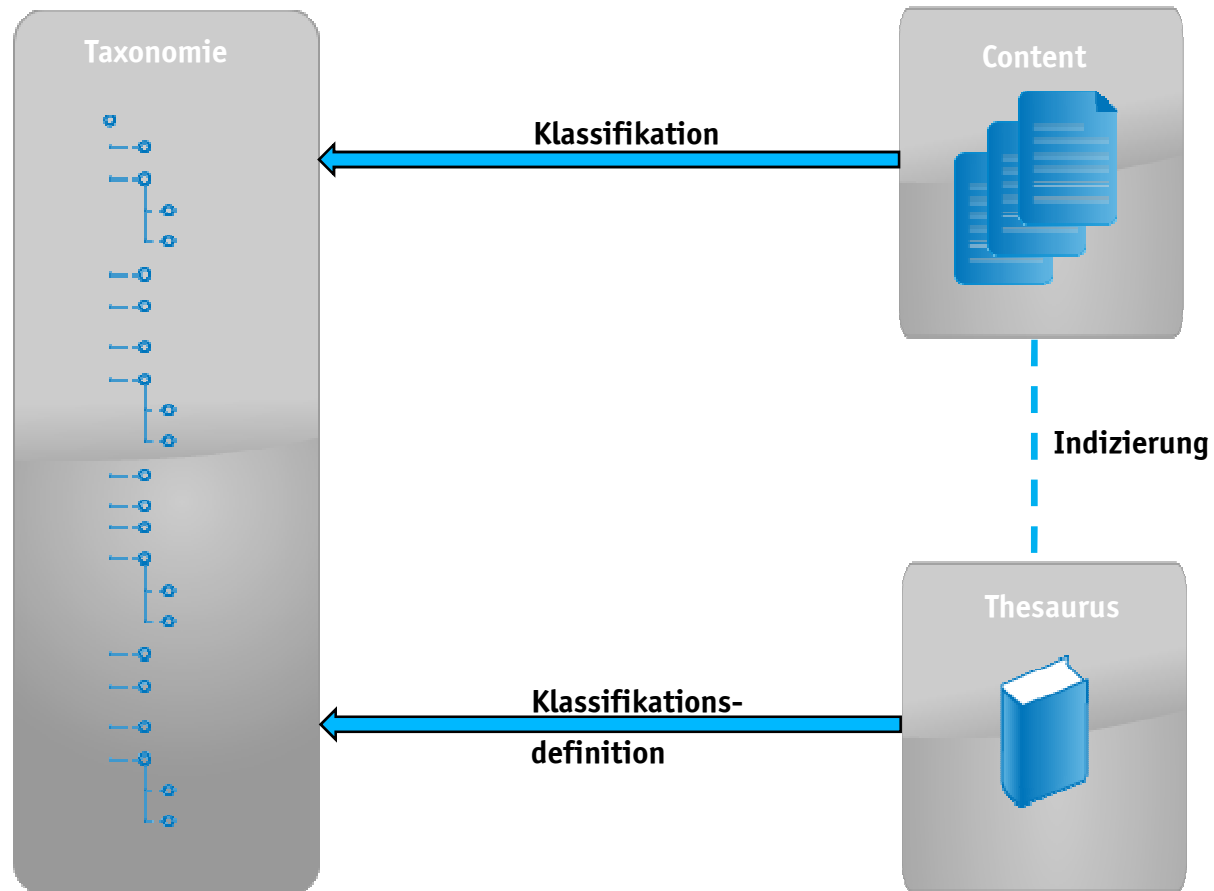
Die Morphologie (Wortbildungsregeln) klassifiziert die Worte in Gruppen (z.B. Substantive, Verben, Adjektive, Abkürzungen, etc.).

Über die Wortsammlung wird pro Dokument ein „Dokumenten-Vektor“ gebaut, indem die Häufigkeit eines jeden Wortes das im Dokument vorkommt gezählt wird.

Durch Differenz-Analysen lassen sich bereits „Schlüsselworte“ (100%) und „relevante Worte“ (>100%) zu einem Dokument oder eine Dokumentengruppe ermitteln.

Clusterverfahren ermöglichen die Gruppierung von Dokumenten aufgrund ihrer Ähnlichkeit.

1. Ziel: Gruppierung von Dokumenten und Worten in ihre logische Relationierung.





## Wissensrepräsentation – Wie?

Durch die Anwendung weiterer Analysen werden weitere Zusatz-Informationen erzeugt, die wiederum weitere Analysen stützen. Dabei werden semantische Tags in die Metadaten der XML-Ströme eingefügt.

Es werden kontextbasierte Ansichten erzeugt und als hierarische **Taxonomie** vorgehalten. Diese können zur **Klassifikation** und **Workflow**-Steuerung herangezogen.

Wort zu Wort-Beziehungen wie z.B. Synonyme und Antonyme werden im **Thesaurus** (Wortnetz) festgehalten.

Der Sinnzusammenhang (**Ontologie**) zwischen den Worten kann im Sinne der Objektorientierung zu Objekten mit ihren Merkmalen und ihren jeweiligen Ausprägungen erzeugt werden. Sogar die Zuordnung von Methoden (Verben) lässt sich automatisiert umsetzen.

2. Ziel: weitere inhaltsbasierte Strukturen erzeugen, um dem Anwenderprozess näher zu kommen.



### Wissensrepräsentation – Ziel?

Durch die Umsetzungen dieser mehrstufigen Wissens-Schöpfung erzeugt man aus Daten zunächst interpretierbare Informationen und zuletzt das Wissen über ihre Anwendung.

Durch eine Inferenzmaschine (**Fallbasiertes Schließen**) können nun die Lern-Prozesse nach **Optimierungsverfahren** automatisiert verbessert werden.

Mit jeder neuen Information wird die Integrität der Gesamtstruktur hinterfragt und aspektorientiert vervollständigt.

Somit nähert sich die Maschinen dem allgemeingültigen oder fachspezifischen Wissen an und kann z.B. wenn nach Auto gesucht wird, die Namen aller Automarken gleichfalls zur Suche heranziehen, da diese eine Klasse bilden. Dies wird als **Induktion** bzw. **Deduktion** bezeichnet.

3. Ziel: Intelligente Prozess-Steuerung zur Unterstützung der Anwender-Prozesse.

