

Universität Hamburg • Institut für Germanistik I • Sommersemester 2006

## Vorlesung Computerphilologie

# Themenfeld Internet

„Wie kann man als Textwissenschaftler das  
Internet nutzen und gestalten,  
und wie ist dieses aufgebaut?“

# Informationserhebung und -filterung

- Das WWW kann man als riesige Datenquelle auffassen, die aber ohne Filterung meist nutzlos ist. Beispiel:
- Eine Anfrage „*Computer + Schule*“ brachte
  - bei Fireball 2.091.105 Fundstellen,
  - bei Yahoo 100 Fundstellen und 14 Nachrichten, darunter auch die Taifun-Nachricht:

„Der Taifun verursachte Stromausfälle sowie Störungen im Luft- und Straßenverkehr. Mehr als 250.000 Haushalte waren ohne Strom. Die Finanzmärkte in der Hauptstadt Taipeh blieben am Montag geschlossen, ebenso Regierungsbüros und **Schulen**. Die Arbeit im Industriegebiet im Norden Taiwans wurde nicht ausgesetzt. Auch das Zentrum der **Computer**chip-Produzenten von Hsinchu war nicht betroffen.“

# Konsequenz

- Die Internet-Texte müssen aufbereitet werden, um die richtige Information und genau diese zu finden. Was wir finden, ist einerseits zu viel und andererseits zu wenig:
- Sei
  - a die Menge der relevanten Treffer einer Suche
  - b die Menge der nicht relevanten Fundstellen und
  - c die Menge der nicht gefundenen relevanten Daten
  - So ist der *Recall* (Vollständigkeit) =  $\frac{a}{a+c}$  und
- Die *Precision* (Genauigkeit) =  $\frac{a}{a+b}$
- Leider ist dieses Maß nur bei Datenbanken wirksam anwendbar, denn nur dort kennt man den Wert von c. Im Internet brauchen wir daher nicht numerische, sondern intelligentere Methoden, um überhaupt zu bestimmen, was „vernünftige“ Ergebnisse sind.

# Was suchen wir überhaupt?

## Fakten?

- Wie groß ist die Entfernung vom Mond zur Erde?
- Welches Bruttosozialprodukt hatte Italien im Jahr 1998?

⇒ **Faktenretrieval in Datenbanken**

## Fundtexte?

- Wo kann ich etwas zur Entwicklung der Chipherstellung lesen?
- Allgemeine Texte zur Steinzeit?

⇒ **Stichwort- und Kategorienretrieval im Internet**

## Antworten?

- Was kann ich in der Schule mit dem Internet machen?
- Wo ist der beste Urlaubsort für mich?

⇒ **Data-Mining**

# Welche Möglichkeiten der Suche haben wir?

1. Web-Texte sind in einer einheitlichen Darstellungs-Sprache organisiert (HTML), Sie enthalten den Text + Gliederungsinformation + wenige allgemeine Daten. Wir haben also mehr als nur die Wörter eines Textes, nämlich auch dessen **Organisationsinformation** (z. B. Überschriften, Institution, Datum, Sprache, Suchbegriffe) 
2. Mit geeigneten Methoden können wir sogar **linguistische Beziehungen** finden:
  - Flexionsformen derselben Wörter zusammenfassen,
  - Relevante Wörter von grundsätzlich irrelevanten trennen,
  - Domänenrelevante Wörter auswählen,
  - Syntaktische (damit u.U. logische) Abhängigkeiten berechnen,
3. Wir können gefundene Wörter aus einem Text mit einem **Begriffsgerüst** (einer Ontologie, z.B. im *Semantic Web*) vergleichen, wir können also (vom Fundwort zur Ontologie) Unter- und Oberbegriffe zur weiteren Suche gegeneinander tauschen oder (von der Ontologie zum Fundwort) Teile einer Ontologie im Netz suchen.
4. Erfolgreiche Suche analysieren und weitere Suchprozesse dadurch steuern (Effiziente **Suchstrategien lernen**)

# Hypertexte, die Grundidee von HTML

- Die Inhaltseinheiten im WWW sind fast ausschließlich Hypertexte.
- Als Hypertext bezeichnet man Texte, die in nicht-linearer Anordnung Inhalte präsentieren. So kann jeder Leser die Reihenfolge und die Granularität des konsumierten Inhalts selbst bestimmen.
- Hypertexte eignen sich daher besonders gut
  - für strukturierte Lerninformation,
  - für Texte, die von sehr unterschiedlichen Lesern auf unterschiedliche Art gelesen werden und
  - Zur synthetischen Präsentation von Informationen aus unterschiedlichen Quellen (virtuelle Webseiten)
  - Aber auch als neue Literaturgattung

# Granularität

- Eigentlich die Korngröße eines Schüttguts,
- metaphorisch aber auch der Grad der Detailliertheit und Strukturiertheit eines Textes und seiner [Inhalte](#)

[zurück](#)

# Nicht-Linearität

Nicht durch die lokale Reihenfolgerelation verbunden

- Linear sind z.B. normale Bücher, die Präsentation normaler Filme, die meisten Instruktionstexte, Hörereignisse, wie Musik oder Hörspiele
- Nichtlinear sind schon immer z.B. Fahrpläne, in denen man Verbindungen heraussuchen kann, die Logik von Kindererzählungen oder Filmen, die ungezielte Wegesuche.



