

DSL

1. Teil (Stefan Müller)

1.1 Was ist eine DSL?

1.2 Beispiele

1.3 Hinweise für die Erstellung einer DSL

1.4 Warum DSL?

2. Teil (Eckhart Arnold)

Verarbeitung von DSL. Vorzüge von DSL.

1.1 Was ist eine DSL?

DSL

domain specific language

‘domänenspezifische Sprache’

‘maßgeschneiderte Annotationssprache’

Vgl. XML, aber mit selbstgesetzten Regeln, die auf den Anwendungszweck hin zugeschnitten sind.

(Sie tun es ja sowieso.)

1.2 Beispiele

„Notizen aus französischen Archiven“

Transkripte von Gesandtschaftsinstruktionen (J. Schulz)

... nulle guerre ne inimitié ne besoing de **[korrigiert aus: de___]**

[5r]

[am Rand:] *l'ayde et despense* desdicts roys de Polonnye et
de Honguerye [Zeichen ohne Gegenzeichen] que devole et
monstre clerement que ledict sr. ne faisons ladicte offre
a autre fin si n'est pour le desir et [gestr: intencion] *affection*
qu'il auvoit ...

„Das machen wir seit 200(0) Jahren so“
Reiches Erbe der philologischen Notation
(Ptolemaeus Arabus et Latinus)

<Liber I>

Liber hic [hic] Almagesti †predasissimo Bolemeist ...

<2> Capitulum secundum: De ordinibus ...

<Q>uod prius de hac scientia est sciendum vel premittendum
est ... ab eis quoque¹ inseparabilium, que promeniunt²
pro/2v/pter declinationes secundum eorum ordines.

¹ quoque *add.* but then *del.*

² Probably corrupt for: proveniunt. Cf. Manitius, I, page 38.

„was Mentschikoff schon zu Lebzeiten nicht genügt hat“
Abhilfe schaffen in der Eingabeumgebung
(Russische Relationen)

```
nun die dieserwegen sonder  
zweifel mit nicht erfor[der]licher  
behutsamkeit genomme <!passus><-:~?; Lesung unsicher.>  
verlaut baret worden,  
so hat auch der König in Preuss[en,]  
welcher wegen sicheren auf  
das Herzogthumb Curland  
vermeint[lich] habenden gelt-  
protektionen schon lange
```

„Alles kann Bedeutung haben“

Tiefenannotation von 1.100.000 Wortformen

(Bonner Korpus der mittelhochdeutschen Grammatik)

***{D*}**ihtens het ich mich verlaubt von d' zeit her,, \$eit min
haubt Manger laye done gewan.(:) Si(=)den, dizzen, \$u\e\$e\-,
\$inge\-, Zwitzern, grelle\-, \$nurren, clingen,, Die do\ene ich
gelernet han,, ...

`tihten @n	``tihten	#n	^GS	~tihten<e>s	*{D*}ihtens
`haben @swv-a	``haben	#swv-a	^1SV	~hæt<e>	\het
`ich @pron	``ich	#pron	^NS	~ich	\ich
`ich @pron	``ich	#pron	^AS®	~mich	\mich
`ver-louben @swv	``ver-louben	#swv pV	^-	~ver-loub<e>t	\verlaubt

1.3 Hinweise für die Erstellung einer DSL

Leitfaden. Worum geht es grundsätzlich?

- Eine DSL ist eine Sprache:
ein Regelwerk, Inhalt auszudrücken und Ausdrücken
Inhalt zu entnehmen.
Auszeichnung: Aussage über Daten mit Bereichsbezug
- Sprachverwendung ist eine Handlung, daraus folgt:
 - Der Wandel gewachsener Sprachen ist geprägt von sprachökonomischen Neigungen.
 - Die Gestaltung einer zu entwerfenden Sprache sollte sprachökonomische Neigungen berücksichtigen.

Daraus folgen zwei Grundfragen:

- Zielbestimmung: Welche Inhalte sollen unterschieden werden?
- Wegbereitung: Wie lassen sich die Inhalte mit Zeichen und Zeichenzusammensetzungen möglichst einfach und dennoch so eindeutig unterscheiden, dass eine automatische Verarbeitung möglich ist?

- Welche Inhalte sollen unterschieden werden?
 - Welche Arten von Inhalten liegen vor?

- paradigmatisch (Klassifizierung):
 - kategoriale Unterscheidung genügt
 - individuelle Unterscheidung nötig → IDs nötig
- syntagmatisch (Bereichsbezug):
 - Bereich ist ein Punkt (Abschnittsgrenzen?)
 - Bereiche sind ausgedehnt, aber nie verschachtelt → Start- und Endbezeichnung nötig (z. B. "...")
 - Bereiche können verschachtelt sein → Start- und Endbezeichnung müssen verschieden sein
 - Bereiche können einander überlappen oder diskontinuierlich sein → IDs

- Wie lassen sich die Inhalte mit Zeichen und Zeichenzusammensetzungen möglichst einfach und dennoch so eindeutig unterscheiden, dass eine automatische Verarbeitung möglich ist?
→ Sprachökonomische Abwägung

- Einheitlichkeit

ein Ausdruck – ein Inhalt (keine Polysemie!),

ein Inhalt – ein Ausdruck (keine Polynymie!).

