



מדינת ישראל

משרד האוצר - אגף החשב הכללי
ועדת האינטרנט הממשלתית
WWW.ITPOLICY.GOV.IL

שם המסמך : דגשים בנוגע לשילוב עברית במסמכי HTML
תאריך כתיבה : 7/11/97 תאריך עדכון : 12.11.97

גירסא : 1.0 (סופית)

מצב המסמך : לפרסום

שם לאחזור : hebrew/siteheb

כותב המסמך : דודו רשתי, איציק ירחי

הערות : המבוא למסמך זה והערות שוליים (1,2) נלקחו מדו"ח הכנסת בנושא "הערכות מדינת ישראל לעידן התקשוב והמידע" שמחבר המסמך היה שותף בכתיבתו.

מבוא

כיום קיימות, דה-פקטו, שתי שיטות ליישום עברית ב-HTML: שיטת סנונית, המבוססת על עברית חזותית¹, ושיטה המבוססת על העברית המשתמעת² של מערכת חלונות. יודגש, שאף אחת משתי השיטות אינה תקנית, משום בחינה. שיטת סנונית מאפשרת שימוש בעברית באינטרנט במערכות הפעלה שונות מבלי להזדקק למערכת הפעלה עברית, ולכן ניתן לעיין בדפי אינטרנט כאלה גם בתפוצות, תחת מגבלות השיטה החזותית. שיטת חלונות מתגברת על מגבלות השיטה החזותית, אולם היא ייחודית למערכת הפעלה מסוימת, אשר כמעט שלא ניתן להשיגה מחוץ לישראל.

כיום שלטת שיטת העברית החזותית, ועל-פי החלטת ועדת האינטרנט הממשלתית (תכ"ס 5.3.4) יש להשתמש בה להצגת אתרי מידע ממשלתיים ברשת האינטרנט.

מסמך זה יציג את הדגשים ואת אופן הצגת העברית במסמך HTML לפי שיטת העברית החזותית.

זיהוי העברית במסמך

על-מנת לאפשר לדפדפן לזהות את הגופן העברי ואת שיטת היישום של העברית שבה נעשה שימוש, יש להשתמש בתגית META בראש המסמך.

בדפדפן Netscape³ ובדפדפן של חברת Microsoft⁴ נעשה שימוש בתגית זו על-מנת לטעון את הגופן המתאים, וכן כדי לאפשר את הצגת העברית והדפסתה בכיווניות המתאימה.

השימוש בתגית מתבצע באופן הבא :

```
<META http-equiv="Content-Type" content="text/html; charset=iso-8859-8">
```

יש לכלול את התגית בחלק הראשון של מסמך ה-HTML, הנמצא בתוך תגיות ה-HEAD, וזאת באופן הבא :

```
<HTML>
```

```
<HEAD>
```

```
<META http-equiv="Content-Type" content="text/html; charset=iso-8859-8">
```

```
</HEAD>
```

```
<BODY>
```

בחלק זה מופיע תוכן המסמך

```
</BODY>
```

</HTML>

יש לשים לב לכך, שגם שרת המידע (Web Server) מצרף מידע META משלו לדף ה-HTML. לכן, יש לוודא שלא קיימת סתירה בין מידע ה-META של השרת, לבין תגיות ה-META בדף ה-HTML עצמו. סתירות אלו גורמות לטעינה כפולה של דף המידע בדפדפן.

ניתן להגדיר את שרת המידע כך שיצרף לכל דף מידע הנשלח ממנו את תגית ה-META המתאימה לתמיכה בעברית. לקבלת מידע נוסף, יש להיעזר בספרות הספציפית לכל שרת מידע.

יישור העברית לימין המסמך

קיימות מספר דרכים ליישור העברית לימין המסמך. חשוב להדגיש, שבכל השיטות האמורות אין אפשרות להתמודד עם שבירה אוטומטית של המלל בהתאם לרוחב המסך, ולכן יש לחשב מראש את רוחב המסך האופטימלי ועל-פיו לשבור את השורות באופן ידני. עובדה זו מהווה את אחד החסרונות העיקריים בשיטת העברית החזותית והיא מקשה גם על בוני דפי המידע.

