fantasai notes that there are a lot of Iranians here today :) 
Err, except i've just looked at http://wiki.freedesktop.org/wiki/Software_2fFonts_ and can't actually see the logs (after a quick scan).

behdad: Are you and zwnj twin brothers :-)

fantasai: lets remove Latin from DejaVu fonts :-D

behdad: :D

behdad: cool!
Advanced OpenType alternate glyph forms!

CSS first-letter styling!

Advanced technology from the Book of Kells, c. 1000 C.E., over 1000 years old!
Superior font technology!

CSS first-word pseudo class!
CSS first-line pseudo class!
Unfortunately the Spell Checker still needed some work ...
And the justification algorithm wasn't perfect yet ...
International Text Layout & Typography: the Big & Future Picture

- Comprehensive review of issues
- Unified text layout
Directionality & Block Progression


- Prelude to CSS3 text layout model
- Comprehensive
- Does not break existing Unicode BIDI or CSS models
Directionality
&
Block Progression

Before we begin writing, we need to look at:

- *Inline progression*
- *Block progression*
Directionality & Block Progression

CSS3 Model

**Inline Progression**

- left-to-right (ltr) - Latin
- right-to-left (RTL) - Arabic
- top-to-bottom (ttb) - 中文, 日文
- bottom-to-top (btt) - Runic
Directionality & Block Progression

CSS3 Model

Block Progression

- top-to-bottom (ttb) - Latin
- right-to-left (rtl) - 中文, 日文
- Left-to-right (ltr) - Mongolian
Directionality & Block Progression

Scripts are generally constrained to a subset of the possible combination of values for these two properties.
Directionality & Block Progression

However unusual cases can and do occur in real life with surprising frequency.
Directionality & Block Progression

Mongolian - Japanese dictionary set ttb-rtl
Directionality & Block Progression

Uyghur - Chinese - Russian dictionary with Uyghur & Chinese both set rtl.
There isn’t just one right way to do it ...
Text Directionality in Titles

Title page of a Chinese dictionary set rtl in order to match content’s rtl block progression
Mirror
Writing ...

... It’s not just on ambulances

...for it’s all in some language I don’t know,” she said to herself. It was like this.

ARBREMOCKY

She puzzled over this for some time, but at last a bright thought struck her. “Why, it’s a
Mirror Writing

... in Hebrew

... goes from left to right
Boustrophedon
A GPL'ed boustrophedon text reader ...

The licenses for most software are designed to take away your freedom to share and change it. By contrast, the GNU General Public License is intended to guarantee your freedom to share and change free software--to make sure the software is free for all its users. This General Public License applies to most of the Free Software Foundation’s software and to any other program whose authors commit to using it. (Some other Free Software Foundation software is covered by the GNU Library General Public License instead.) You can apply it to your programs, too.

When we speak of free software, we are referring to freedom, not price. Our General Public Licenses are designed to make sure that you have the freedom to distribute copies of free software (and charge for this service if you wish) that you receive source code or can get it if you want it, that you can ...
Vertical Typography

Pango now knows how to do it ...
Traditional Mongolian ...
... should be the next target
Mongolian Fonts
... are going the wrong way
Mongolian Fonts

That’s better!

<table>
<thead>
<tr>
<th>𑠭</th>
<th>𑠯</th>
<th>𑠫</th>
<th>𑠬</th>
<th>𑠭</th>
<th>𑠧</th>
</tr>
</thead>
<tbody>
<tr>
<td>ࠬ</td>
<td>ࠫ</td>
<td>ࠬ</td>
<td>ࠬ</td>
<td>ࠬ</td>
<td>ࠬ</td>
</tr>
</tbody>
</table>
Advanced Typographical Features ...

- Glyph Substitution
- Glyph Positioning
- Baseline Adjustments
- Vertical Metrics
- Stylistic Alternates
- Optional Ligatures
- Swash Forms
- etc ...
Advanced Typographical Features ...

- *Smart cursor positioning*
- *Smart highlighting*
# Backspacing Example

*(Going forward would yield similar results)*

<table>
<thead>
<tr>
<th>OpenOffice.org</th>
<th>A Better Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>What you actually see</td>
<td>How it should work</td>
</tr>
<tr>
<td>What is really selected</td>
<td>(A)</td>
</tr>
<tr>
<td></td>
<td>How it should work</td>
</tr>
<tr>
<td>(1) हिंदी</td>
<td>(1) हिंदी</td>
</tr>
<tr>
<td>(2) हिंदी</td>
<td>(2) हिंदी</td>
</tr>
<tr>
<td>(3) हिंदी</td>
<td>(3) हिंदी</td>
</tr>
<tr>
<td>(4) हिंदी</td>
<td>(4) हिंदी</td>
</tr>
<tr>
<td>(5) हिंदी</td>
<td>(5) हिंदी</td>
</tr>
<tr>
<td>(6) हिंदी</td>
<td>(6) हिंदी</td>
</tr>
</tbody>
</table>
The backspacing example just proves ...
The backspacing example just proves ...

