

Markup, XML och namnrymder

Workshop för Svensk Nationell Datatjänst



BIBLIOTEKSHÖGSKOLAN
HÖGSKOLAN I BORÅS

Mikael Gunnarsson

28 april, 2017

BHS — Borås

Översikt

- 1 Markup
- 2 XML
- 3 Namnrymder

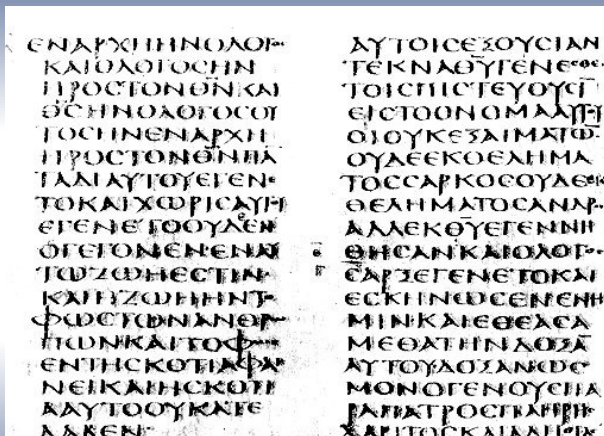
Markup - typologi

- Interpunktion
 - Presentationsorienterad
 - Procedurell
 - Beskrivande
- + Referentiell markup; Metamarkup

Coombs, J. H., Renear, A. H., & DeRose, S. J. (1987). Markup systems and the future of scholarly text processing. *Communications of the ACM*, 30(11), 933–947.



Scriptio continua



Figur: Scriptio continua från Codex Sinaiticus, 300 AD

Interpunktion och “white space”

“Jag heter Mikael”
Vad heter du?

«Je m'appelle Mikael»
¿cómo te llamas?

Interpunktion och “white space”

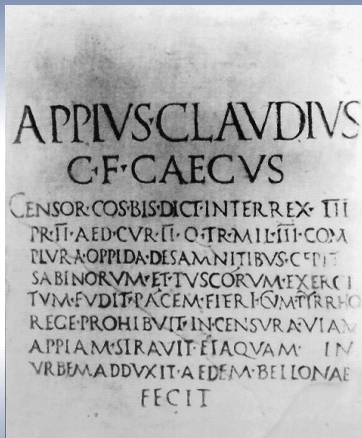
“Jag heter Mikael”
Vad heter du?

«Je m'appelle Mikael»
¿cómo te llamas?

En sjuk system
A web site

En sjuksystem
A website

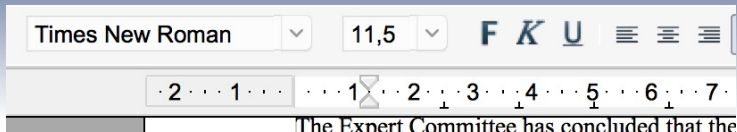
Tidiga tecken på interpunktion



Figur: Romersk inskription

notera punkterna och apex (den akuta accenten) för att markera lång vokal

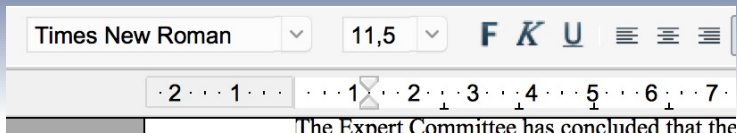
Presentationsorienterad markup



Figur: Funktionalitet för visuellt orienterad formatering i ordbehandlingsbehandling

Om typsnittet behöver ändras från Times New Roman till Verdana, storleksförhållandet mellan brödtext och rubriker behöver justeras, citations-tecknens form behöver förändras ...???

Presentationsorienterad markup



Figur: Funktionalitet för visuellt orienterad formatering i ordbehandlingsbehandling

Om typsnittet behöver ändras från Times New Roman till Verdana, storleksförhållandet mellan brödtext och rubriker behöver justeras, citations-tecknens form behöver förändras ... ???

Visuell form är sällan otvetydig. Vi vet inte alltid **varför** ett ord, en fras eller ett stycke är kursiverat — har givits en viss form?

Procedurell markup

Distinktionen mellan presentationsorienterad och procedurell markup är inte skarp, men den senare kan definieras som **instruktioner** till typsätaren (mänsklig eller icke-mänsklig aktör) för hur text ska bearbetas.

