

Adatbázisok 1.

XML – 1. rész

Document Type Definitions (DTD)

XML séma

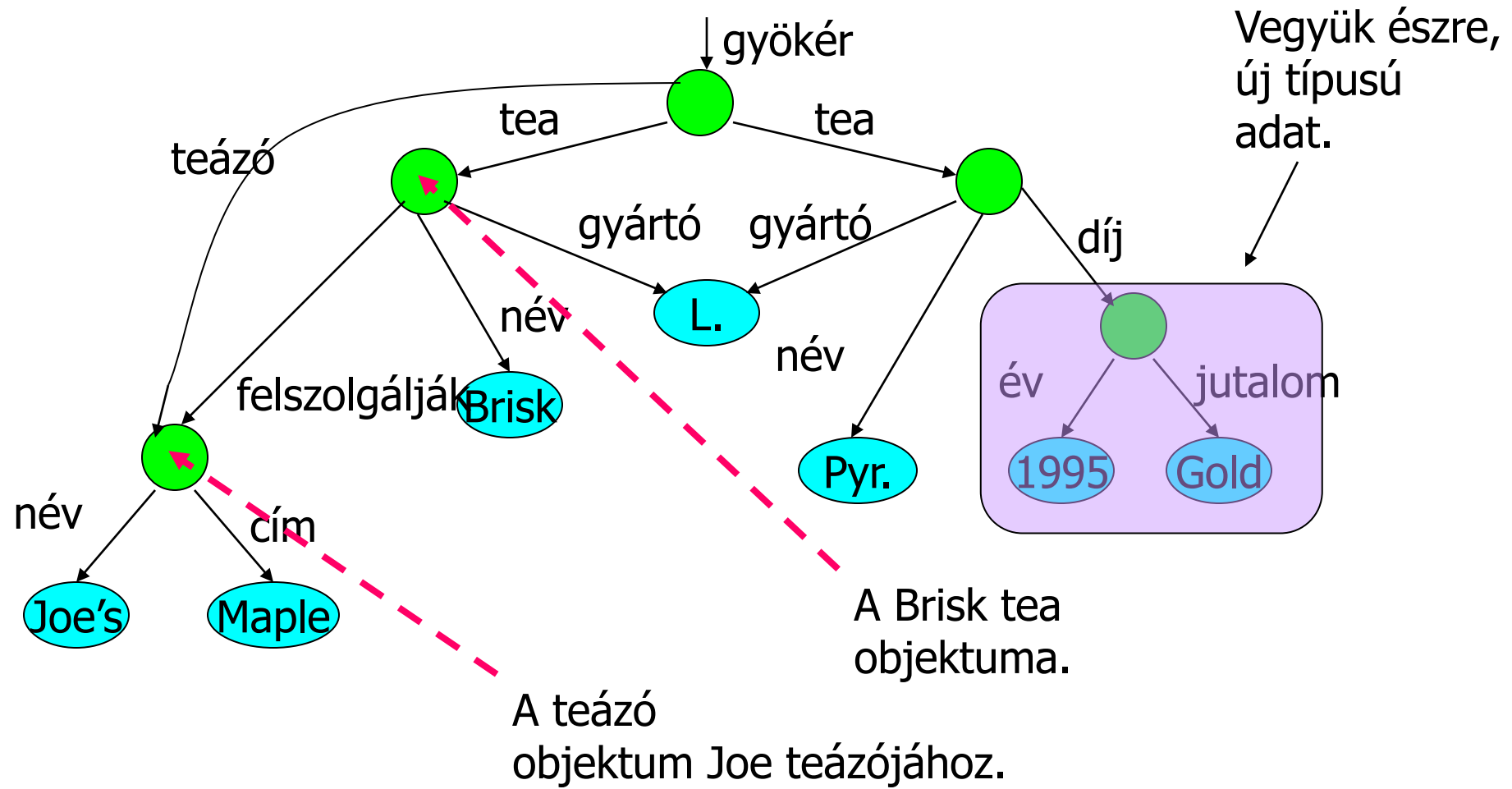
Féligstrukturált adat (*semistructured data*)

- Egy másik, fákon alapuló adatmodell.
- Motiváció: az adatok rugalmas megjelenítése (*flexible representation*).
- Motiváció: *dokumentumok* megosztása (*sharing of documents*) rendszerek és adatbázisok között.

