

IPIS 2

Michal Kvasnica

O predmete

- Prednášky:
 - utorok 8:00 - 10:00, CH17
 - 29.9.2009 bude TAR3
- Cvičenia:
 - streda, 2h, 638, rozdelenie na 2 skupiny
 - Janka Závacka, Rado Paulen

O predmete

- Ukončenie predmetu:
 - zápočet = aktívna účasť na cvičeniach
 - skúška = 50% veľká písomka v polke semestra + 50% posemestrový projekt

Náplň predmetu

- Formát XML
 - základná filozofia ilustrovaná na príklade HTML
 - elementy, atribúty
 - všeobecná štruktúra XML dokumentu
- DTD: validácia
- XPATH: vyhľadávanie
- XSLT: transformácie

HTML

- Hypertext Markup Language
- “Táto **veta** *obsahuje* formátovanie.”
- `<p>Táto veta obsahuje formátovanie</p>`
- HTML: tagy sú predefinované a vždy znamenajú to isté (napr. `...`)

HTML: syntax

Elementy

- Element = **otváračací tag** + obsah + **zatváračací tag**
- “Táto **veta** *obsahuje* formátovanie.”
- `<p>Táto veta obsahuje formátovanie</p>`
- `<p>Táto veta obsahuje formátovanie</p>`

