Standardi in oznaèevanje

SGML: Standard Generalized Markup Language

ISO standard: metode za predstavitev (strukture) teksta v elektronski obliki.

Zagotavlja *trajnost* in *izmenljivost*.

Primer - HTML dokument:

```
<h1 align="center">Standardi in oznaèevanje</H1>
<HR>

<H2>SGML: Standard Generalized Markup Language</H2>

ISO standard: metode za predstavitev (strukture) teksta v elektronski
obliki.
<P>
Zagotavlja <em>trajnost</em> in <em>izmenljivost</em>.
```

TEI: Text Encoding Initiative

Ukvarja se z zapisom in ozna"cevanjem strukture besedil predvsem za namene znanstvene obravnave jezika.

TEI P3 (1200 str.) (1987 -- 1994)

Primeri:

- Skladenjska analiza:

```
<p>
<cl type='finite declarative'
    function='independent'>
 <phr type=NP function='subject'>
    Nineteen fifty-four,
  <cl type='finite relative declarative'
    function='appositive'>
    when
   <phr type=NP function='subject'>
    I
   </phr>
   <phr type=VP function='predicate'>
     was eighteen years old
    </phr>
  </cl>
  ,
```

- Vzorec za glavo radijskega poroèila:

```
<sourceDesc>
 <scriptStmt id=CNN12>
 <bibl><author>CNN Network News</>
    <title>News headlines</>
    <date>12 Jun 1991</>
 </bibl>
 </scriptStmt>
 <!-- this script statement might be
      used to document the parts of a
      spoken transcript which included
      a news broadcast -->
</sourceDesc>
```

CES: Corpus Encoding Standard

Ukvarja se z zapisom in ozna"cevanjem strukture korpusov, predvsem za namene raèunalni¹kih orodij.

MULTEXT-East korpus:

	Orwell			**Leposlovje**			**Èasopisi**		
En:	D	H	C						
Bg:	D	H	C	D	H	C	D	H	C
Cs:	D	H	C	D	H	C	D	H	C
Et:	D	H	C	D	H	C	D	H	C
Hu:	D	H	C	D	H	C	D	H	C
Ro:	D	H	C	D	H	C	D	H	C
Sl:	D	H	C	D	H	C	D	H	C

Metoda

podatki + standardi + orodja + jezikoslovec = korpus

[nazaj]
