Web-technológia II. XML feldolgozás: expat, DOM, SimpleXML

Hatwágner F. Miklós

Széchenyi István Egyetem, Győr

2019. április 11.

XML feldolgozók

Megközelítések:

- eseményvezérelt: expat könyvtáron alapul
 - ullet XML elemek o eseménykezelő fv. hívás
 - SAX (Simple API for XML) konvenciók
 - memóriatakarékos, de bonyolult kód
- fa alapú:
 - az egész dokumentum a memóriába kerül \to gyors, egyszerű kód, de memóriaigényes
 - DOM
 - SimpleXML

Elemző létrehozása:

```
xml_parser_create()
```

Bemeneti kódolás: automatikus felismerés, kimeneti: paraméterként megadható, aé.: UTF-8. Erőforrást ad vissza (nincs type hinting)

Eseménykezelők regisztrálása:

```
xml_set_element_handler()
```

Paraméterek:

parser elemző erőf.

start_element nyitó elem eseménykezelője end_element záró elem eseménykezelője

 Nyitó elem eseménykezelője: startElementHandler() parser elemző erőf. name elem neve, alapértelmezetten NAGYBETŰSRE alakítva attribs attribútumok asszociatív tömbben

```
    Záró elem eseménykezelője:
endElementHandler()
parser elemző erőf.
    name elem neve
```

• Karakteres adatok (CDATA) eseménykezelőjének regisztrálása:

```
xml_set_character_data_handler()
parser elemző erőf.
handler eseménykezelő fv.
```

• Eseménykezelő alakja:

```
handler()

parser elemző erőf.

data adatok string-ként. Lehet üres string, de az is lehet,

hogy egy node teljes szövege csak többszöri fv.

hívással érhető el.
```

Feldolgozási utasítások (Processing Instruction, PI) elhelyezése XML-ben:
 ? utasítás; ?>
 Pl. egy PHP kód elhelyezésére:

<?php echo 5*2; ?>

Eseménykezelő regisztrálása:

```
xml_set_processing_instruction_handler()
parser elemző erőf.
handler eseménykezelő fv.
```

Eseménykezelő alakja:

```
handler()

parser elemző erőf.

target PI célja (pl. PHP)

data a végrehajtandó utasítások. Kiértékelhető lenne pl.

eval() fv.-nyel, de veszélyes!
```

- Entitások: a legtöbb automatikusan feloldásra kerül,
- kivéve: $k \ddot{u} l s \ddot{o}$ és nem értelmezett \rightarrow eseménykezelő fv.
- Külső entitás eseménykezelőjének regisztrálása:

```
xml_set_external_entity_ref_handler()

parser_elemző erőf.

handler_eseménykezelő fv.
```

 Eseménykezelő alakja: handler()

```
parser elemző erőf.
openEntityNames nyitott entitások szóközzel tagolt listája, beleértve a hivatkozott entitást
base a külső entitás rendszer-azonosítójának (systemID) feloldásához használt bázis; mindig üres string systemID rendszer azonosító, ált. fájlnév publicID nyilvános azonosító, ha van ilyen, különben üres string
```

Ha vt. érték nincs vagy FALSE, az elemzés
 XML_ERROR_EXTERNAL_ENTITY_HANDLING hibával leáll.

Nem értelmezett entitásokat jelölés deklaráció kíséri, pl.:

Jelölés deklaráció eseménykezelőjének regisztrálása:

```
xml_set_notation_decl_handler()

parser elemző erőf.

handler eseménykezelő fv.
```

Eseménykezelő alakja:

```
handler()

parser elemző erőf.

notationName jelölésnév

base üres string
```

systemID rendszer azonosító publicID nyilvános azonosító



Nem értelmezett entitás kezelőjének regisztrálása:

```
xml set unparsed entity decl handler()
   parser elemző erőf.
 handler eseménykezelő fv.
```

