

# ניהול נתונים באינטרנט – תרגיל מס' 1 – XML

## הוראות:

יש לעלות את הפתרונות בקובץ ZIP לMOODLE, שכולל קובץ PDF בשם answers.pdf ובו הפתרון וקבצי קוד נוספים (HTML, XML או Python) לפי הדרישה של כל סעיף וסעיף. ההגשה היא בזוגות, ורק אחד מבני הזוג יגיש את התרגיל, אך יש להקפיד לכתוב את השמות והת.ז. של שני בני הזוג בתוך הקובץ. שם של הקובץ ZIP צריך לכלול את הת.ז. של אחד מהמגישים (למשל: HW1\_123.zip). תאריך הגשה: **25.03.2018**

## שאלה 1:

לפניכם קובץ XML וחלק מה DTD המתאים לו. במסמך ישנם מספר שגיאות מסוגים שונים, כגון אי התאמה ל DTD, וטעויות בסינטקס. מצאו 3 שגיאות המופיעות בקובץ:

DTD:

```
<!ELEMENT PARTS (TITLE?, PART*)>
<!ELEMENT TITLE (#PCDATA)>
<!ELEMENT PART (ITEM, MANUFACTURER, MODEL, COST)+>
<!ELEMENT ITEM (#PCDATA)>
<!ELEMENT MANUFACTURER (#PCDATA)>
<!ELEMENT MODEL (#PCDATA)>
```

XML:

```
<PARTS>
  <TITLE>Computer Parts</TITLE>
  <PART>
    <ITEM>Motherboard</ITEM>
    <MANUFACTURER>ASUS</MANUFACTURER>
    <MODEL>P3B-F</MODEL>
    <COST> 123.00 </COST>
  </PART>
  <PART>
    <ITEM>Video Card</ITEM>
    <MANUFACTURER>ATI</manufacturer>
    <COST CURRENCY='USD'> 160.00</COST>
  </PART>
  <PART>
    <ITEM>Sound Card</ITEM>
    <MANUFACTURER>Creative Labs</MANUFACTURER>
    <MODEL>Sound Blaster Live</MODEL>
    <COST CURRENCY='USD'> 80.00
  </PART>
  <PART>
    <ITEM>Mouse</ITEM>
    <MANUFACTURER>Logitech</MANUFACTURER>
    <MODEL>AF180</MODEL>
    <COST> <CURRENCY ID='USD'>8.00</CURRENCY> <CURRENCY ID='EUR'>10.00</CURRENCY> </COST>
  </PART>
  <PART>
    <ITEM> 19 inch Monitor</ITEM>
    <MANUFACTURER>LG Electronics</MANUFACTURER>
    <MODEL> 995E</MODEL>
    <COST CURRENCY='USD'> 290.00</COST>
  </PART>
  <PART />
    <ITEM> Processor </ITEM>
    <MANUFACTURER>Intel</MANUFACTURER>
    <COST CURRENCY='EUR'> 240.00</COST>
  </PARTS>
```

## שאלה 2:

בשאלה זו נתייחס לקובץ: <http://slavanov.com/teaching/wdm1415b/books.xml>

(א) כתבו DTD מתאים לקובץ (ישנה יותר מאפשרות נכונה אחת) והסבירו את ה DTD שכתבתם

(ב) כתבו שאילתת XPATH שמחזירה את הנתונים הבאים:

a. מסמך הכולל את כל החנויות שיש בהם ספרים שעולים מתחת ל 10 דולר

b. מסמך הכולל את כל הספרים שעולים מעל 3 דולר בלפחות אחת החנויות

c. מסמך הכולל את כל הספרים שיש להם יותר ממחבר אחד

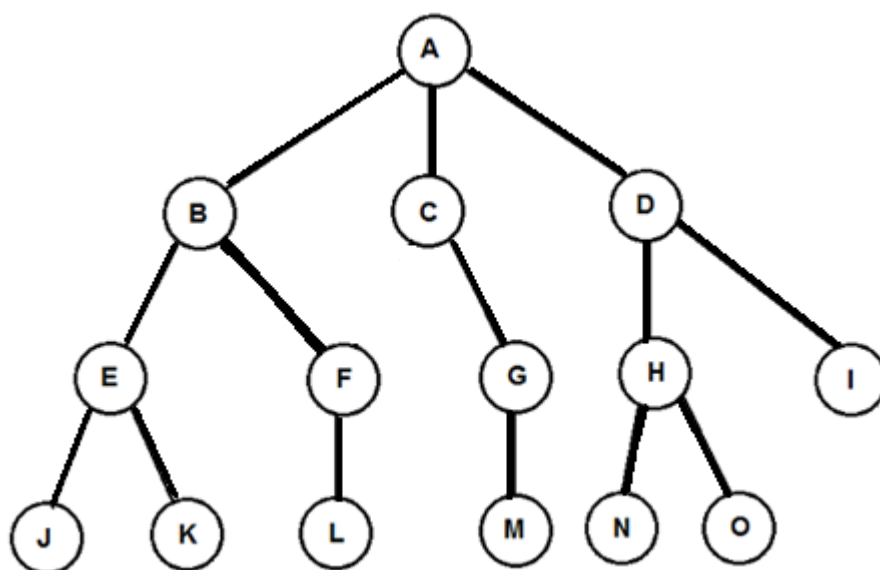
d. מסמך הכולל את כל המחברים שכתבו ספר אחרי שנת 2000

(ג) הריצו את כל אחת מהשאילתות שקיבלתם על הקבצים הנתונים ותגישו את התוצאות.

ניתן להשתמש ב: <http://www.xpathtester.com/xpath>

## שאלה 3:

נתון עץ הבא:



(א) כתבו את ה DOCUMENT ORDER של העץ הנ"ל.

(ב) כתבו את העץ הנ"ל ביצוג של XML

(ג) כתבו את התוצאה של הרצת הצירים הבאים על העץ הנ"ל:

a. Preceding-sibling של קודקוד G

b. Descendant של קודקוד C

c. Following של קודקוד F

d. Ancestor-or-self של N

e. Preceding של M

f. Following של Parent של K

## שאלה 4:

בשאלה זו נעסוק בעצי XML שבהם התוויות על העלים הן: GREEN, RED, BLUE. התוויות על כל הצמתים הפנימיים הן UNDEF. על עץ כזה נאמר שהוא ניתן לצביעה אם ניתן להקצות לכל צומת פנימי את אחד מ-3 הצבעים, כך שיתקיים שצבעו של כל צומת שונה מצבעם של כל ילדיו (המידיים).

תהי L שפת כל העצים (מדרגה לא חסומה) הניתנים לצביעה.

1. כתבו בפירוט אוטומט מטה-מעלה (bottom-up), לא בהכרח דטרמיניסטי, המקבל את השפה L בדיוק. שימו לב שעל האוטומט להיות unranked.

2. נגביל כעת את השפה לעצים טרינאריים בלבד (בדיוק 3 ילדים לכל צומת פנימי) נקרא לשפה החדשה L2. האם ניתן לזהות את L2 ע"י אוטומט מעלה-מטה (top-down) לא דטרמיניסטי? הוכיחו (ע"י בנייה מפורשת של האוטומט) או הפריכו.

3. האם ניתן לזהות את L2 ע"י אוטומט מעלה-מטה דטרמיניסטי? הוכיחו (ע"י בניית האוטומט) או הפריכו.

## שאלה 5:

(א) הכינו קובץ HTML, הכולל תיאור של עיר. דף זה צריך לכלול תמונה של העיר, דגל של המדינה עליה העיר שייכת, וטבלה שכוללת נתונים על העיר (למשל כמות תושבים לפי שנים). כמו כן, הדף צריך לכלול לינק חיצוני לאתר העיר. הדף צריך להיות בנוי בשפה אנגלית ותיקני מבחינת הסטנדרט של HTML. (ניתן לוודא ב <http://validator.w3.org/question5a.html>) קראו לקובץ

(ב) כתבו תוכנית Python שמקבלת כתובת של עמוד אינטרנט ומדפיסה לקובץ:

a. את attributen בשם src של כל התמונות שמופיעות בעמוד שיש להם attribute בשם alt והוא לא ריק.

b. לינקים חיצוניים שמובילים לאתרים בדומיין CO.IL

c. את התוכן של השורה השני ב בטבלה הראשונה שמופיע בעמוד

d. כל המילים המסומנים בITALIC

הגישו את הפתרון בקובץ בשם: question5b.py

(ג) הריצו את התוכנית מסעיף (ב) על האתר הבא:

<http://www.euronews.com/2015/03/20/uber-taxis-overtake-new-york-yellow-cabs>

והגישו את הפלט של התוכנית בקובץ question5c.txt