Vermeidung von Mehrdeutigkeit, vor allem zwischen
Meta- und Objektzeichen. #421 #421# #421;

- Ausdrücklichkeit

Zeichen > Zeichenzusammensetzungen >> Inhalt

- Bequemlichkeit

Möglichst viel automatisch erledigbar machen.

Eingabebequemlichkeit.

Beispiele

Gesandtschaftsinstruktionen:

Pagina-Angaben;

statt Schweifklammern: [#2r]

Ptolemaeus:

Editorentext ←→ Kursivierung

#<i>1.1</i><...>

Bonner Korpus der mhd. Gr.:

Propriale Verwendung eher als z. B. Modalverben.

Komma verwendbar.

XML: <name attr="...>...>

„Notizen aus französischen Archiven“

Transkripte von Gesandtschaftsinstruktionen (J. Schulz)

... nulle guerre ne inimitié ne besoing de [korrigiert aus: de___]

[5r]

[am Rand:] *l'ayde et despense* desdicts roys de Polonnye et
de Honguerye [Zeichen ohne Gegenzeichen] que devole et
monstre clerement que ledict sr. ne faisons ladicte offre
a autre fin si n'est pour le desir et [gestr: intencion] *affection*
qu'il auvoit ...

„Notizen aus französischen Archiven“

Transkripte von Gesandtschaftsinstruktionen (J. Schulz)

... nulle guerre ne inimitié ne besoing de [korrigiert aus: de___]

[#5r]

[am Rand:] *l'ayde et despense* desdicts roys de Polonnye et
de Honguerye [Zeichen ohne Gegenzeichen] que devole et
monstre clerement que ledict sr. ne faisons ladicte offre
a autre fin si n'est pour le desir et [gestr: intencion] *affection*
qu'il auvoit ...

„Das machen wir seit 200(0) Jahren so“
Reiches Erbe der philologischen Notation
(Ptolemaeus Arabus et Latinus)

<Liber I>

Liber hic Almagesti preclarissimi Ptolemei ...

<2> Capitulum secundum: De ordinibus ...

<Q>uod prius de hac scientia est sciendum vel premittendum
est ab eis quoque³ inseparabilium, que promeniunt⁴
propter declinationes secundum eorum ordines.

³ quoque *add.* but then *del.*

⁴ Probably corrupt for: *proveniunt*. Cf. Manitius, I, page 38.

„Das machen wir seit 200(0) Jahren so“
Reiches Erbe der philologischen Notation
(Ptolemaeus Arabus et Latinus)

#<I><Liber I>

Liber hic Almagesti preclarissimi Ptolemei ...

#<I.2> Capitulum secundum: De ordinibus ...

<Q>uod prius de hac scientia est sciendum vel premittendum
est ... ab eis quoque⁵ inseparabilium, que promeniunt⁶
pro/2v/pter declinationes secundum eorum ordines.

⁵ quoque *add. but then del.*

⁶ *Probably corrupt for:* proveniunt. *Cf. Manitius, I, page 38.*

„Alles kann Bedeutung haben“

Tiefenannotation von 1.100.000 Wortformen

(Bonner Korpus der mittelhochdeutschen Grammatik)

***{D*}**ihtens het ich mich verlaubt von d' zeit her,, \$eit min
haubt Manger laye done gewan.(:) Si(=)den, dizzen, \$u\e\$e\-,
\$inge\-, Zwitzern, grelle\-, \$nurren, clingen,, Die do\ene ich
gelernet han,, ...

`tihten @n	``tihten	#n	^GS	~tihten<e>s	*{D*}ihtens
`haben @swv-a	``haben	#swv-a	^1SV	~hæt<e>	\het
`ich @pron	``ich	#pron	^NS	~ich	\ich
`ich @pron	``ich	#pron	^AS®	~mich	\mich
`ver-louben @swv	``ver-louben	#swv pV	^-	~ver-loub<e>t	\verlaubt

„Alles kann Bedeutung haben“

Tiefenannotation von 1.100.000 Wortformen
(Bonner Korpus der mittelhochdeutschen Grammatik)

...

```
<token>  
  <form dipl="het" trans="het" norm="h&#xE6;t&lt;e&gt;"/>  
  <lemma gen="haben" inst="haben"/>  
  <pos gen="VA" inst="VAFIN"/>  
  <inflClass gen="wk" inst="wk"/>  
  <infl val="*_Past_Sg_1"/>  
</token>
```

...

1.4 Warum DSL?

Eigenauszeichnungen treten ohnehin allenthalben auf:

- Irgendwelche Auszeichnungen sind fast immer nötig.
- Eingabeumgebungen sind oft (noch) nicht vorhanden, hinreichend oder unter den geg. Umständen nutzbar.
- Viell. ist eine im Fach übl. Notation zu berücksichtigen.

Es ist besser, seine Auszeichnungen von vornherein so zu gestalten, dass sie automatisch konvertiert werden können (z. B. in XML).

Die dafür angestellten Überlegungen, welche Auszeichnungsfälle wie zu unterscheiden sind, schärfen und klären den Blick auf das, was ausgezeichnet werden soll.