Eseménykezelő alakja:

```
handler()
   parser elemző erőf.
   entity entitás neve
     base üres string
systemID rendszer azonosító
 publicID nyilvános azonosító
 notation az entitáshoz kapcsolt jelölésdeklarációra hivatkozik
```

 Alapértelmezett eseménykezelő: minden egyéb esetben (XML deklaráció, DTD) erre kerül a végrehajtás. Regisztrálása:

```
xml_set_default_handler()
parser elemző erőf.
handler eseménykezelő fv.
```

• Eseménykezelő alakja:

```
handler()

parser elemző erőf.

text szöveges adat
```

 Ha az eseménykezelő fv.-ek metódusok → objektum regisztrálása:

```
xml_set_object()
parser elemző erőf., ezt is illik expliciten felszabadítani
unset-tel
object az eseménykezelőt tartalmazó objektum
```

Az elemző beállításainak lekérdezése:

```
xml_parser_get_option()
parser elemző erőf.
option beállítás azonosítója
Vt. érték: beállítás értéke
```

Beállítások módosítása:

```
xml_parser_set_option()
parser elemző erőf.
option beállítás azonosítója
value beállítás értéke
```

Fontosabb beállítás azonosítók:

```
XML_OPTION_CASE_FOLDING elem- és attribútumnevek nagybetűsre alakítása
XML_OPTION_TARGET_ENCODING kimenet karakterkódolása, alapesetben egyezik a bemenetével
```

 Beállítások megtörténtek, eseménykezelők regisztrálva \rightarrow elemzés megkezdése:

```
xml_parse()
parser elemző erőf.
data adatcsomag; a fv. többször egymás után hívható egy
    fájl egymást követő darabjaival
[final] ez az utolsó adatcsomag?
vt. érték: elemzés sikeres?
```

Hibakezelés: hibakód lekérdezése:

```
xml_get_error_code()
parser elemző erőf.
```

Vt. érték: hibakód, ha minden OK: XML_ERROR_NONE

• Hibaüzenet lekérdezése a kód alapján:

```
xml_error_string()
code_hibakód
```



```
generateXML.php
```

```
<?php
 header("Content-Type: application/xml");
 // XML deklaráció
 $xml = "<?xml version=\"1.0\" encoding=\"UTF-8\" ?>\n";
 // Feldolgozási utasítás, Processing Instruction
 xml = "<?php echo 5*2; ?>\n";
 // receptek tömbjének összeállítása
 $receptek = array(
   array(
     "nev" => "Palacsinta",
     "hozzavalok" => array(
       "125 g liszt",
       "3,5 dl tej",
       "2 tojás"
     ),
     "elkeszites" => "Egy tálba lisztet szitálunk, majd beleütjük ".
       "a tojásokat..."
    ),
```

generateXML.php

```
array(
    "nev" => "Mustáros házinyúl",
    "hozzavalok" => arrav(
      "1,5 kg nyúlhús",
      "4 evőkanál olaj",
      "2 dl tejföl"
    "elkeszites" => "A nyulat nagyobb darabokra vágjuk, és sóval, ".
      "mustárral alaposan bedörzsöljük..."
);
// a tömb alapján xml előállítása
$xml .= "<szakacskonyv encoding=\"UTF-8\">\n";
foreach($receptek as $id => $recept) {
  \mbox{$xml.} = \mbox{$"\t<recept id=\"$id\">\n".}
          "\t\t<nev>{$recept['nev']}</nev>\n".
          "\t\t<hozzavalok>\n";
  foreach ($recept['hozzavalok'] as $hozzavalo) {
    $xml .= "\t\t\t<hozzavalo>$hozzavalo</hozzavalo>\n";
```