Procedurell markup

Distinktionen mellan presentationsorienterad och procedurell markup är inte skarp, men den senare kan definieras som **instruktioner** till typsätaren (mänsklig eller icke-mänsklig aktör) för hur text ska bearbetas.

```
\begin{center} Markup, XML och namnrymder \\  
Workshop för Svensk Nationell Datatjänst  
\end{center}
```

Deskriptiv markup

`\section{Markup, XML och
namrymder}`

`\label{sec:greetings}`

`\par` Lorem ipsum dolor sit amet,
consectetuer ...

`\par` Nam dui ligula, fringilla a,
euismod sodales ...

Figur: \LaTeX markup

`<h1>{Markup, XML och
namrymder</h1>`

`<p>`

Lorem ipsum dolor sit amet,
consectetuer ...

`</p> <p>` Nam dui ligula, fringilla a,
euismod sodales ... `</p>`

Figur: HTML markup

Deskriptiv markup, forts.

- “Ren” beskrivande markup:
Skribenten beskriver titel, kapitel, stycken, titelsidor, **rad- och sidbrytningar** mm i en existerande fysisk utgåva (ex: TEI)

Deskriptiv markup, forts.

- “Ren” beskrivande markup:
Skribenten beskriver titel, kapitel, stycken, titelsidor, **rad- och sidbrytningar** mm i en existerande fysisk utgåva (ex: TEI)
- Performativ markup:
Skribenten deklarerar titel, avsnitt, stycken, titelsidor, citat, ekvationer mm i en text i tillblivelse (ex: HTML)

I det första fallet är det meningsfullt att fråga om uppmärkningens är korrekt eller inte, i det andra fallet inte.

Sammanfattning

- Markup (oavsett typ) tjänar syftet att strukturera en mängd data genom att ge dess olika komponenter “etiketter” som indikerar vilken typ av data vi har att göra med, som avslöjar dess semantik

Sammanfattning

- Markup (oavsett typ) tjänar syftet att strukturera en mängd data genom att ge dess olika komponenter “etiketter” som indikerar vilken typ av data vi har att göra med, som avslöjar dess semantik
- Men sekventiell text är inte den enda typen av data vi har att göra med som behöver struktur ...

Exemplet bibliografiska data

```
@book{
Persson2011ggb,
author={Persson, Leif G. W.},
title={Gustavs grabb : ber{\"}ttelsen om min klassresa},
publisher={Bonnier},
address={Stockholm},
year={2011},
ISBN={978-91-0-012684-1 (inb.)},
}
```

Figur: BibT_EX-post

MARC-XML

```
<?xml version="1.0" encoding="UTF-8"?>
  <record xmlns="http://www.loc.gov/MARC21/slim">
    <datafield tag="020" ind2=" " ind1=" ">
      <subfield code="a">978-91-0-012684-1 (inb.)</subfield>
    </datafield>
    <datafield tag="100" ind2=" " ind1="1">
      <subfield code="a">Persson, Leif G. W.</subfield>
    </datafield>
    <datafield tag="245" ind2="0" ind1="1">
      <subfield code="a">Gustavs grabb :</subfield>
      <subfield code="b">berättelsen om min klassresa </subfield>
      <subfield code="c">Leif GW Persson</subfield>
    </datafield>
    <datafield tag="260" ind2=" " ind1=" ">
      <subfield code="a">Stockholm :</subfield>
      <subfield code="b">Bonnier</subfield>
    </datafield>
  </record>
```

Vokabulärer

Vad är det här?

```
<?xml version="1.0"encoding="UTF-8"?>  
<record xmlns="http://www.loc.gov/MARC21/slim" >  
  <datafield tag="020"ind2= ind1=>
```

...

Vokabulärer

Vad är det här?

```
<?xml version="1.0"encoding="UTF-8"?>  
<record xmlns="http://www.loc.gov/MARC21/slim" >  
  <datafield tag="020"ind2= ind1=>  
    ...
```

En xml-namnrymndsdeklaration!

En sådan knyter ett välformat xml-dokument till en särskild vokabulär (lexikon + grammatik)

Formalismer för vokabulär-kontroll

- Document Type Definition (DTD)
- XML Schema
- Relax NG
- Schematron

Exempel

TEI

```
<?xml version="1.0"encoding="UTF-8"?>  
<TEI xmlns="http://www.tei-c.org/ns/1.0">  
...
```

Exempel

TEI

```
<?xml version="1.0"encoding="UTF-8"?>  
<TEI xmlns="http://www.tei-c.org/ns/1.0">  
...
```

RDF

```
<?xml version="1.0"encoding="UTF-8"?>  
<rdf:RDF  
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">  
...
```

XML is not the end

- XSL: XPath, XSL Transformations, XSL Formatting Objects
- XQuery

<https://github.com/Mikael61/SND>