

חיפוש והחלפה של תווי יוניקוד בוורד

בין עשרוני להקסדצימלי

פתיחה:

העבודה עם חיפוש תווים מיוחדים במסמכי וורד יכולה להיות מורכבת ומתסכלת, במיוחד כאשר מדובר בתווים מערכת היוניקוד. מאמר זה מסביר את האתגרים העומדים בפני משתמשים בבואם לחפש ולהחליף תווי יוניקוד, הנובעים מחוסר עקביות בשיטות סימון התווים בחלקים שונים של תהליך העבודה - שימוש בערכים הקסדצימליים לצד ערכים עשרוניים. נלמד על הרקע לבעיה, נבין כיצד להתמודד עם מגבלות התוכנה הקיימת, ונדון בדרכים אפשריות ליעול התהליך. בסופו של דבר, נראה כיצד שיפורים קלים יחסית מצד מיקרוסופט יכולים להקל משמעותית על העבודה השוטפת ולחסוך זמן יקר למשתמשים.

המאמר המקורי מופיע באנגלית (כאן), ובמסמך זה עשינו תרגום של המאמר, המלווה בהסבר לכל קטע שתורגם מאנגלית לעברית. למעשה אפשר לקרוא את ההסבר בלבד, הוא בהיר וזורם.

תרגום:

יש לי מסמך שבו ספרות בכתב עילי (superscripts) הוכנסו באמצעות שימוש בספרות עיליות מיוחדות מערכת התווים של יוניקוד (Unicode), ולא כהערות עליונות של תוכנת וורד. לדוגמה, אם אני מסמן אחד מהתווים הללו ולוחץ על 'Alt-x', אני לא מקבל את מה שהייתי מצפה.

הערך של הספרה '6' ביוניקוד הוא 0036. לעומת זאת, אם אני מסמן את הספרה 6 ולוחץ על 'Alt-x', אני מקבל את הערך 2076. 2076 הוא ערך הקסדצימלי (hex) ביוניקוד עבור ספרה 6 קטנה ומורמת. ישנם תווים מיוחדים עבור כל הספרות (וייתכן שגם עבור חלק מהאותיות או כולן).

הסבר:

המאמר מתאר מצב שבו במסמך וורד, ספרות בכתב עילי (כמו חזקות במתמטיקה) לא הוכנסו באמצעות הכלים המובנים של וורד, אלא באמצעות תווים מיוחדים מערכת התווים יוניקוד. כשהכותב מנסה לזהות את הערך של התו באמצעות קיצור המקשים 'Alt-x'¹, הוא מקבל ערך שונה מהמצופה.

למשל, הערך של הספרה הרגילה '6' ביוניקוד הוא 0036, אבל כשהוא מסמן ספרה עילית קטנה של 6 ולוחץ על 'Alt-x', מתקבל הערך 2076, שהוא הערך המיוחד ביוניקוד לאותה ספרה עילית. המאמר מציין שישנם תווים מיוחדים כאלו לכל הספרות, וייתכן שגם לחלק מהאותיות או לכולן.

תרגום:

גופני טקסט ישנים יותר נהגו לכלול תווים מיוחדים עבור הערות עליונות של הספרות 1, 2 ו-3 לצורך ביצוע חזקות, הערות שוליים וכדומה, במסכים שלא היו WYSIWYG (What You See Is What You Get) ולא יכלו להרים את התו בפועל (חשבו על VT100 או דומה). בשלב מסוים, שאר ערכת התווים נכללה, ומכאן המספור הלא רציף. זה די דומה לאופן שבו גופני מתכת אמיתיים היו צריכים לעצב תווי הערה עילית בנפרד. ומנקודת מבט עיצובית, ייתכן שמספר שעוצב לשימוש כהערה עילית יראה טוב יותר מאשר תו 'רגיל' שהורם והוקטן. אז אני לא מתלונן על קיומם של המספרים, אבל אני עובר על מסמך ובזק אם ההערות והמראי המקום ממוספרים ברצף, ואני לא יכול לחפש את ההפניות הצולבות/ציטוטים, כך שזה הופך את העבודה למייגעת ורגישה לשגיאות.

זה קל יחסית לגרום לוורד לחפש תו יוניקוד. אלא שקודים אלה בבירור אינם עשרוניים (8900, נניח). Alt-x מספק את הקוד הקסדצימלי (או מכניס את התו לאחר הקלדת הקוד ההקסדצימלי שלו), אבל וורד מחפש אותו באמצעות ערך יוניקוד עשרוני. כל הכבוד מיקרוסופט! הערך העשרוני של יוניקוד (ו-ASCII) עבור '6' הוא 54. אם אני מסמן '6' במסמך וורד ולוחץ Alt-x כדי למצוא את הקוד שלו, אני מקבל 0036, אבל אני צריך להשתמש ב-00054^u כדי לחפש '6'. כמה טיפשי זה? (חיפוש זה עובד עם או בלי שימוש בתווים כלליים.) זו הסיבה שבטבלה שלי למעלה יש גם ערכים עשרוניים (השתמשתי ב-HEX2DEC () ב-Excel).

