Localizing to right-to-left languages: Main Issues and Best Practices

Murhaf Hossari
University College Dublin
Murhaf.hossari@gmail.com

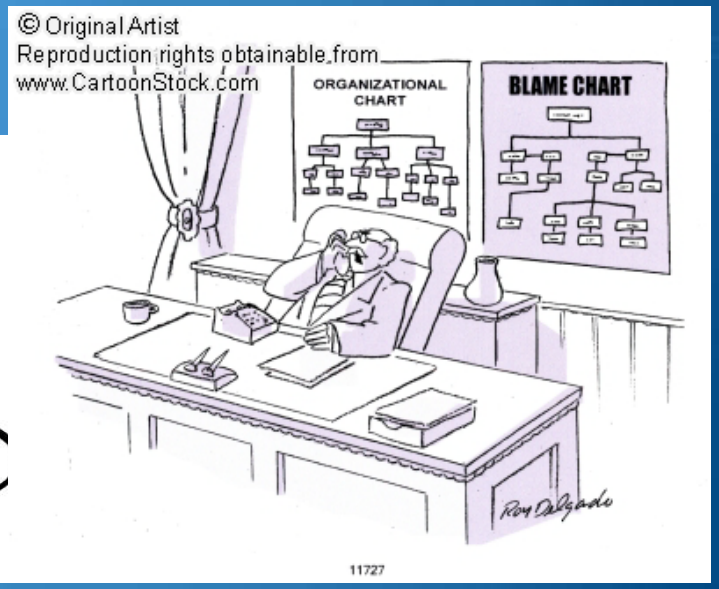
Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Right-to-left languages
Story

