

2010-10-28

**Title:** Standardizing bidirectional emails  
**Author:** Roozbeh Pournader (HighTech Passport)  
**Action:** For discussion at the UTC  
**Date:** 2010-10-28

## Introduction

As of now, there is no guaranteed way to send a bidirectional email and hope that it is read the same way it was sent. The major problems are the sender name, the subject, and the body of plain text emails. The only thing that usually survives, is the body of HTML emails. This greatly reduces the usability of the email in the Middle East and in expatriate communities. This document tries to open the discussion for standardizing the way bidirectional emails are sent and displayed.

The three most popular providers of email software, Google, Microsoft, and Yahoo, treat bidi email very differently.

For example:

- Microsoft's Hotmail service assumes that the body of plain text emails, as well as sender names and subject lines of all emails are in the direction of the UI. For example, a colleague once received a promotional English email from Microsoft in her Hotmail account, with a subject line like "5 ways to reduce debt". But she was using the Arabic UI, and the subject line was displayed as "ways to reduce debt 5".
- Microsoft's Outlook software uses a home brew variation of the bidirectional algorithm at least for subject lines. The algorithm matches parentheses and makes sure they are resolved to the same bidi level. (A presentation about the specifics was made at the recent Unicode Conference by Ayman Aldahleh.)
- Google's Gmail service uses a complicated and proprietary home brew algorithm for determining the direction of plain text. It is based on word counts and 40% RTL words make the whole email RTL. There is also some dependency on the UI direction, as some subject lines are treated differently with LTR UI directions.

Gmail's behavior has an interesting effect on emails sent by Facebook, for example. Facebook sends email notifications when a friend comments on someone's status. There is some English header and footer, explaining why the email was sent, and then in the middle of the email, there is the text of the comment itself.

Interesting things happen when the comment is RTL (or mostly RTL). If the comment is relatively short, the email is resolved to be LTR, and everything including the comment is displayed LTR. But if the comment is longer than a certain threshold, everything is displayed LTR, including the English header and footer. This has happened to the emails received by the author hundreds of times.

- Current versions of Yahoo Mail do not treat bidirectional text especially. Apart from the body of HTML emails using explicit "`dir=rtl`" tags, nothing receives any special treatment. This makes most use cases except the very common to get displayed incorrectly.
- No popular vendor provides a way for a plain text emails to be displayed in multiple directions. For example, a standard-complying Arabic script email containing three lines of a US address has no way of

getting displayed correctly. Either the whole thing is displayed RTL with the address jumbled, or the whole thing is displayed LTR, with the numbers and English words jumbling the RTL lines.

The author has been a victim of such problems: a package he once received from his family in the Middle East had his address listed as “OAKLAND RD STE B202 1590”. He had sent an email to his family with his address, and the whole email, including the lines containing the postal address, was resolved to RTL by Gmail when displayed. The address line was displayed just like that to his brother, who then copied it to the envelope.

There are many other issues too, like the way the email clients send out email. It is not even guaranteed that the email sent from the same vendor’s software with the same settings (for example, in the jumbled postal address example above, both the sender and the recipient were using Gmail, with RTL support enabled).

The author believes that an attempt should be made to recommend some standard practices for bidi email. The alternative, reverse engineering every major player’s behavior and then sending especially tailored email or displayed emails sent from them using a special bidi algorithm is too expensive for any vendor.

For example, the UTC can recommend that in the absence an extra higher level standard from IETF or such, all plain text fields in emails should be treated explicitly according to the Unicode Bidirectional Algorithm (especially steps P1–P3). The recommendation should be independent of the UI direction, as the UI direction is no hint towards the actual information in an email sent from someone else.

This would make bidirectional plain text email possible, as well as predictably displayed subject and sender lines.