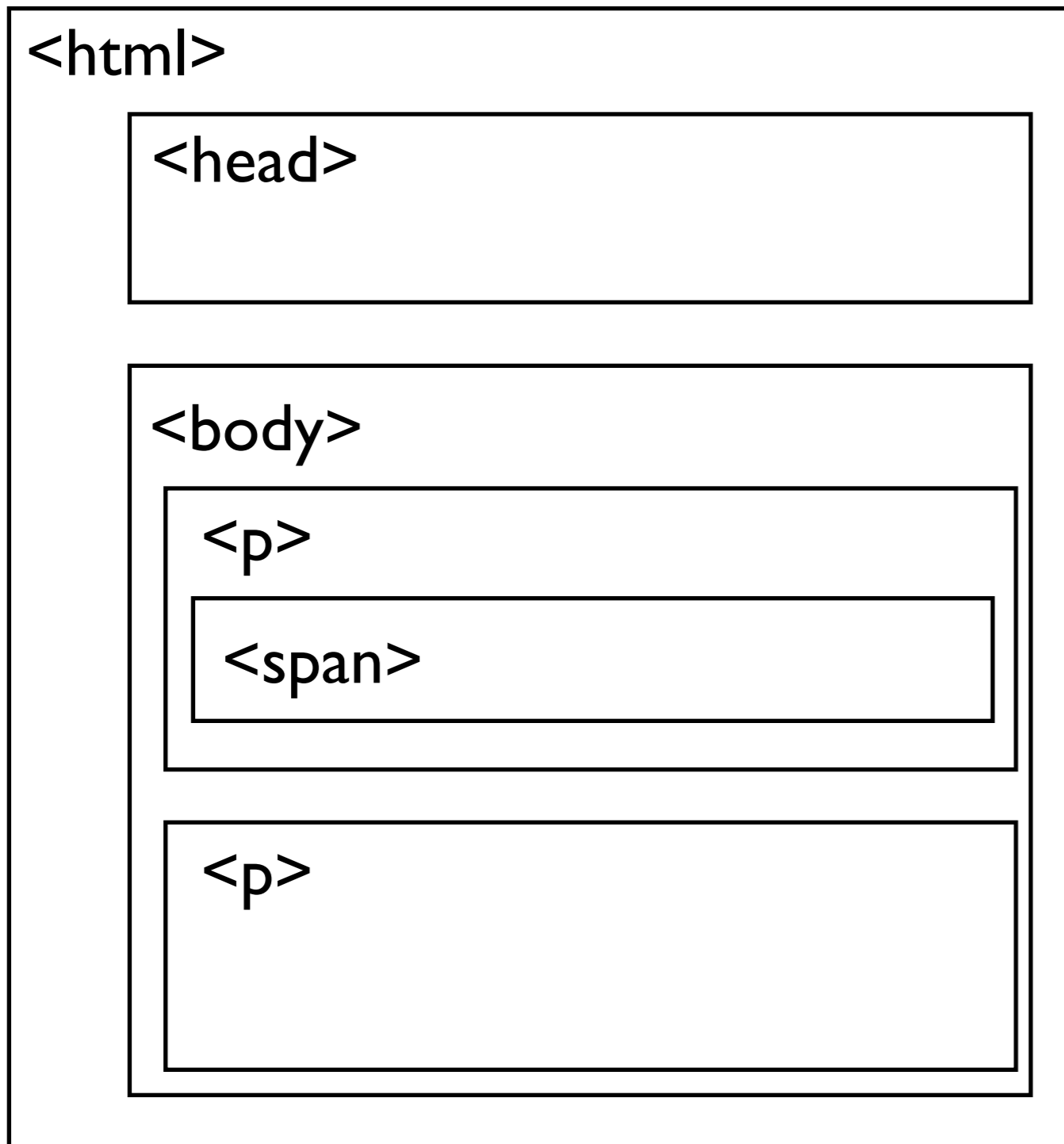
Elementy

- Element = **otváračací tag** + **obsah** + **zavírací tag**
- Otváračací a zavírací tag musia byť zhodné:
 - OK: `<p> Obsah elementu </p>`
 - CHYBA: `<p> Obsah elementu </q>`
- Obsah je nepovinný:
 - OK: `<p> Obsah elementu </p>`
 - OK: `<p></p>`

Elementy

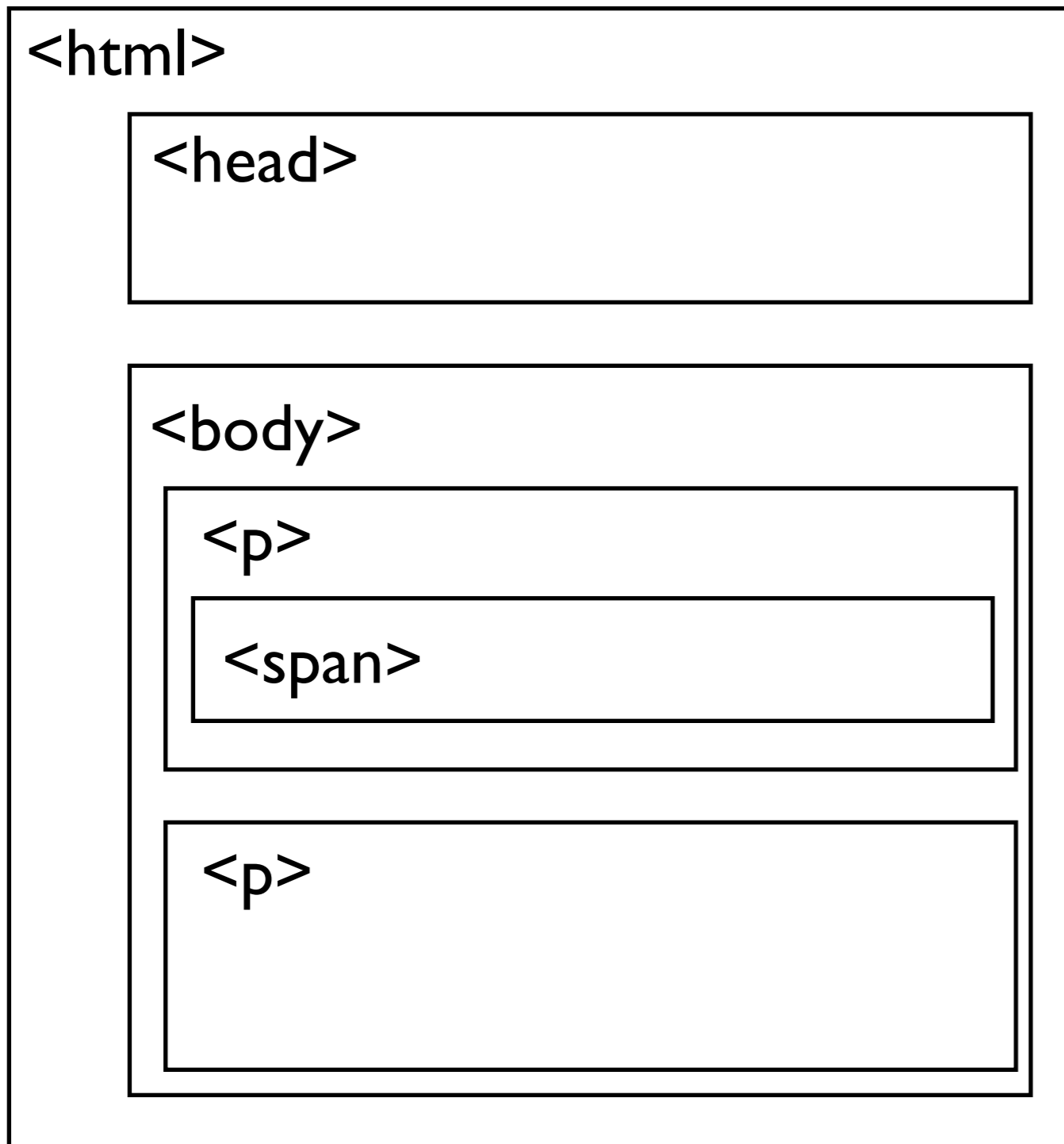
- Element = **otváračací tag** + **obsah** + **zavírací tag**
- Otváračací tag může obsahovat atribúty:
 - ``
 - definují dodatočné vlastnosti elementu
- Všeobecný formát:
 - `<tag nazov_atributu = "hodnota atributu">`
 - `<tag atr1="val1" atr2="val2" ...>`