A féligstrukturált adatok gráfja

- Pontok = objektumok.
- Az élek címkéi tulajdonképpen az attribútumnevek, kapcsolatok.
- Az atomi értékek a levélcsúcsokban (*leaf nodes*) tárolódnak.

Példa: adatgráf (*data graph*)



XML

- XML = *Extensible Markup Language*.
- A HTML a dokumentumok kinézetét írja le (pl. `<i>rózsa</i>` “italic”), az XML-ben a jelentés, szemantika leírására használják a **tageket** (jelölőket) (pl. “ez egy cím”).

XML dokumentumok

- A dokumentum *deklarációval* (*declaration*) kezdődik, amit egy speciális `<?xml ... ?>` tag határol.
- Tipikusan:

```
<?xml version = "1.0" encoding = "utf-8" ?>
```

```
<?xml version = "1.0" standalone = "yes" ?>
```

Jelölők/címkék (*tags*)

- A tagek, mint a HTML esetében is nyitó és záró elemből állnak `<VALAMI> ... </VALAMI>`.
 - Opcionálisan `<VALAMI/>`.
- Tageket tetszőlegesen egymásba ágyazhatjuk.
- Az XML tagek érzékenyek a kis- és nagybetű különbségre.

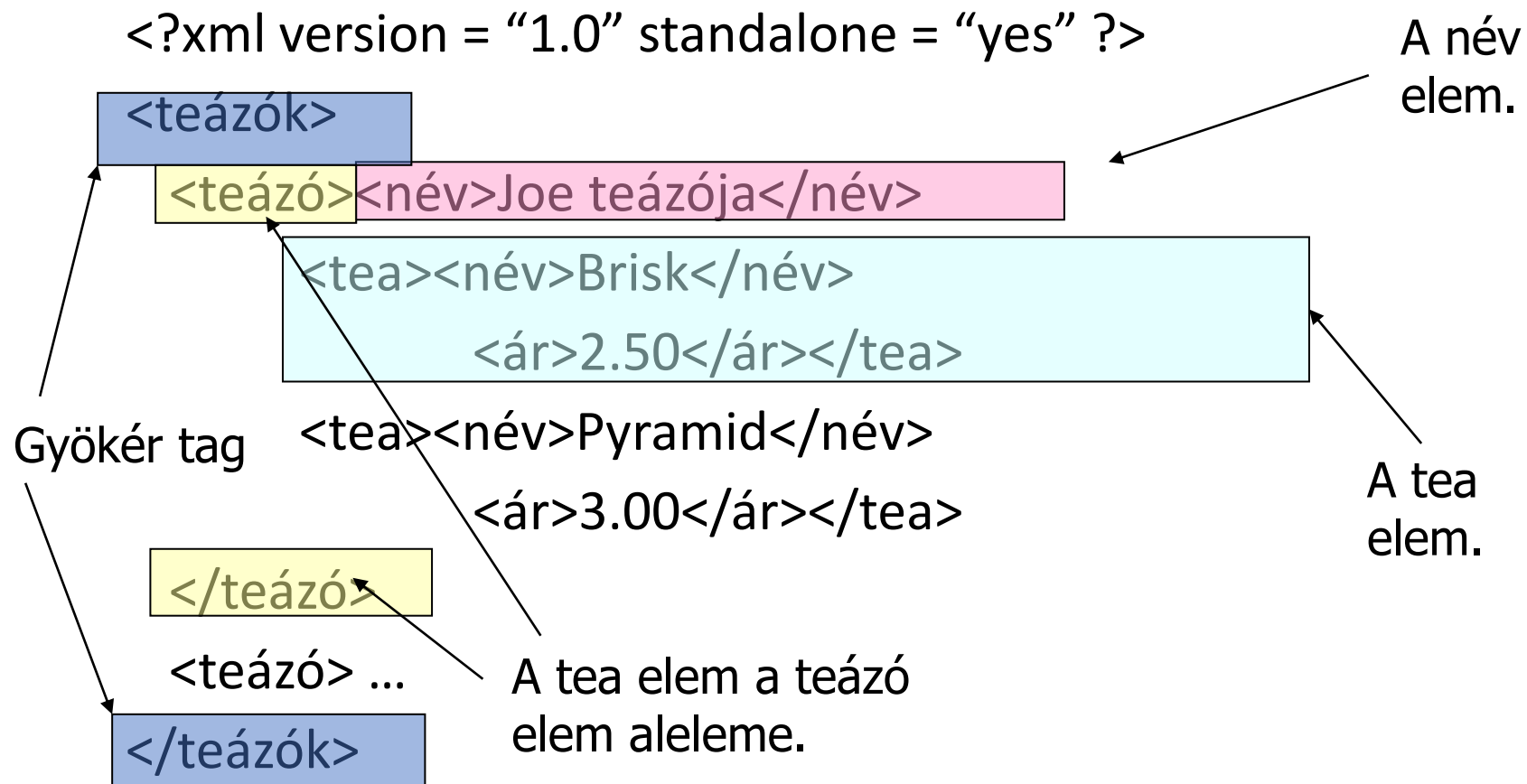
Jól formált és érvényes XML

- *Jól formált XML* (*well-formed XML*) megengedi, hogy önálló tageket vezessünk be.
- *Érvényes XML* (*valid XML*) illeszkedik egy előre megadott sémára: DTD vagy XML séma.

Jól formált XML

- Nem hiányzik a *deklaráció*: <?xml ... ?>.
- Minden nyitó tagnek megvan a záró párja.
- Egymásba ágyazhatók a tag-ek, de nem lehet „átfedés” köztük
- Egy *gyökér jelölő* (*root tag*)

Példa: jól formázott XML



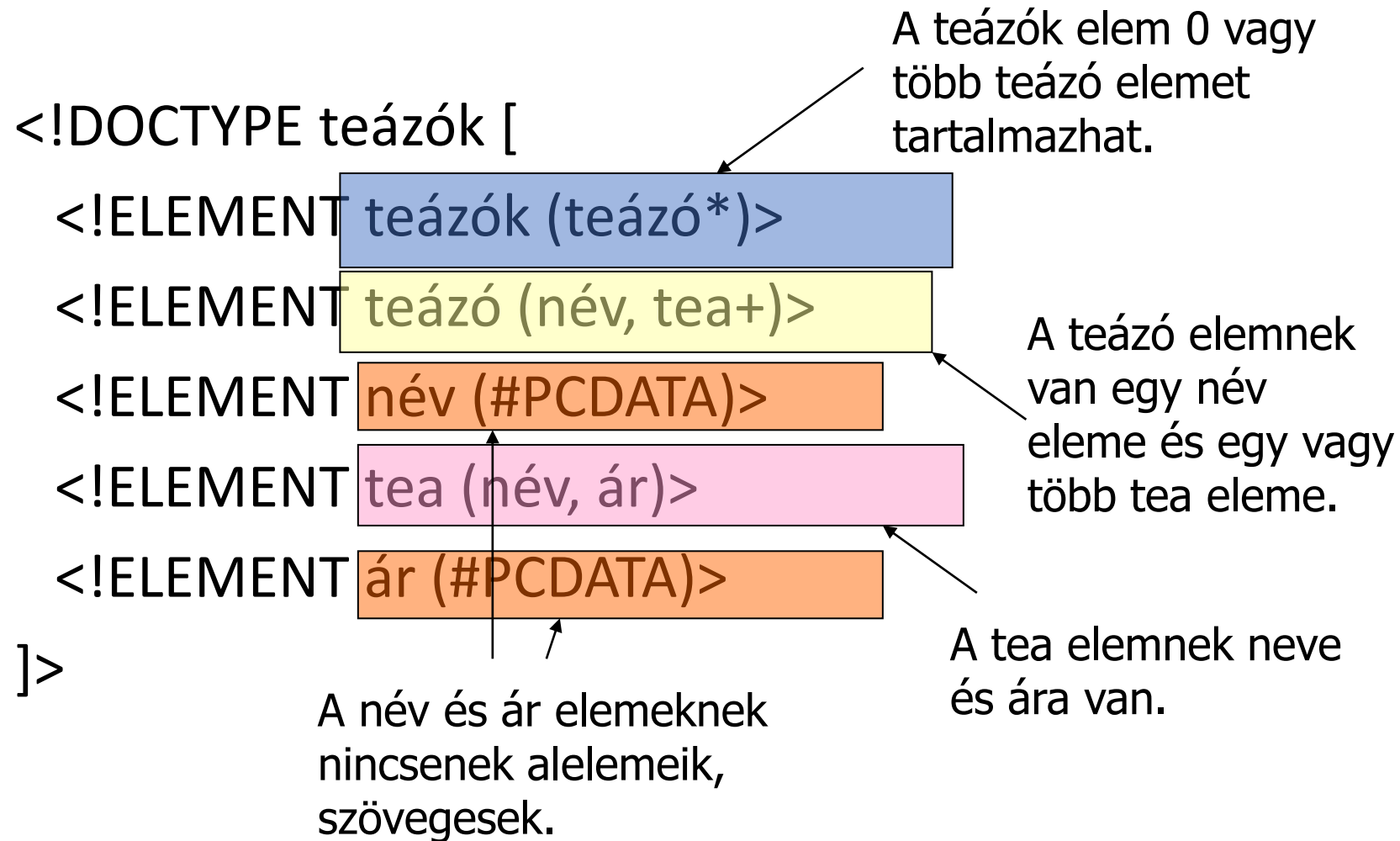
DTD felépítése

```
<!DOCTYPE gyökér elem [  
  <!ELEMENT elem név (összetevők) >  
  ... további elemek ...  
>
```