נמנה את השיטות המאפשרות הצמדת העברית לימין המסך :

1) הצמדת העברית באמצעות התגית `<DIV>`.

באמצעות התגית `<DIV>` ניתן להגדיר פסקה בגוף מסמך ה-HTML, שבה המלל מיושר לימין. הדבר מתבצע באופן הבא :

```
<DIV align=right><NOBR>
```

```
...<BR>שורה ראשונה...
```

```
...<BR>שורה שניה...
```

```
...</NOBR></DIV>
```

שימו לב שבסוף כל שורה בעברית יש להוסיף את התגית `
`, שפירושה שבירת שורה.

התגית `<NOBR>` מונעת את שבירת השורות הארוכות בעברית בעת הקטנת רוחב מסך הדפדפן.

היתרון של השימוש בשיטה זו לעומת שיטות אחרות מתבטא בעובדה שאם מופיעה שורה ארוכה במיוחד בתוך דף המידע, אין הדבר משפיע על היישור לימין של שאר השורות בדף, אלא רק אותה השורה חורגת מהיישור לימין האופטימלי. החיסרון הוא באי-היכולת לקבוע את המיקום היחסי של המלל ביחס לרוחב המסך כמו שניתן באמצעות טבלאות.

2) הצמדת העברית באמצעות התגית `<PRE>`.

באמצעות התגית `<PRE>` ניתן להצמיד מספר רב של שורות בעברית לימין המסך ללא שימוש בתגית `
` בסוף כל שורה. התגית `<PRE>` נועדה לחלקים במסמך ה-HTML שבהם מעוניין המשתמש לשלב קטעי מלל ערוכים מראש. המלל מופיע על המסך באמצעות גופן שווה-רווחים (Fixed Width), לעומת גופן בעל רווחים-משתנים (Variable Width) שבו מופיע המלל בדרך כלל. החיסרון הנוסף הוא הצורך "לדחוף" רווחים בתחילת כל שורה בעברית, על-מנת ליישר את השורות לימין. שיטה זו אינה מומלצת לשימוש, הן בשל העבודה הרבה הנדרשת לשם הכנת המסמך והן בשל בעיית הקריאות של המידע המופיע על מסך הדפדפן באמצעות גופן שווה-רווחים.

3) הצמדת העברית באמצעות התגית `<TABLE>`.

שיטה זו היא אמנם המורכבת ביותר לשימוש, אך היא מספקת יכולת תימרון רבה. היא מאפשרת שליטה גם על מיקום המלל ביחס לרוחב המסך וגם על מניעת שבירת שורות בתוך

המסמך.

הדבר מתבצע באופן הבא :

```
<TABLE width=x%><TR><TD align=right NOWRAP>
```

```
...שורה ראשונה<BR>
```

```
...שורה שניה<BR>
```

```
...
```

```
</TD></TR>
```

```
</TABLE>
```

שימו לב לפרטים הבאים :

x% או x - מאפשר להגדיר את רוחב הטבלה יחסית לרוחב המסך. ניתן בדרך זו למקם את המלל באופן יחסי (באחוזים) או מוחלט (בפיקסלים) במקום מסוים לרוחב מסך הדפדפן. כאשר משתמשים בערכים מוחלטים יש לקחת בחשבון שרוחב מסך ממוצע נע סביב ה-600 פיקסלים.

NOWRAP - מונע שבירת שורות. גם כאשר רוחב מסך הדפדפן קטן מרוחב השורה - לא יישברו שורות המלל, דבר אשר משבש⁶ את המלל העברי המופיע על המסך.

4) הצמדת פסקאות בעברית באמצעות התגית ⁷<P>.

באמצעות התגית <P> ניתן למקם פסקה עברית בתוך דף מידע. השימוש בתגית זו מומלץ רק לקטעים קצרים, שנעשה בהם שימוש בעברית, או לשורה בודדת. השימוש נעשה באופן הבא :

```
<P align=right>
```

```
...שורה
```

```
</P>
```

ממירים רבים של מעבדי תמלילים משתמשים מסיבה עלומה בתגית <P> לשם הצמדת המלל העברי לימין. דבר זה אינו רצוי והופך את הטיפול במסמך למורכב ובעייתי.