Stromová organizácia elementov



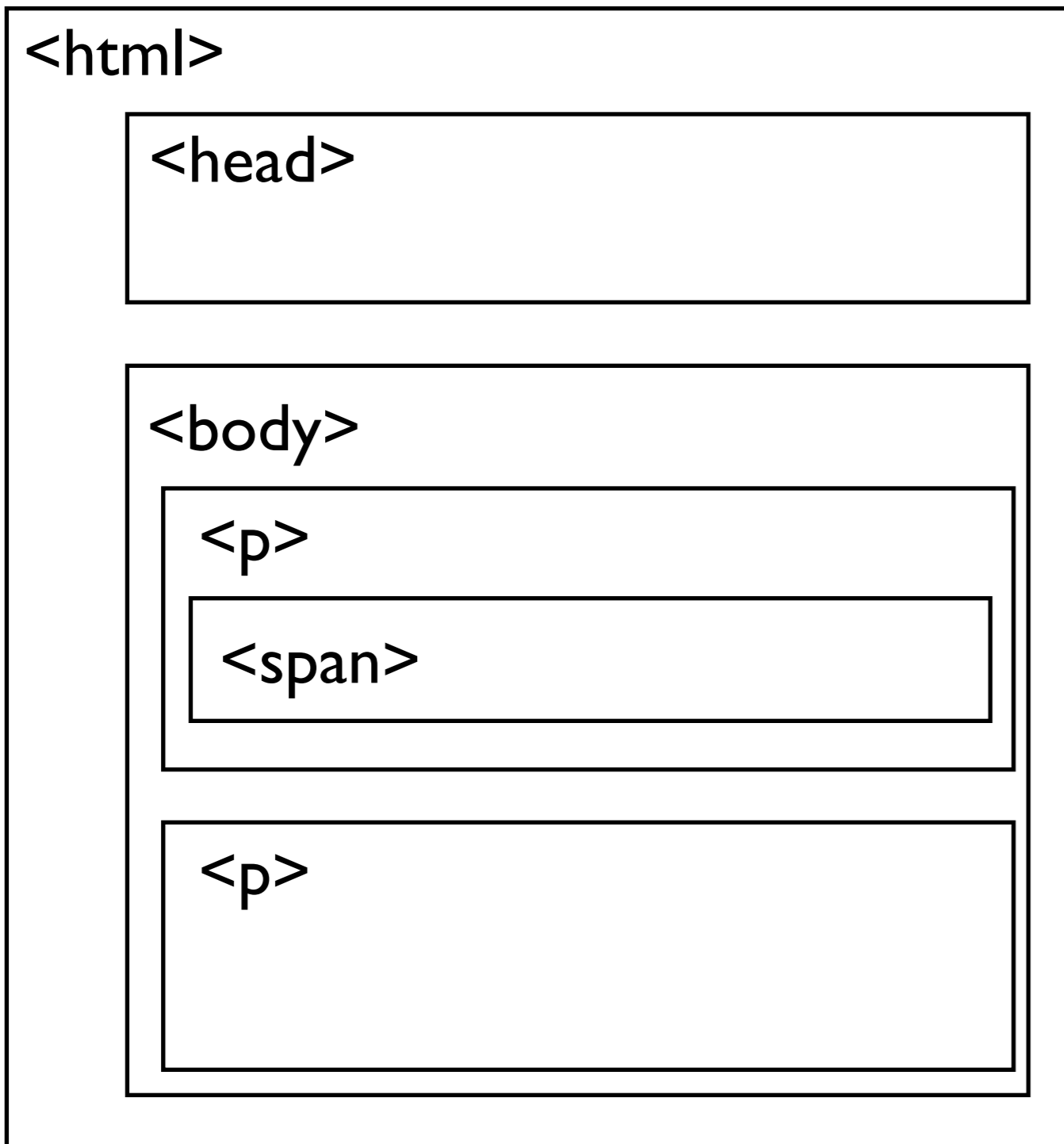
```
<html>
  <head>
</head>
  <body>
    <p>
      <span>
    </span>
    </p>
    <p>
    </p>
  </body>
</html>
```