generateXML.php

```
const MERET = 4192:
private $elemzo; // XML elemző, parser
private $fh; // File handle
private $bejegyzes; // Aktuális recept adatai
private $verem; // Egymásba ágyazott XML elemek
private $id; // Keresett recept azonosítója
private $cdata; // Szöveges adatok összevonása
public function __construct(string $fajl) {
  $this->elemzo = xml_parser_create();
  xml_set_object($this->elemzo, $this);
  $this->fh = fopen($fail, "r");
public function __destruct() {
  fclose($this->fh);
  xml_parser_free($this->elemzo);
```

```
private function feldolgozas() {
  $this->bejegyzes = array();
  $this->verem = array();
  while ($adat = fread($this->fh, self::MERET)) {
    if (!xml_parse($this->elemzo, $adat, feof($this->fh))) {
      echo "Elemzési hiba: ".xml_error_string($this->elemzo).
           " Sor: ".xml_get_current_line_number($this->elemzo);
private function kezdoElem(string $elem, array &$attributumok) {
  array_push($this->verem, $elem);
  if(array_key_exists("ID", $attributumok)) {
    $this->bejegyzes["ID"] = $attributumok["ID"];
  if(array_key_exists("ENCODING", $attributumok)) {
    xml_parser_set_option($this->elemzo,
      XML_OPTION_TARGET_ENCODING, $attributumok["ENCODING"]);
```

```
private function zaroElem(string $elem) {
  array_pop($this->verem);
  if($elem == "RECEPT") {
    $this->bejegyzes = array();
public function showTartalom() {
  // Rossz ötlet, ki ne próbáld!
  // xml_set_processing_instruction_handler($this->elemzo, "pi");
  xml_set_element_handler($this->elemzo, "tartKezdoElem".
    "tartZaroElem"):
  xml_set_character_data_handler($this->elemzo, "tartCdata");
  echo "<h1>Tartalomjegyzék</h1>\n\t\t<u1>\n";
  $this->feldolgozas();
  echo "\t</u1>\n";
```

```
public function pi($elemzo, $cel, $adat) {
  if($cel == "php") {
    echo "<h2>Vajon mi ";
    eval($adat);
    echo "?</h2>\n";
public function tartKezdoElem($elemzo,
  string $elem, array &$attributumok) {
  $this->kezdoElem($elem, $attributumok);
public function tartZaroElem($elemzo, string $elem) {
  if($elem == "RECEPT") {
    echo "\t\t\t<a href=\"".$ SERVER["PHP SELF"].
         "?id=".$this->bejegyzes["ID"].
         "\">".$this->bejegyzes["NEV"]."</a>\n";
  $this->zaroElem($elem);
```

```
public function tartCdata($elemzo, string $szoveg) {
  if(end($this->verem) == "NEV") {
    if(array_key_exists("NEV", $this->bejegyzes)) {
      $this->bejegyzes["NEV"] .= $szoveg;
    } else {
      $this->bejegyzes["NEV"] = $szoveg;
public function showRecept(string $id) {
  this->id = id;
  xml_set_element_handler($this->elemzo, "receptKezdoElem",
    "receptZaroElem");
  xml_set_character_data_handler($this->elemzo, "receptCdata");
  $this->feldolgozas();
```

```
public function receptKezdoElem($elemzo, string $elem,
  array &$attributumok) {
  $this->kezdoElem($elem, $attributumok);
  $this->cdata = false:
  if($elem == "HOZZAVALOK" &&
     $this->bejegyzes["ID"] == $this->id) {
    $this->bejegyzes["HOZZAVALOK"] = array(); } }
public function receptZaroElem($elemzo, string $elem) {
  $this->cdata = false:
  if($elem == "RECEPT" && $this->bejegyzes["ID"] == $this->id) {
    echo "\t< h1>".$this->bejegyzes["NEV"]."</h1>\n".
         ''\t\t<\h2>\Hozzávalók:</h2>\n\t\t<\ul>\n";
    foreach ($this->bejegyzes["HOZZAVALOK"] as $hv) {
      echo "t\t$hv</li\n>";
    }
    echo "t\t</u1>\n\t\t<h2>Elkészítés:</h2>\n".
         "\t\t".$this->bejegyzes["ELKESZITES"]."\n".
         "\t\t<a href=\"".$_SERVER["PHP_SELF"]."\">".
         "Vissza a tartalomjegyzékhez</a>\n";
  $this->zaroElem($elem): }
```

```
public function receptCdata($elemzo, string $szoveg) {
  if(array_key_exists("ID", $this->bejegyzes) &&
     $this->bejegyzes["ID"] == $this->id) {
    $utolso = end($this->verem);
    if($utolso == "NEV" || $utolso == "ELKESZITES") {
      if(array_key_exists($utolso, $this->bejegyzes)) {
        $this->bejegyzes[$utolso] .= $szoveg;
      } else {
        $this->bejegyzes[$utolso] = $szoveg;
    } elseif($utolso == "HOZZAVALO") {
      if($this->cdata) {
        $hv = array_pop($this->bejegyzes["HOZZAVALOK"]);
        array_push($this->bejegyzes["HOZZAVALOK"],
                $hv.$szoveg);
      } else {
        $this->bejegyzes["HOZZAVALOK"][] = $szoveg;
  $this->cdata = true: } }
```

```
$\text{sparsexML.php}

$\text{horvathRozi} = new Szakacskonyv(
    "/home/feltoltes/www/szakacskonyv.xml");
    if(isset(\settinglefta])) {
        $\text{horvathRozi->showRecept(\settinglefta]);}
} else {
        $\text{horvathRozi->showTartalom();}
}
    ?>
    </body>
</html>
```