כתיבת נושא המסמך

המלל העברי המופיע בתוך תגית ה-TITLE מופיע בנפרד מהמסמך עצמו, לרוב על מסגרת החלון. קביעות המלל וסוג הגופן שמוגדרות בדפדפן אינן חלות על המלל של תגית ה-TITLE. משום כך, אופן הצגתו של מלל זה נקבע אך ורק על-ידי מערכת ההפעלה של המשתמש. מלל בעברית חזותית המופיע בתוך תגית זו יופיע הפוך, כאשר דף מידע זה יוצג בחלונות תומכי-עברית של Microsoft. לכן, מומלץ לכתוב את המלל המופיע בתגית ה-TITLE בעברית משתמעת ולא בעברית חזותית.

שימוש בטפסים

טפסים מאפשרים משלוח של מלל מהמשתמש לשרת ועיבודו. קיימים שדות רבים המאפשרים קליטת מלל. על-מנת לאפשר הצגת גופן עברי בשדות אלו יש להשתמש בתגית font לפני השימוש בתגית, וזאת באופן הבא :

עבור שדות בהם מופיע התגית INPUT

```
<FONT face="Arial (hebrew)">
<INPUT TYPE="text" NAME="שם השדה" SIZE="גודל" VALUE="ערך התחלתי">
</FONT>
```

עבור שדות בהם מופיע התגית SELECT

```
<FONT face="Courier New (hebrew)">
<SELECT NAME="שם השדה" >ערך
</FONT>
```

שימו לב שבחלק מהדפדפנים (למשל, Netscape בגרסה 3) קיימת בעיה לקלוט מלל בעברית בשדה ה-TextArea.

ניהול הקלט/פלט בטפסי HTML וסוגי הגופנים שניתן להשתמש בהם תלויים באופן מוחלט במערכת ההפעלה הספציפית של המשתמש. לכן, לא ניתן להשתמש באופן גורף בעברית חזותית או משתמעת, אלא צריך לבנות את הטופס לפחות עבור שני הסוגים של מערכות ההפעלה.

למרות זאת, מתברר כי במערכות שאינן תומכות בעברית משתמעת, לא תופיע תמיכה ממשית⁸ בקלט/פלט בעברית. על כן, במצב העניינים כיום, ניתן להסתפק בבניית טפסים למערכות הפעלה המשתמשות בעברית משתמעת בלבד.

שפת JavaScript

שפת JavaScript מאפשרת לשפר את האינטראקציה בין המשתמש לדף המידע. סוג העברית שבה משתמשים תלוי במערכת ההפעלה שבה פועל היישום. למשל, אם פותחים חלון דפדפן חדש וכותבים לתוכו, יש להשתמש בעברית חזותית. אולם, פתיחה של חלון מסוג alert תחייב בדיקה לפני הצגתו. הבחירה בין שימוש בעברית חזותית או משתמעת תלויה, שוב, במערכת ההפעלה.

במסמך זה תקצר היריעה כדי לפרט את כל האזורים של כתיבת מלל עברי ב-JavaScript. כל אזור עשוי להיות מושפע ממערכת ההפעלה הספציפית של המשתמש.

תמיכה בעברית במערכות הפעלה שונות

מספר בעיות נוספות מתעוררות בכתיבת מסמכי HTML בעברית, והן נובעות מן השוני בטיפול בעברית במערכות ההפעלה השונות. נתייחס כאן לשלוש מערכות ההפעלה הנפוצות: חלונות בגרסאות 3.11 ו-95 התומכות בעברית, חלונות בגרסאות שאינן תומכות בעברית, מקינטוש ויוניקס. ההבדל בין מערכות ההפעלה הוא פונקציה של שלושה מרכיבים: (1) סוג התמיכה בעברית (חזותית או משתמעת); (2) האלגוריתם המשמש ליישום עברית משתמעת; (3) חוסר מוחלט בתמיכה בעברית ברמת מערכת ההפעלה, אשר אינו מאפשר כתיבת מלל עברי מימין לשמאל.