Stromová organizácia elementov



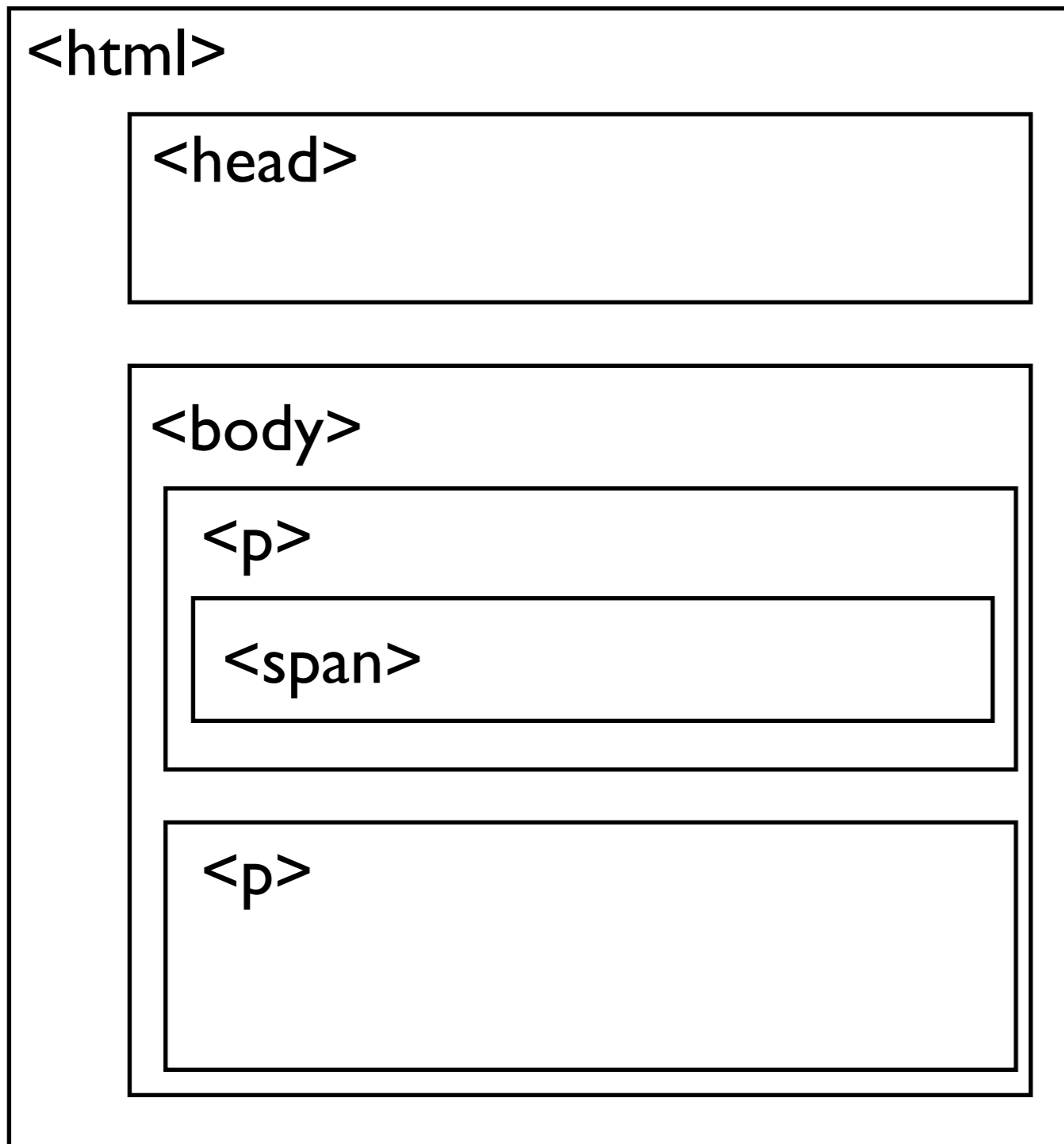
```
<html>
  <head>
</head>
  <body>
    <p>
      <span>
    </span>
    </p>
    <p>
    </p>
  </body>
</html>
```