DOM

DOM referencia

Néhány fontosabb osztály:

- DOMDocument a teljes XML fát ez reprezentálja (gyökér elemet külön kell létrehozni)
- DOMNode egy általános csúcs a fában, többnyire a leszármazottjaival dolgozunk (pl. DOMDocument, DOMElement)
- DOMNodeList DOMElement-ek bejárható listája (Traversable interfész)
- DOMElement egy XML elemet reprezentál
- DOMAttr egy attribútumot reprezentál



DOM

DOMDocument néhány fontosabb metódusa

- Konstruktor opcionális paraméterei: verziószám, karakterkódolás
- Pl csúcs létrehozása: createProcessingInstruction()
- Elem létrehozása: createElement()
- Attribútum létrehozása: createAttribute()
- Elem keresése ID alapján: getElementByld(), de csak a DTD/XSD tudja megmondani, mi az ID attribútum neve! → DOMElement::setIdAttribute()
- Elemek keresése name attribútum alapján: getElementsByTagName()
- Betöltés: load()
- Mentés: save()
- XML szövegként: saveXML()



DOM

DOMElement néhány fontosabb metódusa

- Konstruktor paraméterei: név, opcionálisan érték
- Attribútum értékének lekérdezése: getAttribute(), beállítása: setAttribute()
- Tartalmazó DOMDocument elérése: ownerDocument adattaggal

DOMNode néhány fontosabb metódusa

- Gyerek hozzáadása: appendChild()
- Gyerek beszúrása: insertBefore(), paraméterek: mit, mi elé
- Gyerek eltávolítása: removeChild()
- Gyerekek elérése: \$e->childNodes->item(\$index)



generateDOM.php

```
<?php
header("Content-Type: application/xml");

// XML dokumentum létrehozása
$xml = new DOMDocument("1.0", "UTF-8");
$xml->xmlStandalone = true;
// Feldolgozási utasítás, Processing Instruction
$pi = $xml->createProcessingInstruction("php", "echo 5*2;");
if($pi) {
    $xml->appendChild($pi);
}
```

generateDOM.php

```
$receptek = array(
  array(
    "nev" => "Palacsinta",
    "hozzavalok" => array(
      "125 g liszt",
      "3,5 dl tej",
      "2 tojás"
    ),
    "elkeszites" => "Egy tálba lisztet szitálunk, majd beleütjük ".
      "a tojásokat..."
  ),
  array(
    "nev" => "Mustáros házinyúl",
    "hozzavalok" => array(
      "1,5 kg nyúlhús",
      "4 evőkanál olaj",
      "2 dl tejföl"
    ),
    "elkeszites" => "A nyulat nagyobb darabokra vágjuk, és sóval, ".
      "mustárral alaposan bedörzsöljük..."
  ));
```