# Inhalte und ihre Eigenschaften

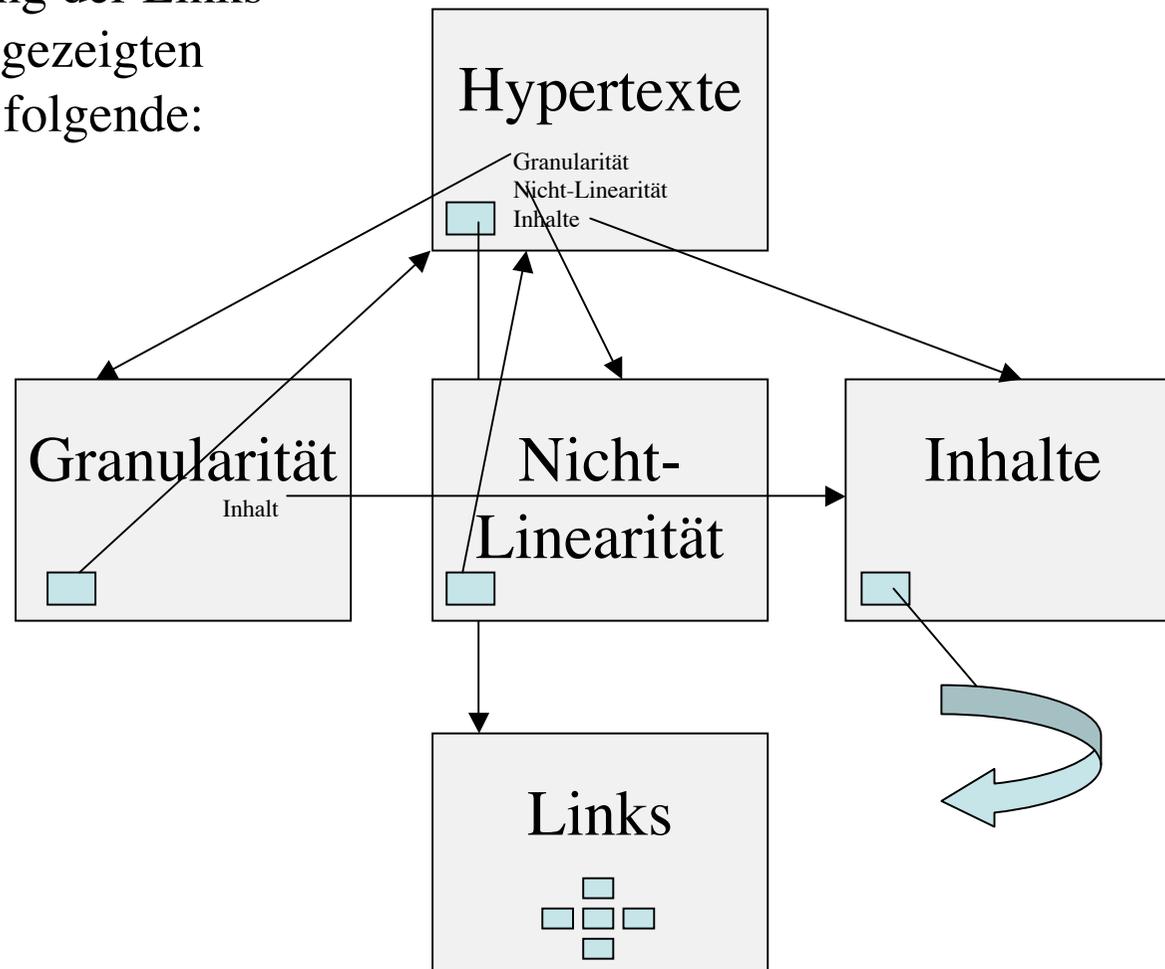
Um Teiltex te beliebig (je nach dem Fortschreiten des Lesers) und beliebig oft (z.B. Definitionen oder Benutzungshilfen) miteinander zu verbinden, haben die Teiltex te in Hypermedien spezifische Eigenschaften:

- Inhalte in Hypermedien sind meist abgeschlossener als in linearen Medien,
- Sie sind einliniger aber durch die Links eingebettet in unterschiedliche semantische Kontexte.
- Meistens sind sie kleinräumiger (stärker modularisiert) als sequentielle Text e, um ihre mehrfache Einbettung an unterschiedlichen Stellen zu erleichtern

[zurück](#)

# Links

Die Anordnung der Links  
auf den eben gezeigten  
Seiten ist die folgende:



# Hypertextsysteme

erlauben die einfache Implementierung (Programmierung) von Hypertexten.

- In gewissem Sinne ist PowerPoint auch ein Hypertextsystem, obwohl es nicht vorwiegend für diesen Zweck entworfen ist.
- HTML-Browser, die ihre Funktionalität durch die implementierten „W3C Recommendations“ erhalten. Hier ist das Hypertextsystem eher der Link-Teil des Browsers.
- Reine Hypertextsysteme mit einer Entwicklungsumgebung, wie z.B. das (relativ alte) HyperCard-System von Apple mit seiner Sprache Hypertalk.
- Multimedia-Systeme mit Hypertext-Funktionen, wie Makromedia Director
- Schwierigere Server-Lösungen z.B. mit PHP (sie laufen nicht auf dem lokalen Rechner ab)
- Auch für andere Endgeräteklassen gibt es Hypertextlösungen.

# HTML

Abkürzung für „Hypertext Markup Language“. Zur Zeit ist die „HTML Specification 4.01“ gültig. Festgelegt ist sie durch die W3C Recommendation vom 24 December 1999.

HTML ist eine Übereinkunft von Web-Entwicklern („W3C“), wie in Web-Browsern plattformunabhängig Webseiten (durch welche Befehle) aussehen sollen.

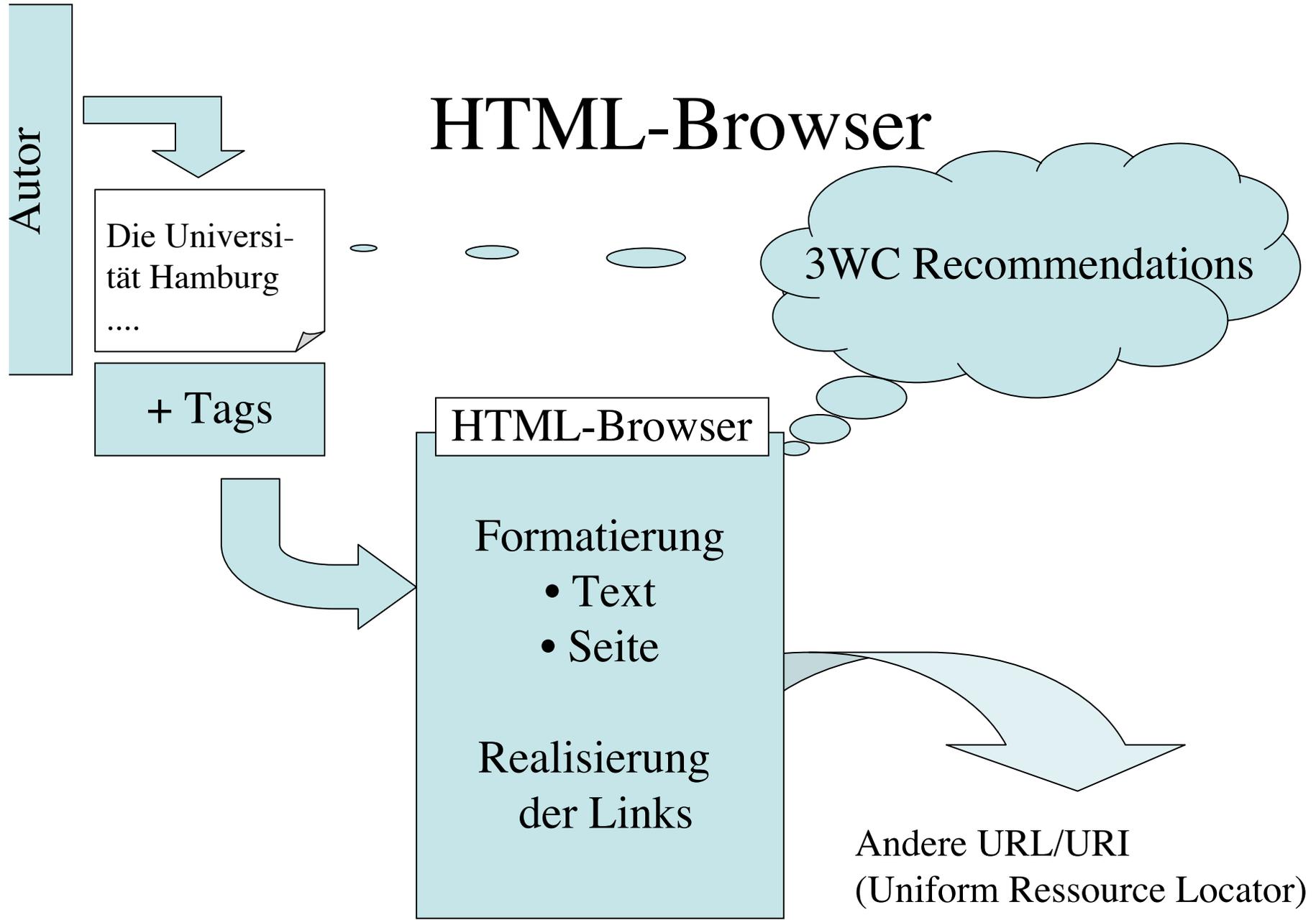
z.B. soll die Zeichenfolge `<hr>` immer eine waagerechte Linie erzeugen:



HTML ist aus der Sicht der Benutzer die formale Sprache, in der man Webseiten durch Auszeichnen von Texten (Markup) mit Meta-Zeichenfolgen (Tags) für die empfangenden Computer schreibt.