Stromová organizácia elementov



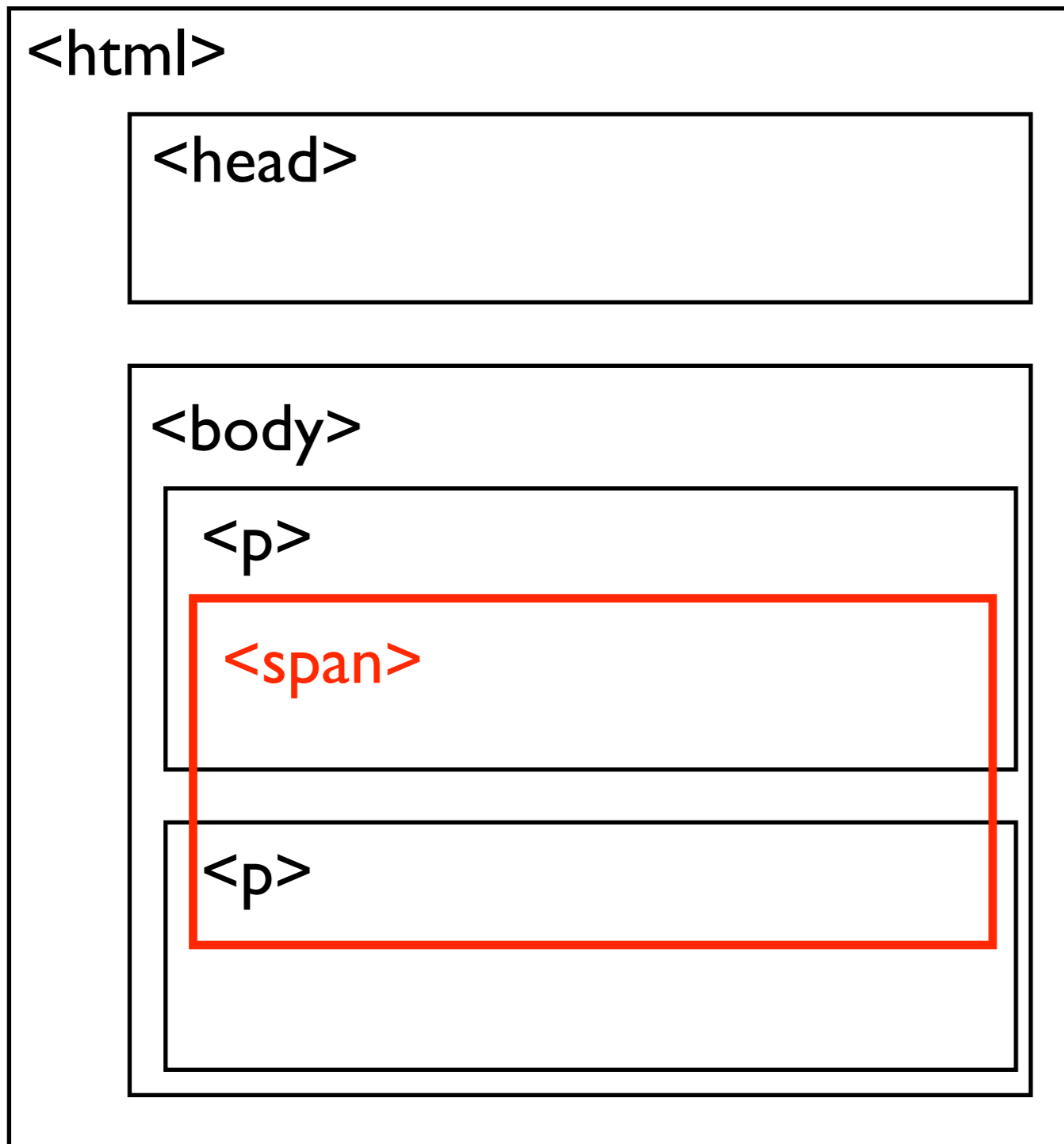
```
<html>
  <head>
</head>
  <body>
    <p>
      <span>
    </span>
    </p>
    <p>
    </p>
  </body>
</html>
```

