

יוניקוד שבור, בואו נתקן אותו!

שי ברגר

shai@kaplanopensource.co.il

@shaib@tooot.im

shai@platonix.com

יוניקוד הוא שיפור גדול

LMDTI LTKNT
BM@RKT B* LA *IW
TWWIM @BRIIM.
NALXNW L@BWD @M
T@TIQ M%WN* .

```
printf("%d", j);
```

Unicode Bidi Algorithm

- מגדיר את התרגום מרצף לוגי של תווים (סדר בזכרון, סדר הקלדה) לרצף של תווים מוצגים



- לוגי: יש כאן 32 כבלי BSU

- מוצג: יש כאן 23 כבלי USB



- (זה עוד יותר סבוך, חלוקה לשורות למשל)

I
B
M

ש
ל
מ
ה

ה
מ
ו
ו

ע
ל

R
H
A
T

איך צריך להציג את זה?

בעצם יש שתי אפשרויות

IBM שלמה המון על RHAT

RHAT שלמה המון על IBM

אולי זה רק תאורטי?



Tomer F. Ashur (cry... @... · Jul 19, 2021 ...

Replying to @GadiAleks

אם הסדרה נכתבת מימין לשמאל כי טוויטר אז האיבר שאחרי 55 הוא 34



Gadi Aleksandrowicz

@GadiAleks

בואו נשחק משחק: הביטו נא בסדרה הבאה (שנכתבת מימין לשמאל כי טוויטר)

55, 34, 21, 13, 8, 5, 3, 2, 1, 1

ואמרו לי - מי האיבר הבא בתור אחרי 55?

וכן, כמובן שזו שאלה מכשילה. אחרת מה הכיף פה.

[Translate Tweet](#)

11:51 PM · Jul 18, 2021

2 Retweets 130 Likes

תקן יוניקוד שבור

- המטרה של (כל) תקן היא תאימות בין מימושים
- טקסט פשוט אמור להיות חד משמעי ברמה הטכנית
- טקסט דו כיווני עלול להיקרא במשמעות הפוכה משנכתב

תקן יוניקוד כולל גם אמצעים לתיקון

- ישנם בתקן תווי בקרת כיווניות

- המוכרים יחסית: LRM, RLM

- השימושיים לפתרון: LRE, RLE, PDF (ועוד)

- הפתרון צריך להיות שקוף

לא לוותר על טקסט פשוט!

- בקבצים סתם
- ב־Markdown ודומיו
- במיילים, עם תוכן `text/plain`
- ב־git