DTD ELEMENT

- Egy-egy elem leírása az elem nevét és zárójelek között az alelemek megadását jelenti.
 - Ez magában foglalja az alelemek sorrendjét és multiplicitását.
- A levelek (szöveges elemek) típusa #PCDATA (*Parsed Character DATA*).

Példa: DTD



Elemleírások (*element descriptions*)

- Az alelemek a felsorolás sorrendjében kell, hogy kövessék egymást.
- Egy elemet a felsorolásban egy további szimbólum követhet megadva a multiplicitását.
 - $*$ = 0 vagy több.
 - $+$ = 1 vagy több.
 - $?$ = 0 vagy 1.
- A $|$ szimbólum jelentése: vagy.

Példa: elemleírás

- A névhez opcionálisan hozzátartozik egy titulus, egy kereszt- és vezetéknév (ebben a sorrendben), vagy egy IP cím:

```
<!ELEMENT név (  
    (titulus?, kereszt, vezetékek) | IP cím  
)>
```

Példa: DTD-k használata (a)

```
<?xml version = "1.0" standalone = "no" ?>
```

```
<!DOCTYPE teázók [  
  <!ELEMENT teázók (teázó*)>  
  <!ELEMENT teázó (név, tea+)>  
  <!ELEMENT név (#PCDATA)>  
  <!ELEMENT tea (név, ár)>  
  <!ELEMENT ár (#PCDATA)>  

```

a DTD

a dokumentum

```
<teázók>  
  <teázó><név>Joe teázója</név>  
    <tea><név>Brisk</név> <ár>2.50</ár></tea>  
    <tea><név>Pyramid</név> <ár>3.00</ár></tea>  
  </teázó>  
  <teázó> ...  
</teázók>
```


Példa: DTD-k használata (b)

- Feltesszük, hogy a teázók DTD a teahouses.dtd fájlban van.

```
<?xml version = "1.0" standalone = "no" ?>
```

```
<!DOCTYPE teázók SYSTEM "teahouses.dtd">
```

```
<teázók>
```

```
  <teázó><név>Joe teázója</név>
```

```
    <tea><név>Brisk</név>
```

```
      <ár>2.50</ár></tea>
```

```
    <tea><név>Pyramid</név>
```

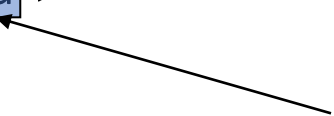
```
      <ár>3.00</ár></tea>
```

```
  </teázó>
```

```
  <teázó> ...
```

```
</teázók>
```

A DTD-t
teahouses.dtd fájlból
nyeri.



