

---

## Beteiligte Projekte

- Ausbau und Pflege der Korpora geschriebener Sprache
- Methoden der Korpusanalyse und -erschließung
- Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen

## Mitarbeiter

- Dipl.-Ing. Cyril Belica
- Dr. Marc Kupietz
- Dr. Harald Längen
- Dipl.-Inf. Rainer Perkuhn
- Heike Stadler
- Dr. Andreas Witt

## Kontakt

Dr. Marc Kupietz  
Programmbereich Korpuslinguistik  
Institut für Deutsche Sprache  
Postfach 10 16 21  
D-68016 Mannheim

Telefon: 0621/1581-0  
Fax: 0621/1581-200  
E-Mail: [korpuslinguistik@ids-mannheim.de](mailto:korpuslinguistik@ids-mannheim.de)

---



 INSTITUT FÜR  
DEUTSCHE SPRACHE

R 5, 6-13

D-68161 Mannheim

[www.ids-mannheim.de](http://www.ids-mannheim.de)

Mitglied der  
  
Leibniz-Gemeinschaft

*Das Institut für Deutsche Sprache (IDS) ist die zentrale Einrichtung zur Erforschung und Dokumentation der deutschen Sprache in ihrem Gebrauch und in ihrer neueren Geschichte. Es gehört zu den 86 außeruniversitären Forschungseinrichtungen und Serviceeinrichtungen für die Forschung der Leibniz-Gemeinschaft.*

---

---

 INSTITUT FÜR  
DEUTSCHE SPRACHE

Programmbereich Korpuslinguistik

# Das IDS- Textmodell

Eine wichtige Voraussetzung für die effiziente automatische Auswertung großer elektronischer Textsammlungen ist, dass die Texte in einem einheitlichen Format vorliegen. Für die Korpus-sammlung am IDS wird ein solches Format durch das IDS-TEXTMODELL definiert, das auf internationalen Standards basiert.

## Überblick

- vorrangiges Ziel: möglichst originalgetreue Abbildung von Inhalt und Struktur des linguistisch relevanten Quellmaterials
- breites Spektrum erfassbarer Textarten
- hierarchische Gliederung des Datenbestandes auf den Ebenen *Korpus*, *Dokument* und *Text*
- Gruppierung von Texten z.B. nach Quellen, chronologischer Abfolge, Themenbereichen und/oder Textarten möglich
- Erfassung von bibliografischen, textstrukturellen und weiteren Informationen (sofern verfügbar), die für Recherche und Auswertung sinnvoll erscheinen
- aufwärts kompatibel zu den Formatstandards TEI (Text Encoding Initiative) und CES (Corpus Encoding Standard)
- gegenwärtig umgesetzt in SGML (Standard Generalized Markup Language), Umstellung auf XML (Extensible Markup Language) in Arbeit