Stromová organizácia elementov



```
<html>
  <head>
</head>
  <body>
    <p>
      <span>
    </span>
    </p>
    <p>
    </p>
  </body>
</html>
```

Nedovolená organizácia elementov



```
<html>
  <head>
</head>
  <body>
    <p>
      <span>
    </p>
      </span>
    <p>
  </p>
</body>
</html>
```

XML

=

eXtensible Markup Language

XML elementy

- XML je ako HTML s vlastnými elementami
- “Táto **veta** *obsahuje* formátovanie.”
- HTML vs XML:
 - `<p>Táto veta obsahuje formátovanie</p>`
 - `<odsek>Táto <farebnyBold farba="red">veta</farebnyBold> <sikmo>obsahuje</sikmo> formátovanie</odsek>`

XML

- XML je textový formát zápisu ľubovoľných dát
 - používa stromovo organizované elementy
 - je porozumiteľný pre človeka aj pre počítač

```
<clovek>  
  <meno>James</meno>  
  <priezvisko>Bond</priezvisko>  
  <telefon>  
    <mobil>0907-777777</mobil>  
    <fax></fax>  
  </telefon>  
</clovek>
```

XML

- XML je jednoducho rozšíriteľný

```
<clovek>
  <meno>James</meno>
  <priezvisko>Bond</priezvisko>
  <titul>Mgr</titul>
  <telefon>
    <mobil>0907-777777</mobil>
    <fax></fax>
  </telefon>
</clovek>
```

Použitie XML

- XML je priemyselný štandard výmeny akýchkoľvek informácií
- Je používaný takmer všade:
 - malá škála: správy zo senzorov do regulátorov
 - veľká škála: opis sekvencie génov
- Umožňuje rôzne spôsoby spracovania:
 - veľmi rýchle spracovania (časovo kritické aplikácie)
 - spracovanie nenáročné na pamäť
 - inkrementálne spracovanie (streaming)

Čo je a čo nie je XML

- XML je:
 - vynikajúci formát na výmenu štrukturovaných dát vo forme stromov
 - nezávislé od platformy
 - základ iných technológií (XHTML, AJAX, SVG)
- XML nie je:
 - programovací jazyk (už ste programovali v CSV?)
 - sám osebe nedáva dátom význam

Kedy použiť XML

- Ked' chcete vymieňať dáta:
 - a ak chcete použiť formát, ktorý je všeobecne používaný a ľahko použiteľný
 - a ak chcete umožniť iným (ľuďom, programom) jednoducho spracovať vaše dáta
- Ked' chcete lacno vymieňať dáta:
 - je jednoduchšie použiť XML ako vymýšľať niečo nové
- Ked' chcete vymieňať dáta v otvorenom formáte

XML vs iné formáty

- Život pred XML vyžadoval:
 - ručne definovať štruktúru pre každý nový typ dát
 - spracovanie takýchto dát vyžadovalo separátne nástroje

```
@misc{xml10fourth,  
  author = "Tim Bray and Jean Paoli and C. Michael  
           Sperberg-McQueen and Eve Maler and  
           Fran\c{c}ois Yergeau",  
  title = "Extensible Markup Language (XML) 1.0  
          (Fourth Edition)",  
  month = aug,   year = 2006,  
}
```

XML verzia: zlý nápad

```
@misc{xml10fourth,  
  author = "Tim Bray and Jean Paoli and C. Michael  
            Sperberg-McQueen and Eve Maler and  
            Fran\c{c}ois Yergeau",  
  title = "Extensible Markup Language (XML) 1.0  
           (Fourth Edition)",  
  month = aug,   year = 2006,  
}
```

<bibtex>

```
@misc{xml10fourth,  
  author = "Tim Bray and Jean Paoli and C. Michael  
            Sperberg-McQueen and Eve Maler and  
            Fran\c{c}ois Yergeau",  
  title = "Extensible Markup Language (XML) 1.0  
           (Fourth Edition)",  
  month = aug,   year = 2006,  
}
```