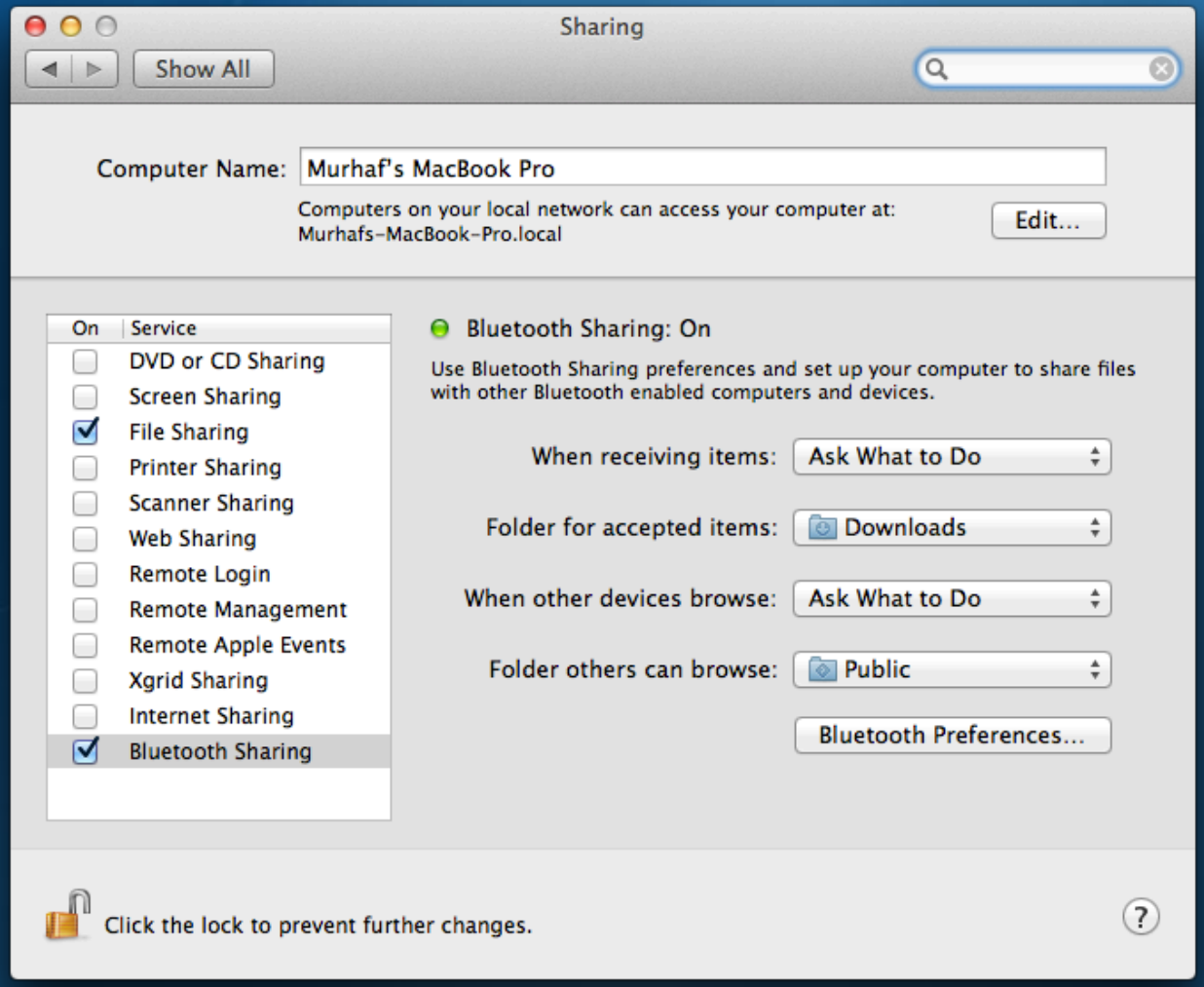


Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

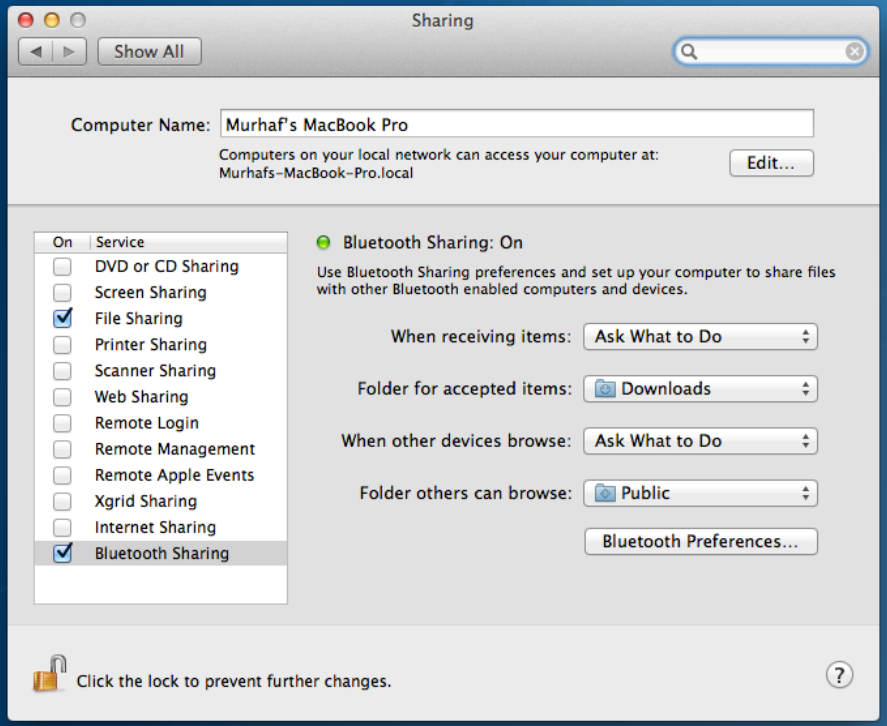
Localizing to RTL Languages

RTL Layout



Localizing to RTL Languages

RTL Layout

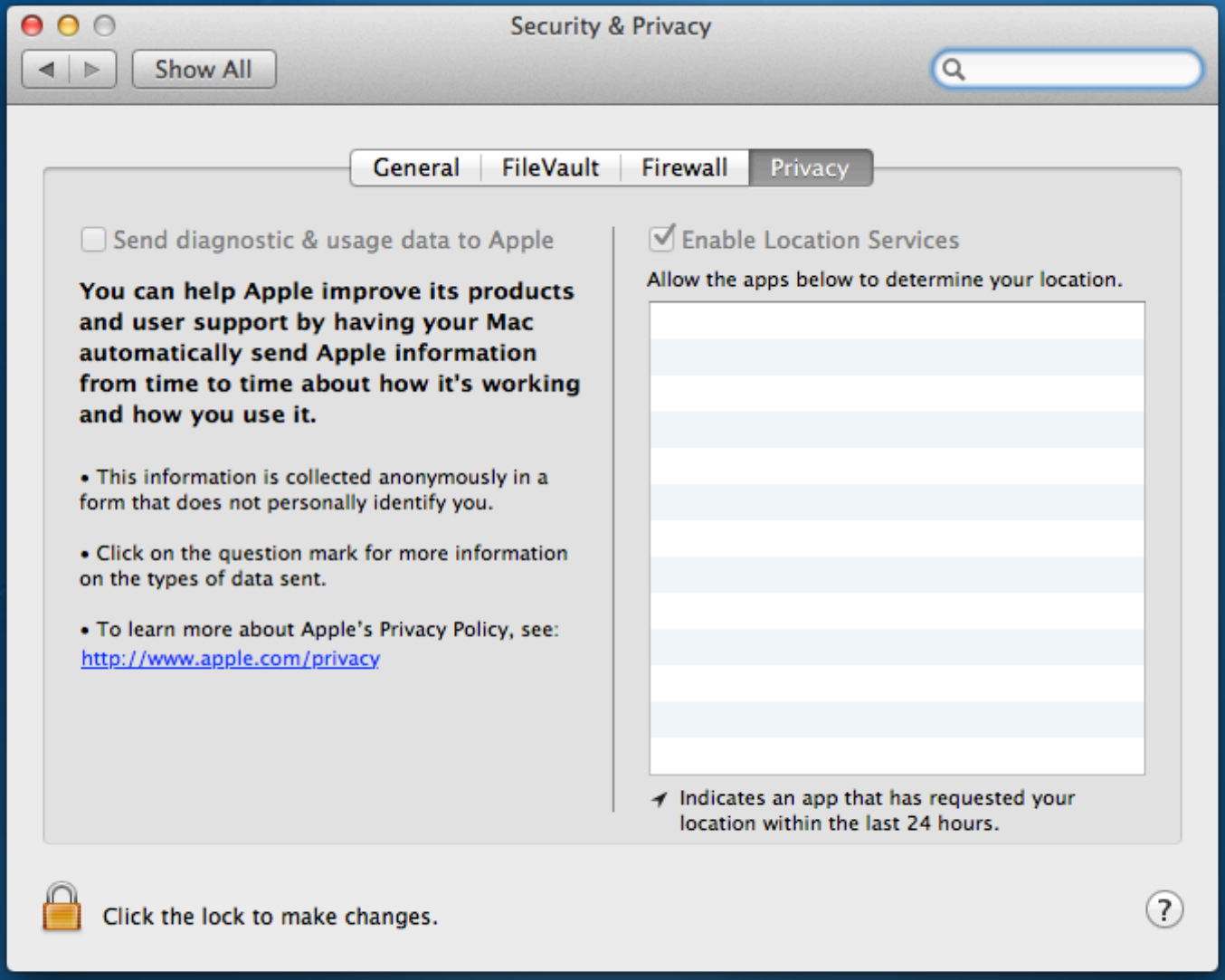


Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Localizing to RTL Languages

Text Justification



Localizing to RTL Languages

Text Justification

Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Directionality Support

Bidirectional Text

Left-to-right text:

Doubt is a pain too lonely to know that faith is his twin brother.

Right-to-left text:

.rehtorb niwt sih si htiaf taht wonk ot ylenol oot niap a si tbuoD

Bidirectional text:

Doubt is a pain too lonely to know rehtorb niwt sih si htiaf taht.