Stand: 04/12

```
flächendeckend versorgt.</s><s>Allein 190 000 schlaue Tonnen rollen in der Hansestadt.</s><s>Vor fünf Jahren hat  
Bremen das 20 Millionen Mark teure Codierte System eingeführt - und sich damit ganz schön verrechnet.</s></p>  
<p><s>Wo die Chips Einzug halten, landet der Müll vor den Haustüren der Nachbarn</s></p>  
<p><s>Denn die Chips funktionieren zwar, die Menschen aber nicht.</s><s>Schon bei der Umstellung liefen die Wohnungs-  
baugesellschaften gegen das Projekt Sturm, weil die Code-Tonne nicht mehr auf das Prinzip der solidarischen Kosten-  
teilung setzt, sondern auf die Einzelabrechnung.</s><s>Wo sich aber viele Personen einen Müllcontainer teilen müssen  
wie in Wohnanlagen, krankt das System, weil nicht "individuell erfasst" werden kann.</s><s>Wohnungsbau-  
gesellschaften müssen also weiterhin personen- oder quadratmeterbezogen abrechnen.</s></p>  
<p><s>Auch bei den Privathaushalten ging die städtische Rechnung nicht auf.</s><s>In kürzester Zeit hatten sich die  
Bürger heruntergespart, ließen ihre Tonnen nur noch ein Dutzend Mal im Jahr leeren.</s><s>Für die Stadt wurde die  
Abfallwirtschaft dadurch unberechenbar, sie nahm viel weniger Geld ein, musste aber Verbrennungsanlagen und Deponien  
bei konstanten Fixkosten unterhalten.</s><s>Die Folge für die Verbraucher waren steigende Müllgebühren, obwohl sich  
das Restmüllaufkommen laut Statistik ständig verringerte.</s></p>  
<p><s>Die Dresdner Zahlen wirken auf dem Papier überzeugend: Die Einführung des Identsystems für die 480 000 Einwoh-  
ner erfolgte von 1996 auf 1997.</s><s>Statt vorher 178 Kilo Restmüll pro Einwohner fielen im Jahr 1997 nur noch 30  
Kilo an.</s><s>Dafür füllten sich die übrigen Müllbehälter.</s><s>In den Transponder-Städten landet der Abfall nicht  
mehr in den schlauen Tonnen, sondern in Gelben Säcken, Biotonnen - oder vor den Haustüren der Nachbarn.</s></p>  
<p><s>"Zu den Papierkörben der Innenstadt oder zu Rastplätzen hat sich ein regelrechter Mülltourismus entwickelt",  
sagt Reinhard Holtin von den Bremer Entsorgungsbetrieben.</s><s>So schaffen es in die Bremer Biotonnen drei Tonnenleerun-  
gen jährlich auszukommen.</s><s>Das "Amt für Schiet und Dreck", wie Wagenfahrer Tönen am Amt, hat dennoch  
mehr zu tun als vorher, weil es die elektronische Datenpflege am "Trash R" einleiten musste.</s><s>Die Bilanz des  
Bremer "Wahnsinnsvorhabens", das mit 40 Müllfahrzeugen, 50 Mitarbeitern, die Müll aufkommen und 20 Millionen  
Mark Einsatz mehr Gebührengerechtigkeit schaffen wollte, ist ein kolossales Scheitern.</s><s>Das neue System, das aber Trittbrettfah-  
rern das kostenlose Mitfahren ermöglicht.</s><s>Trash R ist ein Paradebeispiel für den Übergang von der geregelten Müllentsorgung zum  
Sankt-Florians-Prinzip.</s></p>  
</div>  
</body>  
<back>  
<div n="1" type="text" id="Z99/906.93180">  
<figure id="Z99/906.93180">  
<p><s>Der Müll als Datenpfeiler, die Gebühren steigen</s></p>  
</figure>  
</div>  
<back>  
</text>  
</idstext>  
<idstext version="1">  
<idshdr type="text" id="Z99/906.93180">  
<fileDesc>  
<titleStmt>  
<textsigle>Z99/906.93180</textsigle>  
<t.title assemblage="external">Z99/906.93180 Die Zeit, 10.06.1999, Nr. 24, Ressort: Wissen; Die Moden der Orthopä-  
den</t.title>  
</titleStmt>  
<editionStmt version="0">  
</editionStmt>  
<publicationStmt>  
<distributor></distributor>  
<pubAddress></pubAddress>  
<availability></availability>  
<pubDate></pubDate>  
</publicationStmt>  
<sourceDesc>  
<biblStruct>  
<analytic>  
<h.title type="main">Die Moden der Orthopäden</h.title>  
<h.author>Ruta Tengelhardt</h.author>  
<imprint>  
<pubDate type="year">1999</pubDate>  
<pubDate type="month">06</pubDate>  
<pubDate type="day">10</pubDate>  
</imprint>  
<biblScope type="pp">S. 35</biblScope>  
<biblScope type="issue">24</biblScope>
```

Auszug aus einem gemäß dem  
IDS-Textmodell kodierten Korpus