Ich gehe zunächst nicht auf XHTML, die XML-gerechte HTML-Neudefinition, ein, sondern behandle sie unter SGML und XML

Eine sehr gute Einführung steht unter: <http://de.selfhtml.org/>



# HTML Texte

- Ein HTML-Text ist ein [ASCII](#)-Text, der zusätzliche spezielle Zeichenfolgen zur Darstellung in einem Browser enthält. Ein HTML-Text wird auf allen Plattformen (Betriebssystemen) sehr ähnlich dargestellt. Daher kann man das WWW auch von jedem Rechner aus lesen.
- Die HTML-spezifischen spezielle Zeichenfolgen („Tags“), mit denen der Text ausge-zeichnet wird, beschreiben letztlich die logische Struktur einer Webseite, wie etwa:
  - Seitenaufteilung, Überschriften, Absätze, Listen, Tabellen, Zeichenform, Farben, Bildverankerung, und
  - vor allem, die (Hyper-)Links zu anderen Punkten,aber weniger deren konkrete Präsentation.

# Links

Links können ausgehen von

- Textelementen,
- graphischen Objekten,
- Bildern,
- definierten Regionen einer Seite
- definierten Regionen eines Bildes.

Sie führen zu

- andern Absätzen derselben Seite,
- anderen Dokumenten desselben Rechners oder
- Dokumenten anderer weltweiter Rechner (URLs, allgemein URIs)

# ASCII

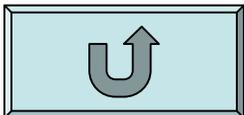
Zeichensatz nach ISO 8859-1, auch „Western“ oder „Latin 1“ genannt.

- Steuerzeichen wie Zeilenwechsel, Leer, Tab, etc.
- Satzzeichen: , . ! ? ; : - " ` ´
- Sonderzeichen § \$ % & / \ ( ) < > @ ^ \_ [ ] { } | ~
- arithmetische Zeichen: + - \* / =
- Zahlen von 0 - 9
- Großbuchstaben: A - Z
- Kleibuchstaben: a - z

In einer zweiten erweiterten Hälfte dieser Tabelle können weitere Zeichen stehen, wie die deutschen Umlaute, das französische ç das dänische å, etc.

Alle für das Notieren von HTML reservierten Zeichen und die Zeichen der zweiten Hälfte müssen in HTML aber umschrieben werden:

< = &lt;      Ä = &Auml;    oder &nbsp; für **Leerzeichen**



# Die Wirkung von HTML-Tags

HTML:	<code>&lt;h1&gt;Dies ist eine &amp;Uuml;berschrift &lt;/h1&gt;</code>
Sieht so aus:	<b>Dies ist eine Überschrift</b>
HTML:	<pre> Man kann eine Liste schreiben &lt;ol&gt;     &lt;li&gt;mit Zahlen&lt;/li&gt;     &lt;li&gt;mit Spiegelstrichen &lt;/li&gt;     &lt;li&gt;mit "bullets"&lt;/li&gt; &lt;/ol&gt; </pre>
Sieht so aus:	<p><b>Man kann eine Liste schreiben</b></p> <ol style="list-style-type: none"> <li><b>1. mit Zahlen</b></li> <li><b>2. mit Spiegelstrichen</b></li> <li><b>3. mit "bullets"</b></li> </ol>

# HTML von Kopf bis Fuß

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN"  
  "http://www.w3.org/TR/1999/REC-html401-19991224/loose.dtd">
```

```
<html lang="de">
```

Prolog

```
<head>
```

```
<meta http-equiv="content-type" content="text/html; charset=iso-8859-1">
```

```
<title>Eine leere Seite</title>
```

```
<meta name="generator" content="BBEdit 6.1.1">
```

Kopf

```
</head>
```

```
<body>
```

```
Hier steht der Text der Web-Seite
```

Körper

```
</body>
```

```
</html>
```



Prolog Einträge ....

```
<html>
<head> Head-Einträge ..... </head>
<body>
<h1>Vorlesung "Computerphilologie"</h1>
<h2>W.v.Hahn</h2>
<h4>Inhalt:</h4>
<ul>
  <li>01 Einführung</li>
  <li>02Theorie</li>
  <li>03Internet</li>
  <li>04Multimedia</li>
  <li>05H&ouml;here Textverarbeitung</li>
  <li>06Lexikalische Repr&auml;sentation</li>
  <li>07Syntaktische Repr&auml;sentation</li>
</ul>
</body>
</html>
```

# Ausführlicheres Beispiel

# Das sieht dann so aus:

## Vorlesung "Computerphilologie"

**W.v.Hahn**

Inhalt:

- Einführung
- Theorie
- Internet
- Multimedia
- Höhere Textverarbeitung
- Lexikalische Repräsentation
- Syntaktische Repräsentation

# Wie schreibt man HTML-Texte?

ASCII-Editor

1