</bibtex>

XML verzia: správne použitie

```
@misc{xml10fourth,  
  author =      "Tim Bray and Jean Paoli and C. Michael  
                Sperberg-McQueen and Eve Maler and  
                Fran\c{c}ois Yergeau",  
  title =      "Extensible Markup Language (XML) 1.0  
                (Fourth Edition)",  
  month =      aug,      year =      2006,  
}
```

```
<bibtex typ="misc">  
  <label>xml10fourth</label>  
  <author> <name>Tim</name> <surname>Bray</surname> </author>  
  <author> <name>Jean</name> <surname>Paoli</surname> </author>  
  <title> Extensible Markup Language (XML) 1.0 (Fourth  
Edition)</title>  
  <published> <month>aug</month> <year>2006</year> </published>  
</bibtex>
```


XML: Elementy

- Element = **otvárací tag** + **obsah** + **zatvárací tag**
- HTML: fixné elementy, XML: variabilné
- Názov tagu môže byť takmer ľubovoľný:
 - OK: <kniha>, <krstneMeno>, <Element_1>, ...
 - CHYBA: <1.kniha>, <krstne Meno>, <Element&1>

XML: Špeciálne znaky

- Element = **otvárací tag** + **obsah** + **zatvárací tag**
- `<vyrok> Pre x=1 a y=2 plati x < y </vyrok>`
 - Problém pri spracovaní! Program by si myslel, že otvárame nový tag
- `<elem attr="Jednoducha ' alebo dvojita "
uvodzovka">`
 - nie je dovolené mixovať ' s "
 - nie je dovolené predčasne ukončiť úvodzovky

XML: Špeciálne znaky

- Element = **otvárací tag** + **obsah** + **zatvárací tag**
- Mapovacia tabuľka špeciálnych znakov:

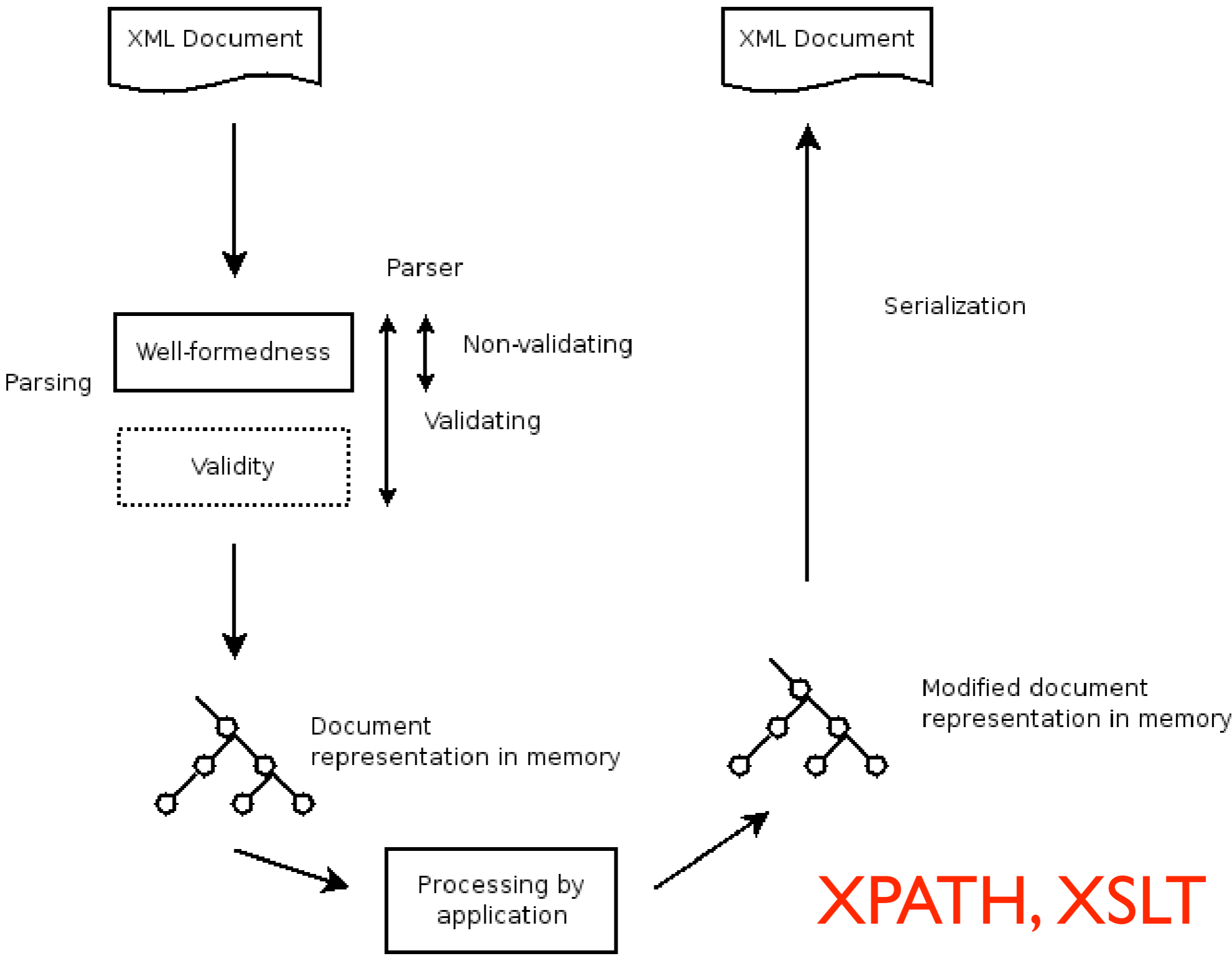
<	<
>	>
&	&
'	'
“	"