האינטרנט חידד את הבעייתיות שבשימוש בעברית. עד כה לא נדרשו יישומים תומכי-עברית לפעול בו-זמנית על מערכות הפעלה שונות, שכן בעידן האינטרנט דפי מידע שנוצרו על מחשבי מקינטוש יכולים לנדוד לסביבת חלונות ולהיפך. השוני ביישום העברית בין מערכות ההפעלה השונות צץ ועלה על פני השטח כאשר משתמשים גילו להפתעתם שכותרת כלשהי, אשר במערכת הפעלה אחת הופיעה בצורה תקינה, לא הופיעה כך במערכת הפעלה שנייה. חמור מכך הוא המצב במה שנוגע לאפשרות לעבוד עם טפסים במערכות הפעלה שונות. אין בכוונת מסמך זה

להיכנס לעומקם של ההבדלים, אך חשוב להביא לתשומת-ליבו של המשתמש את דבר קיומם ולעורר את עירנותו לנושא. הפתרון, לעניות דעתנו, הוא לבחון את אופן הצגת דפי המידע במספר מערכות הפעלה שהן פוטנציאליות למשתמש הקצה, לפני שחרור היישום לרשת.

נושאים נוספים

מספר בעיות נוספות מתעוררות בכתיבת מסמכי HTML בעברית חזותית:

1) **העתקת מלל מחלון לחלון** - העתקה (Cut & Paste) של מלל מחלון הדפדפן לחלון של מעבד התמלילים במערכת החלונות היא בעייתית. משתמשים מגלים זאת בהפתעה רבה ולא מבינים מדוע המלל יוצא הפוך. הדבר נובע משימוש בעברית החזותית, שהיא הפוכה בסדר כתיבתה מן העברית המשתמעת, שבה משתמשים רוב היישומים במערכת חלונות. פתרון הבעיה אינו פשוט ויש להיעזר בתוכנות ביניים, המאפשרות להפוך את העברית. קיימים בשוק כמה פתרונות - חלקם מסחריים וחלקם בתחום ה-Public Domain. יישום של חברת RTLSoft מחולק ללא תשלום וניתן להורידו מאתר סנונית בכתובת <http://www.snunit.k12.il/heb.html>.

2) **הדפסת המלל** - הדפסת המלל מחלון המכיל עברית חזותית היא בעייתית בגלל הגופן שבו משתמשים בעת ההדפסה. הגופן המשמש לעברית חזותית אינו גופן רגיל במערכת החלונות העברית, ולפיכך הוא יוצר קשיים בעת ההדפסה. קיימים מספר פתרונות לבעיה. בדפדפן של חברת Microsoft נפתרה הבעיה לחלוטין, וזאת בכך שהדפדפן תומך בהדפסת קבצים בעברית חזותית. למידע נוסף על פתרון בעיית ההדפסה בדפדפנים אחרים או בגרסאות של דפדפן חברת Microsoft, אשר אינו תומך בעברית, ראו מידע באתר סנונית בכתובת <http://www.snunit.k12.il/heb.html>.

3) **כתיבת מסמכי HTML בעברית חזותית** - דבר זה הוא בעייתי בגלל מחסור ביישומים הפועלים על מערכת חלונות (Windows3.x, Windows95), המאפשרים כתיבת עברית חזותית. בעבר, הפתרון היה כתיבת העברית בצורה הפוכה... בשלב הבא נוצרו יישומים המאפשרים לכתוב עברית ישרה ולהפוך אותה על-מנת להעתיקה לחלון עורך הקבצים. כיום, קיימים מספר עורכי קבצים המאפשרים ייצור של עברית חזותית במערכות חלונות, כמו יישום בשם Emacs המופץ ללא תשלום. כמו-כן, ניתן להמיר מסמך ממעבד התמלילים Word למסמך HTML המופיע בעברית חזותית. למידע לגבי יישומים אלו, ראו אתר סנונית בכתובת <http://www.snunit.k12.il/heb.html>.