```
<html>
<head>
<title>Beispiel</title>
</head>
<body>
<ul>
<li>ASCII-Texte sind</li>
<ul>
<li>Kurz und schnell
&uuml;bertragbar</li>
<li>&Uuml;berall
lesbar</li>
</ul>
<li>Textsysteme wie Word,
WordPerfect etc. erzeugen
intern eine
riesige Menge von
„M&uuml;ll“ und z.T.
plattformabh&auml;ngige
interne
Steuerungen</li>
</ul>
</body>
</html>
```

1a

- ASCII-Texte sind
  - Kurz und schnell übertragbar
  - Überall lesbar
  - Überall darstellbar
- Textsysteme wie Word, WordPerfect etc. erzeugen intern eine riesige Menge von „Müll“ und z.T. plattformabhängige interne Steuerungen.

- ASCII-Texte sind
  - Kurz und schnell übertragbar
  - Überall lesbar
  - Überall darstellbar
- Textsysteme wie Word, WordPerfect etc. erzeugen intern eine riesige Menge von „Müll“ und z.T. plattformabhängige interne Steuerungen.

Im Browser ansehen

2

WYSIWYG-Editor, z.B. NVu

**Nicht Explorer nehmen!**

# Warum nicht einfach ein Textsystem?

<i>ASCII-Texte</i>	<i>Schreibsystem-Texte</i>
Kurz, daher schnell übertragbar	Mehr als 100 mal längere Texte durch aufwendige interne Codierung
Mit jedem Editor les- und schreibbar	Gebunden an lizenzierte Software
Plattformunabhängig	Plattformabhängig
Versionsunabhängig	Versionsabhängig

- Nie versuchen: Einen Word-Text (ab Version 2000) „als HTML speichern“, denn WORD produziert eine entsprechend riesige Menge von HTML-Müll, den man zwar nicht sieht, der aber für Internet-Versendung viel zu sperrig wird.
- Da der MS Internet Explorer keinen Editor (mehr) hat, sondern WORD aufgerufen wird, kann man mit dem Explorer keine effizienten HTML-Texte schreiben oder editieren (schon das öffnen reicht, um den Text aufzublähen).
- Das Lesen von HTML-Texten ist aber bei allen Browsern (ziemlich) ähnlich. Ausnahme z.B.: JavaScript-Interpretation ist im Internet Explorer teilweise anders (dort gibt es JS).

# Codierungs-“Müll“

Der Text der vorigen Folie (ohne Tabelle) verbraucht:

<i>Bei Codierung als</i>	<i>Anzahl Zeichen</i>	<i>Anzahl Wörter</i>
ASCII-Text	429	62
HTML:	1116	179
Word-Code (ohne Editieren)	41.984	5956
WORD → HTML	6255	1002

# Was tut der Nvu? (Netscape Composer)

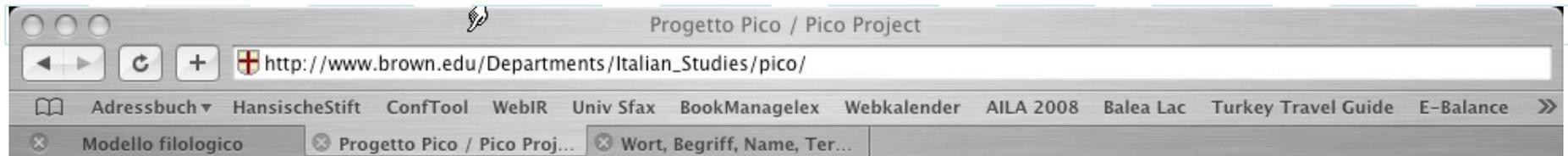
- Man schreibt den Text ungefähr wie mit einem Textsystem,
- Für Formatierung hat man eine Menüauswahl,
- Nvu (der Composer) produziert intern HTML,
- Einzelne Markups werden durch Icons angezeigt,
- Man kann sich den eigentlichen HTML-Text durch die Funktion „Quelltext“ ansehen und auch dort editieren
- Man kann in Nvu (im Composer) jederzeit bestehenden Text editieren
- Die Version 1.0 ist noch etwas klapprig

# Konsequenz

1. Als Browser für eigene HTML-Versuche am besten nur Firefox (oder Netscape) benutzen. Es gibt ihn kostenlos im Netz unter  
<http://home.netscape.com/download/archive.html>
2. HTML-Texte schreiben
  - in einem einfachen Editor oder
  - In Nvu (im Netscape Composer).
3. Alternativ: Zwar schreiben in einem Schreibsystem dann aber speichern als „Nur Text“ / „ASCII“ oder ähnlich und im Nvu/Netscape als HTML-Text bearbeiten.
4. HTML ruhig ausprobieren für Offline-Texte, z.B. Exzerpte für eine Seminararbeit
5. Es gibt zahllose (meist teure) spezielle HTML-Editoren, die HTML nach dem WYSIWYG-Prinzip produzieren. Man sollte aber immer den Code verstehen können.

# Ein reales Beispiel

- Hypertext-Ausgabe der Rede *de hominis dignitate* von Giovanni Pico della Mirandola (1463 - 1494), eines typischen Dokuments des Lebensgefühls der Renaissance.
- [http://www.brown.edu/Departments/Italian\\_Studies/pico/index.html](http://www.brown.edu/Departments/Italian_Studies/pico/index.html)



#### Presentazioni/Introductions

[latine](#)  
[italiano](#)  
[English](#)

#### L'apparato critico/The Critical Apparatus

[Testo annotato/Annotated Text](#)  
[Bibliografia](#)

#### Il progetto/The Project

[Collaboratori](#)  
[Project Participants](#)  
[Newsgroup](#) (limited access)

[Il modello filologico](#)  
[Project Guide & Templates](#)

#### Gli incunaboli/The Incunabula

[Editio princeps \(1496\)](#)  
[Exemplar Brunense \(ca. 1498\)](#)  
[Exemplar Bononiense \(ca. 1498\)](#)  
[Apologia \(1487\)](#)

[La trascrizione del testo/  
The Transcription of the Text](#)

#### Documenti ausiliari/Auxiliary Documents

[La cronologia della vita di Pico](#)  
Altri testi di Pico

[Saggi, contributi vari](#)

Avvertenza: alcune sezioni di questo sito sono ancora in fase di costruzione. Al più presto saranno completate l'edizione del testo dell'*Oratio* con i relativi apparati critici e le traduzioni italiana e inglese.

Notice: some sections of this site are still in the construction phase. The entire edition of the *Oratio*, complete with

# Layout

## Testing inside the four-part window:

The window is composed of a series of frames. Let us imagine the Latin text side-by-side with the English translation (precisely what you find in the [prototype](#)):

Latin text	English translation
variants or linguistic notes	contextual notes

Oratio

http://www.brown.edu/Departments/Italian\_Studies/pico/text/bori/frame2.html

Adressbuch HansischeStift ConfTool WebIR Univ Sfax BookManagelex Webkalender AILA 2008 Balea Lac Turkey Travel Guide E-Balance

Modello filologico Oratio

§ 1.

1. [132r] Legi,<sup>1</sup> Patres Colendissimi, in Arabum monumentis,<sup>2</sup> interrogatum Abdalam<sup>a</sup> sarracenum,<sup>b</sup> quid in hac quasi mundana scena<sup>c</sup> admirandum maxime spectaretur, nihil spectari homine mirabilius respondisse.<sup>d</sup>

2. Cui sententiae illud Mercurii adstip miraculum est homo».<sup>3</sup>

§ 2.

3. Horum dicatorum rationem cogitant quae multa de humanae naturae prae hominem creaturarum internuntium inferiorum; sensuum perspicacia, rationis naturae interpretem; stabilis evi et fluxus Persae<sup>e</sup> dicunt) mundi copulam, immo Davide, paulo deminutum.<sup>7</sup>

<sup>a</sup> prophetam add. P

<sup>b</sup> cum eum rogarent eius discipuli add. P

<sup>c</sup> scena: scaena P. machina legebatur antea

<sup>d</sup> Nihil ... respondisse: post aliquas litteras imperite scriptum) idest hominem respondi

<sup>e</sup> Persae: magi P

<sup>f</sup> sibi iure: iure sibi P

§ 1. Preambolo<sup>1</sup>

1. Ho letto, molto venerabili Padri,<sup>2</sup> nelle fonti degli Arabi<sup>3</sup> che Abdalla Saraceno<sup>4</sup> interrogato su che cosa, in questa sorta di scena del mondo, scorgesse di sommamente mirabile, rispose che non scorgeva nulla di più mirabile dell'uomo.

... illo di Mercurio: «Grande miracolo, o ...  
... ni correnti circa la superiorità umana  
... queste affermazioni non erano  
... olti sono addotte circa l'eccellenza  
... è principio di comunicazione tra le  
... sovrano sulle inferiori; per la  
... ne razionale e per il lume  
... atura<sup>7</sup>; interstizio tra la fissità  
... (come dicono i persiani)<sup>8</sup> copula, anzi  
... ageli (e dà testimonianza Davide)<sup>9</sup>

... ana, in cui si discuteranno le sue 900 tesi, e la  
... o concilio (cfr. par. 164).  
... ida di Flavio Mitridate. Ne parla nella lettera

... con l'identità di questo "Asclepius" sono state fatte varie ipotesi. Cfr. Bori 1997 e Bori, *Pluralità delle vie*, Milano 2000, 43 s.

<sup>5</sup> «Grande miracolo; o Asclepio, è l'uomo, essere animato da adorare e da onorare.» (Asclepio VI. 1.2). Gli scritti oracolari sono attribuiti al dio Thot, in greco Emata