${\sf generateDOM.php}$

```
// a tömb alapján xml előállítása
$szakacskonyv = $xml->createElement("szakacskonyv");
foreach($receptek as $id => $r) {
  $recept = $xml->createElement("recept");
  $recept->setAttribute("id", $id);
  $nev = $xml->createElement("nev", $r['nev']);
  $recept->appendChild($nev);
  $hozzavalok = $xml->createElement("hozzavalok");
  foreach ($r['hozzavalok'] as $h) {
    $hozzavalo = $xml->createElement("hozzavalo", $h);
    $hozzavalok->appendChild($hozzavalo);
  $recept->appendChild($hozzavalok);
  $elkeszites = $xml->createElement("elkeszites", $r['elkeszites']);
  $recept->appendChild($elkeszites);
  $szakacskonyv->appendChild($recept);
$xml->appendChild($szakacskonyv);
// eredmény megjelenítése, fájlba írása
echo $xml->saveXML();
$xml->save("/home/feltoltes/www/szakacskonyv.xml"); ?>
```

```
<!DOCTYPE html>
<html>
  <head>
    <meta charset="utf-8" />
    <title>Szakácskönyv</title>
  </head>
  <body>
    <?php
    class Szakacskonyv {
      private $konyv;
      public function __construct($fajl) {
        $this->konyv = new DOMDocument();
        $this->konyv->load($fajl);
```

parseDOM.php

parse DOM.php

```
public function showRecept($id) {
  $osszes = $this->konyv->getElementsByTagName("recept");
  // DTD hiánya miatt
  foreach($osszes as $recept) {
    $recept->setIdAttribute("id", true);
  $recept = $this->konyv->getElementById($id);
  echo "\t\t<h1>". $recept->getElementsByTagName("nev")->
    item(0) - nodeValue."</h1>\n";
  echo "t\t< h2>Hozzávalók:</h2>\n\t\n";
  $lista = $recept->getElementsByTagName("hozzavalo");
  foreach($lista as $hv) {
    echo "\t\t\t$hv->nodeValue</li\n>";
  echo "\t</u1>n":
  echo "\t\t<h2>Elkészítés:</h2>\n":
  echo "\t\t".$recept->getElementsByTagName("elkeszites")->
       item(0) - nodeValue."\n";
  echo "\t\t<a href=\"".$_SERVER["PHP_SELF"]."\">".
       "Vissza a tartalomjegyzékhez</a>\n";
} }
```

\$\text{parseDOM.php} \$\text{horvathRozi} = new Szakacskonyv("/home/feltoltes/www/szakacskonyv.xml"); if(isset(\$_GET["id"])) { \text{\$horvathRozi->showRecept(\$_GET["id"]);} } else { \text{\$horvathRozi->showTartalom();} } ?> </body> </html>

SimpleXML: elsősorban ismert szerkezetű fájlok feldolgozásához, de új XML is készíthető vele

- Konstruktor
- Létező fájl betöltése:

```
$xml = new SimpleXMLElement($location, null, true);
```

```
xml = simplexml_load_file(slocation);
```

- XML string elemzése:
 - \$xml = new SimpleXMLElement(\$xmlAsString);

```
$xml = simplexml_load_string($xmlAsString);
```

Új fájl készítés:

```
xml = new SimpleXMLElement('< root/>');
```

SimpleXMLElement legfontosabb metódusai:

- Feldolgozási utasítást nem kezeli, de SimpleXMLElement → DOMElement: dom import simplexml()
- Gyerek hozzáadása: addChild(), paraméterek: név, opcionális érték, vissza: gyerek csúcs
- Attribútum hozzáadása: addAttribute(), paraméterek: név, opcionális érték
- Lekérés stringként: asXML()

generateSimpleXML.php <?php header("Content-Type: application/xml"); // XML dokumentum létrehozása \$xml = new SimpleXMLElement('<?xml version="1.0" encoding="UTF-8" standalone="yes"?>'. '<szakacskonyv/>'); // Feldolgozási utasítás, Processing Instruction \$domDoc = dom_import_simplexml(\$xml) ->ownerDocument; \$pi = \$domDoc->createProcessingInstruction("php", "echo 5*2;"); \$domDoc->insertBefore(\$pi, \$domDoc->childNodes->item(0)); // receptek tömbjének összeállítása \$receptek = array(//...