Attribútumok

- Az XML-ben a nyitó tag-ek mellett szerepelhetnek *attribútumok*.
- Egy DTD-ben,

`<!ATTLIST E A1 type1 value1 A2 type2 value2 ... >`

Az *E* taghez tartozó attribútumokat (*A1*, *A2* ...) adja meg.

Mindegyik attribútumhoz meg kell adni, hogy mi legyen a típusa és mi legyen igaz az értékére (lehet egy default vagy rögzített/konstans értéke, lehet kötelező vagy opcionális)

Példa: attribútumok

- A teázó elemhez hozzátartozik egy típus attribútum.

```
<!ELEMENT teázó (név tea*)>
```

```
<!ATTLIST teázó típus CDATA #IMPLIED>
```

Sztring.



Az attribútum opcionális,
ennek ellentéte: #REQUIRED

Példa: attribútum használatára

```
<teázó típus = "skót">  
  <név>McGuiver teázója</név>  
  <tea><név>Rooibos</név>  
    <ár>5.00</ár></tea>  
  ...  
</teázó>
```

ID és IDREF attribútumok

- Az attribútumok segítségével egyik elemről a másikra mutathatunk.
- Emiatt az XML dokumentum reprezentációja inkább gráf, mintsem egyszerű fa.

ID-k létrehozatala

- Adjuk meg egy E elemet és egy A attribútumát, aminek típusa: ID.
- Amikor az E elemet ($\langle E \rangle$) egy XML dokumentumba használjuk, az A attribútumnak egy máshol nem szereplő értéket kell adnunk.
- Példa:

$\langle E \quad A = \text{"xyz"} \rangle$

IDREF-ek létrehozatala

- Egy F elem az IDREF attribútum segítségével hivatkozhat egy másik elemre annak ID attribútumán keresztül.
- Vagy: az IDREFS típusú attribútummal több másik elemre is hivatkozhat.

A DTD

```
<!DOCTYPE teázók [
```

```
<!ELEMENT teázók (teázó*, tea*)>
```

```
<!ELEMENT teázó (felszolgál+)>
```

```
<!ATTLIST teázó név ID #REQUIRED>
```

```
<!ELEMENT felszolgál (#PCDATA)>
```

```
<!ATTLIST felszolgál melyikTea IDREF #REQUIRED>
```

```
<!ELEMENT tea EMPTY>
```

```
<!ATTLIST tea név ID #REQUIRED>
```

```
<!ATTLIST tea kapható IDREFS #IMPLIED>
```

lásd
később

A teázó elemek neve
ID attribútum, emellett
felszolgál alelemeik
lehetnek.

A felszolgál
elemhez
teákra
vonatkozó
hivatkozások
tartoznak.

A tea elemeknek is van egy ID típusú attribútuma
(név), a kapható attribútum pedig teázók nevét
tartalmazza.

Példa: a DTD-re illeszkedő dokumentum

<teázók>

 <teázó név = "JoeTeázója">

 <felszolgál melyikTea = "Brisk">2.50</felszolgál>

 <felszolgál melyikTea = "Pyramid">3.00</felszolgál>

 </teázó> ...

<tea név = "Brisk" kapható = "JoeTeázója SuzyTeázója ..." />

...

</teázók>

Példa: üres elem (*empty element*)

- A DTD:

<!ELEMENT felszolgál EMPTY>

<!ATTLIST felszolgál melyikTea IDREF #REQUIRED>

<!ATTLIST felszolgál ár CDATA #REQUIRED>

- Példa a használatra:

<felszolgál melyikTea = "Brisk" ár = "2.50" />

a záró tag itt lerövidül