```

<html>
<head><title>Oratio</title></head>
<FRAMESET frameborder="0"
framespacing="0" border="0" rows="*,55">
<frameset cols="50%, 50%" border="3"
bordercolor="#9d9d9d">
<frame src="latin.html" name="left">
<frame src="italian.html" name="right">
</frameset>
<frame src="credit.html" name="credit"
scrolling="no noresize">
</frameset>
</html>

```

Oratio

http://www.brown.edu/Departments/Italian\_Studies/pico/prototipo/frame.html

Modello filologico | Oratio | Wort, Begriff, Name, Ter...

§ 11.

54. Sed qua ratione, aut quid tandem agentes?

55. Videamus quid illi agant, quam vivant vitam.

56. Eam si et nos vixerimus (possumus enim) illorum sortem iam equaverimus.<sup>a</sup>

57. Ardet Saraph charitatis igne; fulget Cherub intelligentiae splendore; stat Thronus iudicii firmitate.<sup>1</sup>

58. Igitur si actuosae<sup>2</sup> ad[d]icti vitae inferiorum curam recto examine susceperimus, Thronorum stata soliditate firmabimur.

59. Si ab actionibus feriat, in opificio opificem, in opifice opificium meditantem,<sup>3</sup> in contemplandi ocio<sup>b</sup> negociabimur, luce Cherubica undique coruscabimus.

60. Si charitate ipsum opificem solum ardebimus,<sup>c</sup> est, in Saraphicam effigiem repente flammabimur.

<sup>a</sup> P *aequaverimus*

<sup>b</sup> P *otio*

<sup>c</sup> P *saeculorum*

<sup>d</sup> *anteluchanis hynnīs* (Tognon, 12).

<sup>e</sup> P *delevit totam hanc sententiam*

§ 11

54.

55.

56.

57.

sple  
judg

58.

infe  
stret

59.

creation and creation in the Creator; we will be engaging in the leisure of contemplation; we will shine on all sides with Cherubic light.

Edictio princeps, § 1-5 (132r)

ORATIO IOANNIS PICI MIRAN. CONCORDIAE COMITIS.

... Egi Patres Celestissimi in Arabi monumētis interrogati Abdā Abdala farra  
... qd i hac quasi Mundana scena admirādū maxie cenus  
... Nihil spectari hoie admirābilis respōdit. Cui frāe il  
... Mercurius  
... cogitā mihī nō latīs illa faciebāt; multa de huma  
... nature p̄lātā afferuntur a multis esse hominē creaturas; fieri unūq̄ lupis fa  
... Hominis di  
... gnat  
... Perfe  
... Magna Dauid  
... ubi ure unūq̄  
... Cur. n. nō ipsos angelos & beatissimos cali choros magis admitemur? Tādē  
... Intellexit mihī lum ulcus cur felicissimū; prōindeq̄ dignū ad admiratione aial fit  
... hō; Et quae fit demū illa cōditio quā i uniuersū ferre fortis sit, non brutis mō; sed  
... astris; sed ultramundanis mentibus inuoluam. Res supra fidem & mira. Quid  
... nī; Nā & p̄pa magnū miraculū & admirādū p̄fecto aial iure hō & dicitur & exi  
... flumatur. Sed quae nā ea sit audire patres & benignis aurbus p̄ ueltra humanita  
... te hanc mihī opam condonare. Iam fumus Pater architectus deus; hāc quā uidet  
... mus mundanā domū diuinitatis ep̄li augustissimū archana legibus sapientia la  
... lececeat. Supercedetlem regionē mentibus decorat; ethericos globos aeternis ani  
... mis uegetat; excrementis ac feculentis inferioris mundi partes omnigena aialū  
... turba cōpleat. Sed ope cōsumato; desiderabat artifex esse aliquid q̄ tātū rōnē  
... p̄deret; pulchritudinem amaret; magnitudinem admiraretur. Ideo iam rebus  
... eibus; ut Moses Timeuq̄; telantur abolutis; de p̄ducendo hoie postremo cogita  
... uit. Ves; nec erat in archetypis unde nouā fobolē effingeret; nec in thesauris; qd no  
... Timeus  
... uo filio hereditariū largiretur; nec in sublellis totius orbis; ubi uniuersū cōteplate  
... iste federet. Iam plena oia tota firmis; mediis; inlimiq̄ ordinibus fuerant distribu  
... ta. Sed nō erat parietū potelatis in extrema factura; quasi effera delictiffimē erat  
... sapientia; Conlitiū inq̄ in re nec essera fluduaife. Non erat benedicti amonitus  
... q̄ in aliis esset diuinā liberalitatē laudaturus in fellā dānare cogere. Statuit tū  
... dē optimus opifex; ut cui dari nihil p̄p̄tū poterat; cōe effer q̄q̄d priuatiū singulis  
... fuerat. Igitur hōiem accepit inchofreta opus imaginis; atq̄ i mūdū postū medital  
... ho; sic est alloquutus. Nec certā fedē nec p̄p̄riā lacit; nec munus ullū peculiaritē  
... bi dedimus o. Adātur quā fedē; quā faciē; quā mūnera tute optaueris; ea p̄ uoto;

<a href="http://www.brown.edu/Departments/Italian\_Studies/pico/incunab/dh03b.jpg" target="\_top"></a>

rep. 5. Man's task is the emulation of the highest angels (action, contemplation, communion with God) and dialectic is the necessary intermediate stage in this mystic process (Bori 1996, 557).

<sup>2</sup> This unusual phrase, in which God «broods», is glossed by Pico himself in

Oratio

http://www.brown.edu/Departments/Italian\_Studies/pico/prototipo/frame.html

Adressbuch HansischeStift ConfTool WebIR Univ Sfax BookManagelex Webkalender AILA 2008 Balea Lac Turkey Travel Guide E-Balance

Modello filologico Oratio Wort, Begriff, Name, Ter...

§ 11.

54. Sed qua ratione, aut quid tandem agentes?

55. Videamus quid illi agant, quam vivant vita

56. Eam si et nos vixerimus (possumus enim) equaverimus.<sup>a</sup>

57. Ardet Saraph charitatis igne; fulget Cherub stat Thronus iudicii firmitate.<sup>1</sup>

58. Igitur si actuosae<sup>2</sup> ad[d]icti vitae inferiorum curam recto examine susceperimus, Thronorum stata soliditate firmabimur.

59. Si ab actionibus feriat, in opificio opificem, in opifice opificium meditantes,<sup>3</sup> in contemplandi ocio<sup>b</sup> negociabimur, luce Cherubica undique coruscabimus.

60. Si charitate ipsum opificem solum ardebimus,<sup>4</sup> illius igne, qui edax est, in Saraphicam effigiem repente flammabimur.

inferior things with proper consideration of their worth, we will be strengthened by the steadfast solidity of the Thrones.

59. If we, unconcerned by actions, meditate on the Creator in His creation and creation in the Creator, we will be engaging in the leisure of contemplation; we will shine on all sides with Cherubic light.

60. If we burn for the Creator alone, with charity, with its all-consuming fire, we will burst into flame in the likeness of the Seraphim.<sup>1</sup>

<sup>a</sup> P *aequaverimus*

<sup>b</sup> P *otio*

<sup>c</sup> P *saeculorum*

<sup>d</sup> *anteluchanis hynnis* (Tognon, 12).

<sup>e</sup> P *delevit totam hanc sententiam*

<sup>1</sup> In accordance with the scheme laid out by Pseudo-Dionysius (cf. ¶ 73 of the *Oratio*) and confirmed by St. Thomas (*Summa* I, q. 108, art. 6) and St. Gregory (*In evang.* II hom. 34), Pico represents here the three divisions of the highest order of angels: the Seraphim, the Cherubim and the Thrones. These should not be confused, however, with the three types discussed in *Heptaplus* 3.3 since the Seraphim, Cherubim and Thrones are all members of the highest division, that which enjoys the leisure of contemplation and which is distinguished by a superabundance of goodness. The fact that Pico specifically calls the Cherubim contemplators does not mean that the Seraphim and Thrones are not. See St. Thomas, *Summa* I, q. 108, art. 5, rep. 5. Man's task is the emulation of the highest angels (action, contemplation, communion with God) and dialectic is the necessary intermediate stage in this mystic process (Bori 1996, 557).

<sup>2</sup> This unusual phrase, in which God «broods», is glossed by Pico himself in



Intro

Theorie

**Internet**

Dokument

TextVer

Lexikon

Syntax

Semantik

Texte

Korpus

# Eine Webseite „verzieren“: DHTML

- Webseiten mit drehenden Objekten, mit Besucherzählern, wandernder Schrift oder allerhand sonstigem Spielkram, aber auch sinnvollen erweiterten Funktionen für die Navigation, enthalten Skripts in der Programmiersprache JavaScript oder andere Programme. Global nennt man diese Erweiterungen DHTML (Dynamic HTML). Das ist nicht so einfach, es gibt aber Websites, die nützliche Bausteine mit narrensicheren Anleitungen zum herunterladen anbieten:

<http://javascript.internet.com/>

- Höchste Ansprüche kann man nur mit Java Applets der Programmiersprache Java befriedigen. Das ist aber wirklich schwer zu lernen.
- Alle Formulare, Ausfüllfelder, Knöpfe mit der entsprechenden Datenübertragungen etc. kann man zwar in HTML schreiben muß sie aber mit CGIs auswerten. Das ist Profi-Sache ...
- PHP Code wird direkt in HTML-Dateien notiert. Beim Aufruf führt zunächst der serverseitige PHP-Interpreter den Code aus und erzeugt daraus den endgültigen HTML-Code, der schließlich an den Browser gesendet wird.

# HTML 4/ CSS

- Seit HTML4 kann man systematisch den Inhalt eines Dokuments (den Text) und seine logische Gliederung („stylesheet“) trennen. Auf diese Weise kann man
  - existierenden Dokumenten (Texten oder HTML Dokumenten) nachträglich ein gleiches Aussehen zuweisen,
  - für weitere Dokumente zentrale Stilvorgaben formulieren, ohne Templates für Dokumente zu schreiben,
  - Inhalte in andere Dokumente und deren Stil übernehmen
  - Die Standard-Darstellung eines Browsers korrigieren
- Dieses Prinzip heißt CSS („Cascaded Style Sheet“)

# 3 Arten von Stylesheets

Style information können Sie

1. An Ort und Stelle in einem HTML-Ausdruck schreiben („direktformatieren“),
2. Im Kopf eines Files zentral für das File zusammenfassen,
3. In einem getrennten CSS-File von anderen HTML-Seiten aus benutzen.

Beispiel zu 1.:

```
<p style="background-color:#808040; color:#D8FD02;  
font-family:'Century Schoolbook',serif; font-size:24pt; letter-spacing:3px;  
padding:40px; border:double #D8FD02 4px;"  
title="Zitat von Francis Picabia">  
Unser Kopf ist rund, damit das Denken die Richtung wechseln kann!  
</p>
```

Alle Beispiele sind aus: <http://de.selfhtml.org>

# Zentrale Style Syntax

Syntax

`<STYLE> ... </STYLE>`

Attribute

TYPE=ContentType (Mime-Typ und Sprache)

man kann auch angeben MEDIA=MediaDesc (Medienspezifik)

TITLE=Text (title of style sheet)

Inhalt

ein stylesheet

Erscheint in

HEAD

---

TYPE ist meist „text/css“,

MEDIA kann sein: {screen, tty, tv, projection, handheld, print, braille, aural, all}

Man kann also für verschiedene Medien unterschiedliche Stylesheets machen.

# Zentrales HTML Stylesheet

Das Beispiel definiert den Stil für die h1 und h2 Überschriften und führt eine neue Klasse „chap“ ein:

```
<head>
  <title>CSS Example< /title >
  <style type="text/css">
    h1 { font-size: x-large; color: red }
    h2 { font-size: large; color: blue }
    .chap { font-family:Arial,sans-serif; font-size:20pt; color:blue;
            border-bottom-style:solid; border-bottom-width:3px;
            border-color:red; margin:0px; margin-bottom:16px; }
  < /style >
< head >
  <body> ...
  <div class="chap">Eine Kapitelüberschrift</div> ...
```

# Externes Stylesheet

Man kann das „link“-Element in <head> benutzen, um auf externe Stylesheets zu verweisen:

```
<html>
<head>
<title>Titel der Datei</title>
<link rel="stylesheet" type="text/css" href="formate.css">
<style type="text/css">
<!--
... hier sind zusätzlich auch datei-spezifische Formate erlaubt ...
-->
</style>
</head>
<body>
</body>
</html>
```

# Beispiel für externes css File

(Keine HTML-Syntax!)