Directionality Support

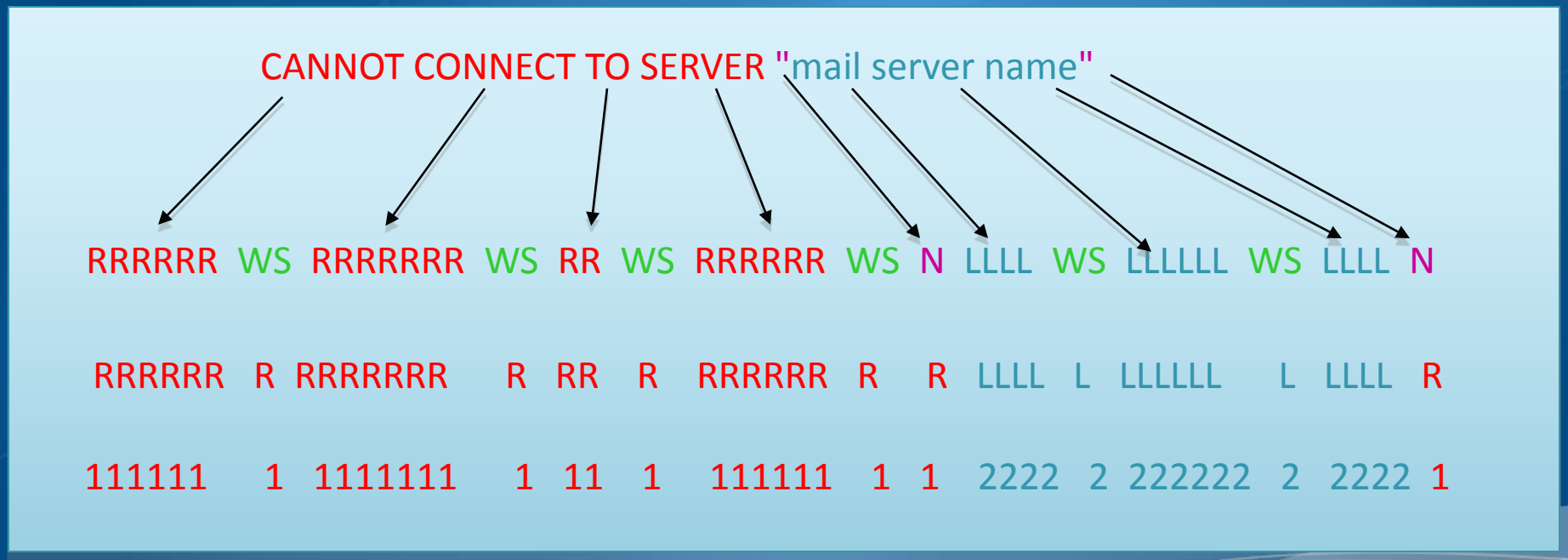
Unicode Bidirectional Algorithm (UBA)

- Text is always entered in the same way
- Different languages have different character flow
- Reorder the characters and define their directionality properties (R, L, N…etc)
- Set of rules to produce the final correct visual representation

Directionality Support

Unicode Bidirectional Algorithm

Example



Directionality Support

Unicode Bidirectional Algorithm - Issues

- Correct layout for most of the cases
- Issues with some cases due to:
  - *Paragraph direction isn't detected correctly*
  - *Complicated nesting of strings of different types*
  - *Strings with special nature such as part numbers, names …etc*
  - *Ambiguous strings even for human eye*

Directionality Support

Unicode Bidirectional Algorithm - Issues

Examples

Incorrect Paragraph Direction (Base Direction)

Apple شركة أمريكية متعددة الجنسيات ✔

شركة أمريكية متعددة الجنسيات Apple

Directionality Support

Unicode Bidirectional Algorithm - Issues

Examples

Weak Characters

شركة Yahoo!

تطبيق.doc

تطبيق.doc

Directionality Support

Unicode Bidirectional Algorithm - Issues

Examples

Weak Characters 2

قطع الاتصال VPN (PPTP) ✔

قطع الاتصال (VPN (PPTP

Directionality Support

Unicode Bidirectional Algorithm – Directional Codes

- Invisible characters to help UBA showing the correct layout
- Explicit directional codes manually added to enforce the correct layout

Directionality Support

Unicode Bidirectional Algorithm – Directional Codes - Pitfalls

a)Manually Added
b)Not trivial to use
c)Invisible!
d)Need to check them at runtime

Directionality Support

Areas of improvements

- Finding patterns in the cases that usually need help
- Detect those patterns in the early stages by parsing the strings
- Deal with those cases in a special way by UBA

Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Localizing to RTL Languages
Numbers

- Some scripts use different numbers
- Localizing numbers can be tricky
- Support for localized numerals in web is very little

(١, ٢, ٣, ٤, ٥, ٦, ٧, ٨, ٩, ٠)

(1, 2, 3, 4, 5, 6, 7, 8, 9, 0)

Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Best Practices

- Support of RTL Layout
- Avoid composed strings
- Avoid weak and neutral characters
- Don’t enforce text direction
- Support of localized numbers
- Support of different locales

Outline

- Right-to-Left Languages
- Localizing to right-to-left languages
  - Right-to-left layout
  - Text justification
  - Directionality support
    - Unicode Bidirectional Algorithm
    - Issues and areas of improvements
  - Numbers
- Best Practices
- Summary

Summary

- Need to pay extra attention when localizing to RTL languages
- Current support for RTL languages is very good
- Some improvements can be added to overcome persistent issues

Localizing to RTL Languages

Thank you

Murhaf Hossari