XML: Atribúty

- Môžu byť ľubovoľne definované
- Príklady:
 - `<priemer jednotka="km">12756</priemer>`
 - `<telefon typ="mobil">0907-777777</telefon>`
- Vždy sa dajú nahradiť elementami:
 - `<priemer> <km> 12756 </km> </priemer>`

Spracovanie XML

DTD



XPATH, XSLT

DTD = Document Type Definition

- DTD definuje “model dokumentu”
 - zaisťuje, že XML dáta spĺňajú ohraničenia
 - zabráňuje vloženiu “zlých” informácií

```
<zaznam>  
  <meno>povinné</meno>  
  <priezvisko>povinné</priezvisko>  
  <telefon>voliteľné  
    <mobil>bud'/alebo, číslo</mobil>  
    <fax>bud'/alebo, číslo</fax>  
  </telefon>  
</zaznam>
```

XPATH

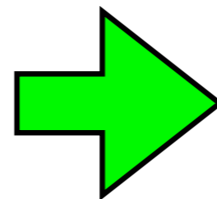
- Spôsob navigácie v dokumente
 - ako zistíme číslo knihy alebo jej titul?

```
<kniha>  
  <cislo>581</cislo>  
  <autor>  
    <meno>Janko Mrkvicka</meno>  
    <titul>Ing</titul>  
  </autor>  
  <titul>O Jankovi Mrkvickovi</titul>  
  <kapitola><cislo>1</cislo></kapitola>  
  <kapitola><cislo>2</cislo></kapitola>  
</kniha>
```

XSLT

- Spôsob transformácie XML dokumentu z jednej štruktúry na inú

```
<zaznam>
  <meno>James</meno>
  <pr>Bond</pr>
  <telefon>
    <mobil>0907</mobil>
    <fax></fax>
  </telefon>
</zaznam>
```



```
<h2>Bond, James</h2>
<b>Mob:</b> 0907 <br/>
<b>Fax:</b> nema <br/>
```


Príklady na cvičeniach

- Vytvorte XML dokument opisujúci telefónny zoznam, naplňte ho dátami a zvalidujte výsledok

```
zoznam
  clovek
    meno
    priezvisko
    adresa
      ulica
      cislo_domu
    mesto
    telefon
    mobil
    fax
```

Príklady na cvičeniach

- Nájdite a opravte chybu

```
<zoznam>
  <clovek>
    <meno>
      Janko
    </clovek>
  <clovek>
    </meno>
    <priezvisko>
      Mrkvicka
    </clovek>
</zoznam>
```

Príklady na cvičeniach

- Navrhните vlastnú štruktúru XML dokumentu na zápis dát z prechodovej charakteristiky

```
system typ="prenos"  
  citatel: 1  
  menovatel: [1 1]  
data  
  t          x  
  0          0  
  0.1060     0.1005  
  0.2119     0.1910  
  0.3179     0.2723  
  0.4239     0.3455  
  0.5298     0.4113
```

XML tutorial

<http://www.zvon.org/>