```
#Kapitel {  
font-size:300%; text-transform:capitalize; line-height:80pt;  
}  
#Definition {  
border-width:3px 16px;  
border-style:dashed;  
border-color:red;  
margin-right:10px;  
padding-top:25px;  
padding-bottom:20px;  
padding-left:20px;  
padding-right:20px;  
text-align:justify;
```

Hier werden in einem File die Styles  
„Kapitel“ und „Definition“ definiert

# HyperCard-Beispiel

Inflection Generator 0.6.3

*German*  
**NatS** *Inflexion*  
*Generator*

Your input (Infinitive):

Part of Speech:

Number:

Case:

Stem:

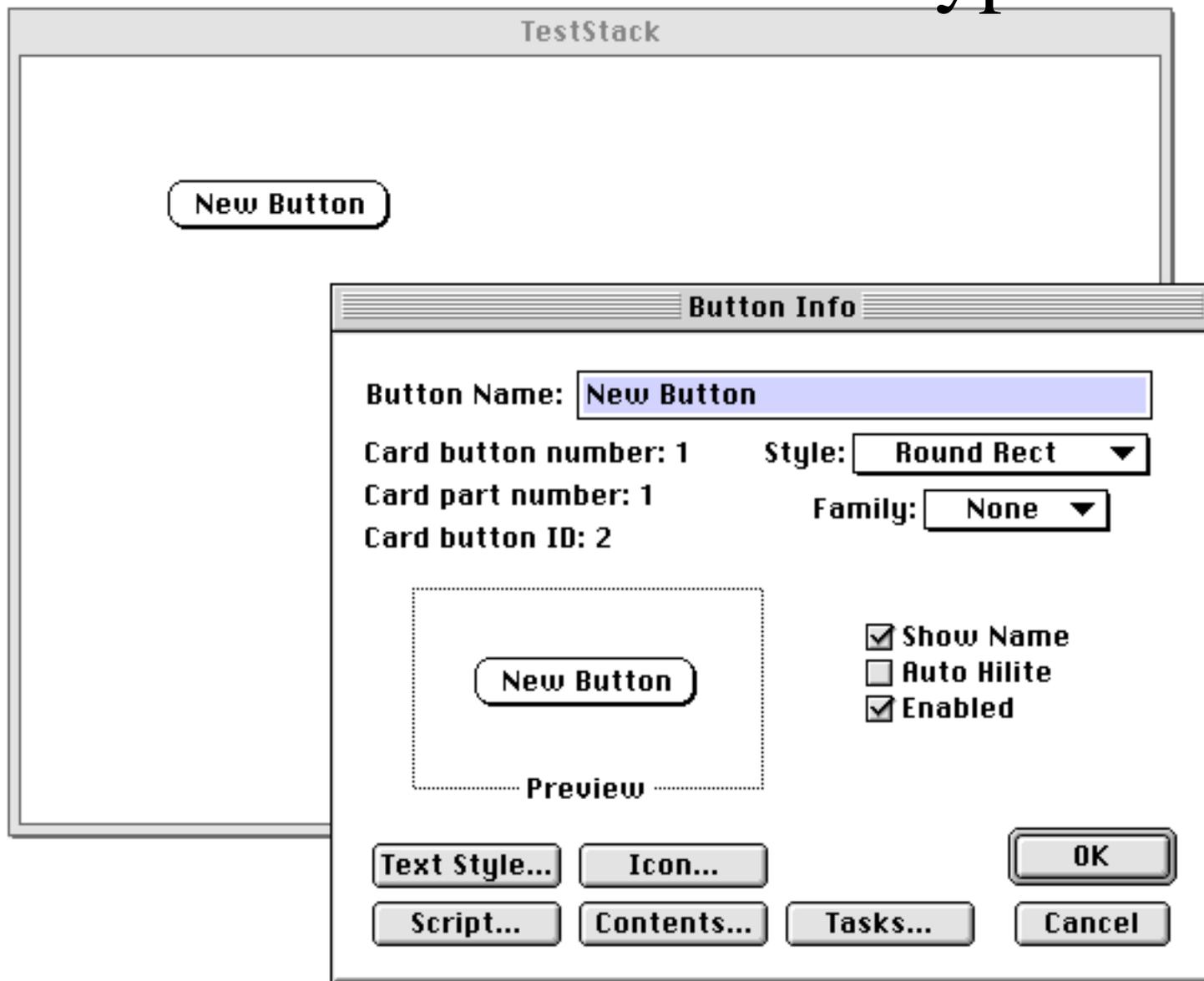
Pluralstem:

Gender:

Description:

The correct form is:

# HyperCard



# Hypercard

- Leicht zu lernen
- Ein guter Vorgeschmack für OOP
- Einfachste Definitionen von Objekten
- Nicht zu stark eingeschränkte Funktionalität
- Modular durch eingeschränkte Message-Kommunikation
- Automatische Unterstützung elementarer Funktionen
- Fünf vordefinierte Klassen



Stacks  
Cards  
Backgrounds  
Fields  
Buttons

# HyperCard Script (Sprache: Hypertalk)

on mouseUp