generateSimpleXML.php foreach(\$receptek as \$id => \$r) { \$recept = \$xml->addChild("recept"); \$recept->addAttribute("id", \$id); \$recept->addChild("nev", \$r['nev']); \$hozzavalok = \$recept->addChild("hozzavalok"); foreach (\$r["hozzavalok"] as \$h) { \$hozzavalok->addChild("hozzavalo", \$h); } \$recept->addChild("elkeszites", \$r["elkeszites"]); }

// eredmény megjelenítése, fájlba írása

"/home/feltoltes/www/szakacskonyv.xml",

echo \$xml->asXML();
file_put_contents(

\$xml->asXML());

?>

Gyerek elemek elérése:

- adattagok az elemével egyező névvel (Traversable)
- Pl. echo \$xml->childElements[0]->childElement; // tartalom
- children()

Attribútumok lekérése:

- mintha tömbelemek lennének
- Pl. echo \$xml['attribute'];
- attributes()

Elem neve: getName()

parseSimpleXML.php

```
<html>
 <head>
   <meta charset="utf-8" />
   <title>Szakácskönyv</title>
 </head>
 <body>
   <?php
   class Szakacskonyv {
     private $konyv;
     public function __construct($fajl) {
       $this->konyv = new SimpleXMLElement($fajl, 0, true);
     public function showTartalom() {
       echo "<h1>Tartalomjegyzék</h1>\n\t\t\n";
       foreach($this->konyv->recept as $recept) {
         echo "\t\t\t<a href=\"".$_SERVER["PHP_SELF"].
              "?id=".$recept["id"].
              "\">".$recept->nev."</a>\n";
       echo "\t</u1>\n";
```

parseSimpleXML.php

```
public function showRecept($id) {
  foreach($this->konyv->recept as $recept) {
    if($recept["id"] == $id) {
      // XPath használatával közvetlenül is kiválasztható lenne
      // $recept = $this->konyv->xpath(
      // "/szakacskonyv/recept[@id=\"$id\"]");
      // $recept = $recept[0];
      echo "\t<h1>".$recept->nev."</h1>\n";
      echo "\t\t<h2>Hozzávalók:</h2>\n";
      echo "\t\n":
      foreach($recept->hozzavalok->hozzavalo as $hv) {
       echo "\t\t\t<li>$hv</li\n>";
      }
      echo "t\t\n";
      echo "\t\t<h2>Elkészítés:</h2>\n";
      echo "\t\t".$recept->elkeszites."\n";
      echo "\t\t<a href=\"".$_SERVER["PHP_SELF"]."\">".
           "Vissza a tartalomjegyzékhez</a>\n";
      break;
```

Feladat

Készítsen weboldalt egy rendezvény lehetséges időpontjának egyeztetéséhez! Az oldalnak tartalmaznia kell

- a rendezvény címét,
- rövid ismertetését,
- a lehetséges időpontokat, amelyek közül tetszőleges számú megjelölhető,
- a résztvevő nevének megadására szolgáló szövegbeviteli mezőt (kötelező kitölteni).

Az adatok megadását követően az adatokat gombnyomásra egy jól formázott (well-formed), de egyébként tetszőlegesen megválasztott szerkezetű XML fájlba kell menteni.

Megjelenítendő továbbá áttekinthető formában az is, hogy eddig ki milyen időpontokat jelölt meg számára megfelelőként.

Adjon lehetőséget a résztvevők által korábban megadott időpontok megváltoztatására, és a leadott szavazatainak teljes törlésére is (minden időpont törlését követően már a neve se látszódjék)!