¹ כאשר בוחרים תו מסויים ולוחצים על Alt-x, מוצג הערך יוניקוד של התו הנבחר.



אז עכשיו אני הולך לחפש, נניח, $\text{u}8313^{\wedge}$ ולהחליף בספרה עילית 9. אולי ניתן לבצע זאת באופן יותר אוטומטי, אבל יש רק 10 ספרות אפשריות, אז זה מספיק פשוט לבצע 10 החלפות.

הסבר:

הכותב ממשיך לתאר את האתגרים בעבודה עם ספרות של כתב עילי במסמך וורד שהוכנסו כתווים מיוחדים מיוניקוד. הוא מסביר שבעבר, גופני טקסט כללו תווים מיוחדים רק עבור הספרות 1, 2 ו-3 בהערות עליונות, אך בשלב מסוים הורחבו לכלול את כל הספרות ואולי גם אותיות. זה דומה לעיצוב נפרד של תווי הערה עילית בדפוס מתכת.

הבעיה היא שקשה לחפש ולמצוא את ההערות העליונות הללו במסמך כדי לוודא שהן ממוספרות ברצף. וורד יכול לחפש תווי יוניקוד, אבל הוא משתמש בקודים עשרוניים, בעוד שקיצור המקשים Alt-x מספק את הקוד ההקסדצימלי. כדי לחפש את הספרה הרגילה '6', יש להשתמש בקוד $\text{u}0054^{\wedge}$, למרות שהקוד ההקסדצימלי המתקבל מ-Alt-x הוא 0036.

הכותב מתכוון לחפש את הקוד $\text{u}8313^{\wedge}$ ולהחליפו בהערה עילית של הספרה 9. הוא מעיר שאולי ניתן לבצע זאת באופן אוטומטי יותר, אבל מכיוון שיש רק 10 ספרות אפשריות, זה פשוט לבצע 10 החלפות ידניות.

תרגום:

כדי להשתמש בקוד יוניקוד בתיבת 'החלף ב', הדבר הפשוט ביותר הוא להזין את התו למסמך (או לשטח טיפוסה), ואז להעתיק אותו מהמסמך לחלון 'החלף ב'; סימון u^{\wedge} לא יעבוד בחלון ההחלפה. נפש לא נדיבה תאמר שחיפוש והחלפה בוורד "שבורים" מעצם תכנונם. אני לעולם לא הייתי אומר דבר כזה. למרות מה שרואים כאן, דברים כאלה ניתן לבצע באופן אוטומטי באמצעות מאקרו של וורד, אבל זה נראה ככלי די כבד למה שאמור להיות משימה שגרתית. התהליך היה מפושט בהרבה ועומד בכלל 80:20 אם ההתנהגות של Alt-x וחיפוש הייתה מסונכרנת, ואם ניתן היה להשתמש בסימון u^{\wedge} בתיבת 'החלף ב'.

הסבר:

בסיום המאמר, הכותב מסביר כיצד להשתמש בתווי יוניקוד בתיבת 'החלף ב' של פונקציית החיפוש וההחלפה בוורד. הדרך הפשוטה ביותר היא להקליד את התו הרצוי במסמך עצמו או בשטח טיפוסה, ואז להעתיק אותו ישירות לתיבת 'החלף ב'. השימוש בסימון u^{\wedge} לא עובד בחלון ההחלפה.

הכותב מציין בהומור שאדם ביקורתי יאמר שהחיפוש וההחלפה בוורד אינם בנויים נכון כבר מעצם תכנונם, אך הוא עצמו לא היה אומר דבר כזה. הוא מפנה לקישור חיצוני לדיון נוסף בנושא - כאן.

לדבריו, ניתן לבצע פעולות כאלו באופן אוטומטי יותר באמצעות מאקרו של וורד, אך זה נראה ככלי מסורבל מדי למשימה שאמורה להיות שגרתית. התהליך היה יכול להיות פשוט יותר ולעמוד בכלל 80:20 (פארטו) אילו ההתנהגות של Alt-x והחיפוש הייתה אחידה, וניתן היה להשתמש בסימון u^{\wedge} גם בתיבת 'החלף ב'.

סיכום:

לסיכום, המאמר מתאר את האתגרים והמורכבות בעבודה עם הערות עליונות שהוכנסו כתווים מיוחדים מיוניקוד במסמכי וורד, ומציע דרכים להתמודד עם בעיות אלו, תוך ביקורת על חוסר האחידות והאינטואיטיביות של הכלים הקיימים לנושא בתוכנה. הכותב טוען שעם מעט שיפורים והתאמות, התהליך יכול היה להיות הרבה יותר פשוט ויעיל עבור המשתמשים.

בברכה,

ישורון קובי – צפת.

<http://www.kubi.co.il>