```
global wordString, slanguage, fromWhere, stopCounting, entryNr
show cd field "Collect Entry"
hide cd field "Short Entry Saved"
-- Concatenating all fields for the lexicon entry:
-- Prepare common header:
if the visible of cd field "Incomplete Entry" is true then
  put cd field "Incomplete Entry" into wordString
else
  put cd field "Main Entry" into wordString
end if
-- enter the word and the Id into the word list:
put wordString&&cd field "Entry Id"&return^
after last char of cd field id 1 of card
  "WordList_"&slanguage
sort cd field id 1 of card "WordList_"&slanguage
```

...

# Virtuelle Webseiten

Sind Webseiten, die man für eine Auskunft erhält, die aber von einem Programm und nicht von einem Web-Designer zusammengestellt wurden. D.h. ein Informations-Extraktionsprogramm schreibt die gefundene Information als (eine Art) HTML-Daten nieder und ein Designprogramm entscheidet (oft basierend auf Vorlagen) über die Anordnung und ggf. die Einbindung von Bildern.

- Noch selten angewandte Technologie

# Chatbots und Virtuelle Assistenten

- Chatbots sind Programme auf Webseiten, mit denen man sich in freier natürlicher Sprache unterhalten kann. Beispiele:
  - <http://www.talkingservant.com/> nur zum Plaudern
  - <http://www.medical-tribune.de/000drelectric/> verbunden mit Gesundheits-Information
  - Meistens haben sie nur ein Personenbild in wenigen Posen.
- Virtuelle Assistenten (gelegentlich auch Avatare genannt) beraten aufgabengebunden Internet-Benutzer oder potentielle Käufer. Die technische Skala reicht von Standbildern über gespeicherte Standardposen und Mimik bis hin zu aufwendigen Graphikprogrammen für natürlich wirkende fließende Bewegungen. Sie können Dialoge führen und sind teilweise ganz clever in der Gesprächsführung.

# Multimodalität,

- Webseiten bieten in der Regel nicht nur
  - Text, sondern auch
  - Graphiken/Bilder,
  - Bewegtbilder (Bildfolgen, Filme, Animation),
  - und Klänge.
- Diese Darstellungs- und Präsentationsmethoden sind teilweise leider durch die Werbung in Verruf geraten, können aber für computerphilologische Zwecke sehr wichtig sein, um zum Beispiel
  - einen Text, ein Video des Vorlesers und ein Video der Zuhörer synchron zu zeigen oder
  - für eLearning-Zwecke animierte Graphiken mit Ton in Lernobjekten zu zeigen.

# Multilingualität,

- Im Internet findet man Seiten in allen bekannten Sprachen, aber
- Internet-Übersetzungen (z.B. Babel/Altavista) sind bisher noch kümmerlich:
- Bei eigenen Seiten sollte man unbedingt, wenn man im Ausland gelesen werden will, eine englische Alternative anbieten. Auf die deutsche Seite sollte man trotzdem nicht verzichten.

## Vorträge, Kurse und Seminare

SommerBezeichnung '01

WinterBezeichnung ' 00/'01

SommerBezeichnung '00

WinterBezeichnung ' 99/00

SommerBezeichnung '99

ARMKREUZ : Öffnen Sie die Arbeitsgruppe, die die Steuerung der Roboter durch natürliche Sprache beschäftigt (auf Deutsch)

# Web-Übersetzung?

## Vorträge, Kurse und Seminare

SommerBezeichnung '01

Winterbezeichnung '00/'01

SommerBezeichnung '00

Winterbezeichnung '99/00

SommerBezeichnung '99

ARMKREUZ : Öffnen Sie die Arbeitsgruppe,  
die die Steuerung der Roboter durch natürliche  
Sprache beschäftigt (auf Deutsch)

## Lectures, Courses, and Seminars

Summer Term '01

Winter Term '00/'01

Summer Term '00

Winter Term '99/00

Summer Term '99

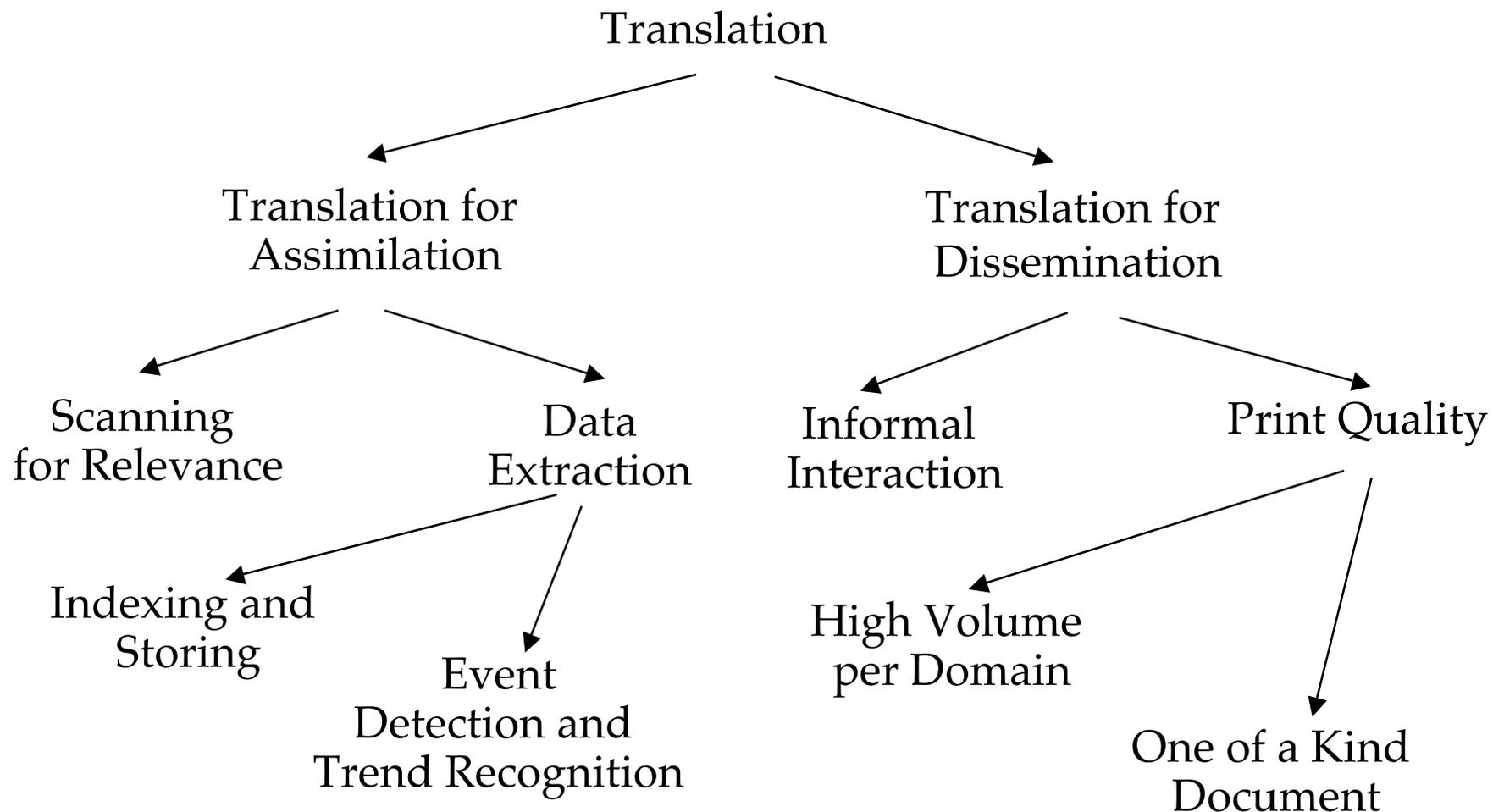
SPIDER: Open working group  
dealing with the control of  
robots by natural language (in  
German)

# Maschinelle Übersetzung

Stand der Kunst:

- Eine größere Zahl von Produkten ist auf dem Markt. Auch als Web-Dienste werden sie angeboten. Diese MT-Systeme sind allerdings entweder auf ein enges Sachgebiet und bestimmte Textsorten eingeschränkt oder qualitativ nur zur Orientierung über den Inhalt geeignet. Manche standalone-Systeme kann man gut trainieren und dann werden sie immer besser.
- Die nächste Technologie-Generation sind speech-to-speech-Systeme zum automatischen Dolmetschen (Noch nicht produktreif)
- Zum Ausprobieren:
  - <http://www.freetranslation.com>      Enterprise Translation Server
  - <http://www.wordlingo.com>      Wordlingo

# Typology of MT-Systems (Carbonell 1995)



# Übersichtsliteratur zu MT im Web

- "Machine Translation - an Introductory Guide", D. Arnold, L. Balkan, NCC Blackwell London, 1994  
[\(http://clwww.essex.ac.uk/MTbook/PostScript/\)](http://clwww.essex.ac.uk/MTbook/PostScript/)
- "A Survey of current Paradigms in Machine Translation", B. J. Dorr, P. W. Jordan, J. W. Benoit, 2000 (<http://www.umiacs.umd.edu/~bonnie/>)
- Unter EAMT (European Association of Machine Translation) nachsehen für die neuesten Entwicklungen, Produkte und websites
- Produktübersicht:
  - <http://www.word2word.com/free.html>

# Fachrelevante Suchmaschinen

- Die Technischen Fächer haben Informationszentren (z.B. FIZ Karlsruhe) und Datenbanken zum Recherchieren.
- Alle geisteswissenschaftlichen Fächer haben (eine oder mehrere) zentrale Sites entwickelt, die eine Vielzahl weiterer brauchbarer Links pflegen.
  - Für die CP ist dies z.B. <http://ilex.cc.kcl.ac.uk/wlm/hc/>
  - Für Gesprächsforschung: <http://www.gespraechsforschung.de/technik.htm> (sehr informativ!)
  - Auf der Website des IfGI finden Sie die Links-Sammlung: <http://www.sign-lang.uni-hamburg.de/fb07/GermS/germ-links.html>
  - Es gibt neuerdings Zusammenstellungen z.B. für Literaturwissenschaft auf CD:

Baasner, Rainer ; Koebe, Kristina: Tools fürs Literaturstudium. wozu. was. wie. Literaturrecherche und Internet. Stuttgart 2000

# Nl Interfaces

- Natürlichsprachliche (nl) Interfaces zu Leistungen des Internet sind noch relativ selten. Es herrscht immer noch
  - die Schreibtischmetapher,
  - die Interaktion über Maus, Buttons, Menüs und Textfelder.

# Texteingabe und Analyseanforderungen

## Eingabe

- Stichwörter
- Suchphrasen
- Restricted Language
- Freitext

## Analyse

- Keyword spotting
- Pattern matching
- Restricted language analysis
- Komplette linguistische Analyse