

## Encoding Basics

Stephen Miller  
Texts & Files: Using Corpora and Digital  
Archives / Editions  
Institut für Germanistik, Universität Wien  
SS 2008 /100118

---

---

---

---

---

---

---

---

## Descriptive Markup

```
<div1 type="chapter">  
<head type="main">Seasons and Months</head>  
<div2 type="section">  
<head type="sub">The New Year</head>  
<p>  
<hi rend="strong">I.</hi> Manx, Yn Vlein Noa. <hi rend  
="emphasis">The New Year</hi>. January 1st. &ldquo;&hellip;  
anciently the first day of November was the first day of the  
year: and the Mummings on the eve of All Saints&apos; Day still  
begin their petition with these remarkable words, &lsquo;Fo-  
night is New Year's night. Og-u-naa. The Moon shines fair and  
bright. Tro-la-la.&rsquo;&rdquo; Kelly (<hi  
rend="emphasis">circa</hi> 1790). S.v. Saal-saun. <hi  
rend="emphasis">See</hi> November 1st.  
</p>  
</div2>
```

(TEILite DTD)

---

---

---

---

---

---

---

---

## How is Content Encoded?

- Elements
  - Attributes
    - Values
- Character Entities
- Character Data
- Non Character Data
  - Notations
- Comments

---

---

---

---

---

---

---

---

## Elements

- Format  
    <element\_name>  
    text  
    </element\_name>  
    Element consists of a pair of tags (start/end) surrounding text
- Elements can be "empty"  
    <element\_name/>

---

---

---

---

---

---

---

---

## Attributes & Values

- Elements can have attributes
- Attributes have values
- <pb n="12"/>
  - element pb
    - empty element
  - attribute n
  - value of 12

---

---

---

---

---

---

---

---

## Attributes

- Global
- Specific

### Values:

- Enforced
- Optional

---

---

---

---

---

---

---

---

### Character Entities

- Accented characters, symbols,
- Encoded using Unicode
  - ß &#xDF;
  - „ &#x201E;
- Unicode Standard 4.0  
[www.unicode.org](http://www.unicode.org)

---

---

---

---

---

---

---

---

### Notations

- Handle for the handling of non character data (eg, images)
- To reference JPEG files:

<!NOTATION jpeg PUBLIC "ISO DIS 10918//NOTATION JPEG Graphics Format//EN" >

---

---

---

---

---

---

---

---

### Example

- Entity declared such as:
  - <ENTITY name SYSTEM "bf01fig01.jfg" NDATA jpeg>
- Entity used in text:
  - <figure entity="name"/>

---

---

---

---

---

---

---

---

## Comments

- Can be inserted anywhere in the text  
    <!-- comment text -->

---

---

---

---

---

---

---

---