¹ בשיטה **החזותית** (Visual) מיוצג הטקסט במחשב באותו הסדר שבו הוא מוצג על נייר או על צג המחשב. לדוגמא, הרצף "שמו Joe" ייוצג על-ידי: J, אח"כ o, אח"כ e, אח"כ רווח, אח"כ ו, אח"כ מ, ולבסוף ש. אם תרצו, זוהי שיטתה של מכוונת הכתיבה. גישה זו קלה ליישום, אך גורמת לקשיים רבים בעיבוד הטקסט (עריכת שורות, גלישה משורה לשורה, שינוי גודלי תווים, הקלדת טקסט, מיון טקסט ועוד). ייצוג טקסט ויזואלי גורם לתוצאות "מוזרות" ובלתי קריאות אם ננסה לערוך מחדש את השורות (באורך שונה), למשל. גישה זו אומצה כפתרון אד-הוק לייצוג עברית ברשת האינטרנט, בדואר אלקטרוני ובאתרי אינטרנט (WWW) לא מעטים.

² בשיטה **המשתמעת** (Implicit) מיוצג הטקסט במחשב באותו הסדר שבו הוא נכתב או נקרא, כלומר הסדר בו הוא מקבל משמעות. הרצף "שמו Joe" דלעיל, בהנחה שכיוון הטקסט הראשי הוא עברי (מימין לשמאל), ייוצג על-ידי: ש, אח"כ מ, אח"כ ו, אח"כ רווח, אח"כ J, אח"כ o, ולבסוף e. תוכנת המחשב אחראית להצגה והדפסה נכונות של הטקסט. שיטה זו מתאימה לטיפול מתקדם בטקסט, אך יישומה בתוכנה אינו טריביאלי. בנוסף, יש להגיע להסכמה לגבי השיטה (אלגוריתם) שבו תוגדר הכיווניות.

³ בדפדפן חברת Netscape בגרסאות 1 עד 4 לא היתה התייחסות כלשהי לשפה העברית. התמיכה בעברית היא פועל יוצא של השימוש בעברית חזותית. למרות זאת, קיים הבדל בתמיכה של המוצר בהיבטים שונים של עבודה בעברית ברשת, כמו שימוש בטפסים בין הגרסאות השונות של המוצר. חברת Netscape הודיעה שבכוונתה להוסיף תמיכה בשפה העברית בגרסאות מתקדמות יותר של המוצר. ניתן להוריד עותק של הדפדפן מאתר החברה

⁴ נכון לזמן כתיבת מסמך זה, הדפדפן של חברת Microsoft, המכונה Internet Explorer, תומך בעברית בגרסא 3.02. ניתן להוריד גרסא זו של הדפדפן מאתר החברה בכתובת <http://www.microsoft.com/israel>. בגרסאות מתקדמות יותר של הדפדפן אין תמיכה אוטומטית בעברית ותמיכה זו משווקת לרוב רק מספר חודשים לאחר יציאת הגרסא באנגלית.

⁵ למידע נוסף ראה :

<http://developer.netscape.com/library/documentation/htmlguid/tags1.htm#1091526>

⁶ שיבוש המלל העברי בעת שבירת שורות מתבטא בכך ששורה ארוכה, לאחר שבירתה, הופכת לשתי שורות עם רצף קריאה משובש. לדוגמא, השורה הארוכה הבאה: "הסיפור שסיפר המשורר היה מלא חוויות אישיות מימי הילדות", לאחר שבירתה, תהפוך לשתי שורות שרצף הקריאה שלהן יוצר את המשפט הבא: "היה מלא חוויות אישיות מימי הילדות הסיפור שסיפר המשורר". השבירה היא מקרית ויכולה ליפול על כל מילה בשורה.

⁷ למידע נוסף ראה:

<http://developer.netscape.com/library/documentation/htmlguid/tags3.htm#1006012>

⁸ במערכות שאינן כוללות תמיכה בעברית, לא ניתן להקליד את המלל בעברית, מכיוון שלא קיימת במערכות אלו תמיכה בכתובת עברית מימין לשמאל.