... that the software was designed for Western scripts, not Eastern Scripts ...
Available Technologies:

• **OpenType:**
  – FLOSS HarfBuzz library
  – Originally part of FreeType
  – Forked into Pango & QT
  – Now merged again

• **Graphite:**
  – Developed by SIL
  – Advanced system for complex scripts like Myanmar & Khmer
Graphite

- Full range of character-to-glyph mappings:
  - one-to-one
  - one-to-many
  - many-to-one
  - many-to-many
- x-y position adjustments (i.e., for diacritical mark stacking)
- advance width modification
- glyph attachment chains
- glyph clusters
Graphite

- Ligatures are not monolithic
- Underlying characters are accessible
- Access rectangular areas corresponding to glyph components
- Highlighting & mouse selection handled by Graphite
- No complex programming required
Floss Developers Mandate

- Integrate & provide access to OpenType features
- Also seamlessly integrate Graphite
- Also include Apple's AAT
- Provide a single, unified API for access to the common features of all three technologies
Word Breaking & Syllabification

No spaces occur in:

- Thai
- Lao
- Khmer (Cambodian)
- Myanmar (Burmese)
- ...
Word Breaking & Syllabification

No spaces occur in:

- Lanna (Unicode proposal voted on Sept. 2006)
- Tai Dam (Viet Tai, Unicode proposal exists)
- Tai Le (Dehong Dai, now in Unicode)
- New Tai Le (now in Unicode)
Look at the following sentence and compare the Burmese writing with the version in transcription:

აერთეჭირთ აერზე დაახ
/qayin-ga. sa-dai'hma qalou' lou'te/

"Formerly (I) worked in the post office."

The first thing you will probably have noticed is that there are no spaces separating words in the Burmese writing. At first this may give you some difficulty in reading, but as soon as you are sure of the individual words you will have no difficulty in knowing how to divide a sentence into its parts correctly. Where there is a major break, such
Word Breaking in Thai

การทำการบ้าน

ทำ การ บ้าน
3 words, no errors, shorter word length

ทำ การบ้าน
2 words, no errors, longer word length
Word Breaking & Syllabification

- Need for a lightweight, object-oriented, pluggable framework for word boundary analysis:
- for Southeast Asian “spaceless” scripts ...
- ... but also for Western scripts:
  - syllabification analysis
  - hyphenation analysis
Fonts are the first step in making advanced typography both possible and convenient.

Arabic OpenType font with set of extended ligatures

Arabic OpenType font without set of extended ligatures
Fonts are the first step in making advanced typography both possible and convenient.

The computer as a tool should open up new avenues for beautiful and creative typography rather than simply reinforce the limitations of the technology of the previous era.
Preparing for New Scripts in Unicode

UC Berkeley Scripts Encoding Initiative:

• Over 100 scripts listed
• Only one-fifth approved so far for inclusion in Unicode
• Some don't have proposals written yet
• Some proposals exist with insufficient quality
• ...

Preparing for New Scripts in Unicode

UC Berkeley Scripts Encoding Initiative:

• Some scripts quite simple ltr alphabets
• But others have special features
• And some have unique features not shared by other scripts
Preparing for ... Lanna
Font Configuration & Customization

Vietnamese expect more vertical spacing ...

**DejaVu Sans**
Đọc cha ăn mắm,
dôi con khát nước. 7pt.
Đọc cha ăn mắm,
dôi con khát nước. 8 pt.
Đọc cha ăn mắm,
dôi con khát nước. 10 pt.
Đọc cha ăn mắm,
dôi con khát nước. 12 pt.

**VU Pho Tho**
Đọc cha ăn mắm,
dôi con khát nước. 7 pt.
Đọc cha ăn mắm,
dôi con khát nước. 8 pt.
Đọc cha ăn mắm,
dôi con khát nước. 10 pt.
Đọc cha ăn mắm,
dôi con khát nước. 12 pt.
Font Configuration & Customization

Chinese & Japanese become acrimonious when forced to use each other's fonts ...
Font Configuration & Customization

What’s wrong here?

<alias>
  <family>DejaVu Serif</family>
  <family>Bitstream Vera Serif</family>
  <family>Times New Roman</family>
  ...
  <family>Luxi Serif</family>
  <family>Kochi Mincho</family>
  <family>Sazanami Mincho</family>
  <family>AR PL ZenKai Uni</family>
  <family>AR PL SungtiiL GB</family>
  <family>AR PL Mingtii2L Big5</family>
  <family>MS 明朝</family>
  <family>Baekmuk Batang</family>
  <family>FreeSerif</family>
  <family>MgOpen Canonica</family>
  <default><family>serif</family></default>
</alias>
Font Configuration & Customization

A Better Way:
Configuring Fonts by Script & Orthography

<script>
  <name>latn</name>
  <orthography>
    <name>vn</name>
    <alias>
      <family>VU Pho Tho</family>
      <family>DejaVu Sans</family>
      <default><family>sans</family></default>
    </alias>
  </orthography>
</script>
Enhancing Development Tools for Better Typography
Enhancing Development Tools for Better Typography

(I thought I had this brilliant idea, but it turns out someone else thought of it before me: )

“... I'm primarily thinking about network access to the font repository. This basically boils down to which version control tool is chosen, and how it is set up. Should each font be one big text file, or perhaps a directory with a separate file for each glyph?”

- Raph Levien, April 2006
Enhancing Development Tools for Better Typography

Special “Filesystem”

Save Format:

- Extended SVG XML font format
- One file per glyph
- Use git, SVN, SVK, any SCM system you want
Wrap Up:

Opportunities for the FLOSS community in the Development Pipeline

- Take a more active role in pushing the Unicode Standard forward to the next milestone
- Collaboration opportunities with non-traditional partners is already proving fruitful
- Move into “high gear” to implement new scripts as they are approved in Unicode
Thanks!

I would especially like to thank the following people, in random order, for answering my questions, providing feedback on this paper, and providing good advice:

- Peter Moulder
- Andreas Vox
- Behdad Esfahbod, Pango
- Sharon Correll, SIL
- Ngwe Tun, Myanmar Unicode & NLP Center
- Elika J. Etemad, W3C CSS Working Group
- Martin Hosken, SIL
- Dr. Deborah Anderson, Scripts Encoding Initiative (SEI), UC Berkeley
- Nicolas Spalinger, SIL
- Simos Xenitellis
- Clytie Siddall, Vietnamese free-software translation team
- Daniel Glassey
- Have I forgotten anyone?