Markdown to PDF

Sitaram Chamarty

TCS Innovation Labs
Hyderabad

22 July, 2009 / bragging

TATA CONSULTANCY SERVICES

Outline

1. the problem
   - office software sucks
2. the pieces start to fall in
   - text to HTML
   - LaTeX to PDF
3. the last piece
   - HTML to LaTeX
4. bonus: images
   - aka: text to graphics
5. the end

TATA CONSULTANCY SERVICES

Outline

1. the problem
   - office software sucks
2. the pieces start to fall in
   - text to HTML
   - LaTeX to PDF
3. the last piece
   - HTML to LaTeX
4. bonus: images
   - aka: text to graphics
5. the end

TATA CONSULTANCY SERVICES

all office software sucks

- but presentation software sucks more
- it requires too much mousing around
  - doesn't matter if it is OpenOffice or MS Office
  - I hate them both equally
- and I hate mice
  - nasty, smelly, things
  - never even have a bath, what a life...

TATA CONSULTANCY SERVICES

all office software sucks

- but presentation software sucks more
- it requires too much mousing around
  - doesn't matter if it is OpenOffice or MS Office
  - I hate them both equally
- and I hate mice
  - nasty, smelly, things
  - never even have a bath, what a life...

TATA CONSULTANCY SERVICES

all office software sucks

- but presentation software sucks more
- it requires too much mousing around
  - doesn't matter if it is OpenOffice or MS Office
  - I hate them both equally
- and I hate mice
  - nasty, smelly, things
  - never even have a bath, what a life...

TATA CONSULTANCY SERVICES

My presentations

- all these years, my presentations have consisted of
  - plain text on plain white backgrounds
  - with hardly any pictures
    - I start sweating if I have to make a picture or a chart
    - even if I'm using OpenOffice

TATA CONSULTANCY SERVICES

My presentations

- all these years, my presentations have consisted of
  - plain text on plain white backgrounds
  - with hardly any pictures
    - I start sweating if I have to make a picture or a chart
    - even if I'm using OpenOffice

TATA CONSULTANCY SERVICES

My presentations

- all these years, my presentations have consisted of
  - plain text on plain white backgrounds
  - with hardly any pictures
    - I start sweating if I have to make a picture or a chart
    - even if I'm using OpenOffice

TATA CONSULTANCY SERVICES

My preferred editor

- for everything and anything under the sun
- is "vim"
- even my firefox browser uses "vimperator", an extension which lets me use vi keystrokes instead of the mouse :-)

TATA CONSULTANCY SERVICES

My preferred editor

- for everything and anything under the sun
- is "vim"
- even my firefox browser uses "vimperator", an extension which lets me use vi keystrokes instead of the mouse :-)

TATA CONSULTANCY SERVICES

My preferred editor

- for everything and anything under the sun
- is "vim"
- even my firefox browser uses "vimperator", an extension which lets me use vi keystrokes instead of the mouse :-)

TATA CONSULTANCY SERVICES

so the problem is this:

- I want to make presentations using plain text
- and if they can look prettier and feel slicker, that's a bonus

TATA CONSULTANCY SERVICES

so the problem is this:

- I want to make presentations using plain text
- and if they can look prettier and feel slicker, that's a bonus

TATA CONSULTANCY SERVICES

so the problem is this:

- I want to make presentations using plain text
- and if they can look prettier and feel slicker, that's a bonus

TATA CONSULTANCY SERVICES

in the interest of time

(...and your sanity)

I will not bore you with all the stuff that failed...

TATA CONSULTANCY SERVICES

Outline

1. the problem
   - office software sucks
2. the pieces start to fall in
   - text to HTML
   - LaTeX to PDF
3. the last piece
   - HTML to LaTeX
4. bonus: images
   - aka: text to graphics
5. the end

TATA CONSULTANCY SERVICES

simple HTML made even simpler

- I'd been using Markdown for a year or so now
- Markdown is one of the seventeen thousand or so markup languages in the world
  - very simple text to HTML conversion
  - indentation based for easy lists
  - italics is like `*italics*`
  - bold is like `**bold**`
    - ...and so on; more details here

TATA CONSULTANCY SERVICES

Outline

1. the problem
   - office software sucks
2. the pieces start to fall in
   - text to HTML
   - LaTeX to PDF
3. the last piece
   - HTML to LaTeX
4. bonus: images
   - aka: text to graphics
5. the end

TATA CONSULTANCY SERVICES

detour: LaTeX

- in the beginning, Don Knuth created TeX
  - (yes, I know it's blasphemy to not format that correctly)
- then Leslie Lamport created LaTeX
  - most popular and powerful text processing language in academia
- then Till Tantau came up with Beamer

TATA CONSULTANCY SERVICES

beamer kicks ass

- it produces PDF
- some fantastic PDF actually
- look at the slide navigation on this one and the other two
- try clicking around to go to other parts of the PDF
- I defy anyone to come up with this kind of navigation in MS or OpenOffice!
  - I'm only showing three themes; there are many more
  - and you can make your own (in fact all these have a subtle mod that my `mdbeamer` produces)

TATA CONSULTANCY SERVICES

beamer kicks ass

- it produces PDF
- some fantastic PDF actually
- look at the slide navigation on this one and the other two
- try clicking around to go to other parts of the PDF
- I defy anyone to come up with this kind of navigation in MS or OpenOffice!
  - I'm only showing three themes; there are many more
  - and you can make your own (in fact all these have a subtle mod that my `mdbeamer` produces)

TATA CONSULTANCY SERVICES

beamer kicks ass

- it produces PDF
- some fantastic PDF actually
- look at the slide navigation on this one and the other two
- try clicking around to go to other parts of the PDF
- I defy anyone to come up with this kind of navigation in MS or OpenOffice!
  - I'm only showing three themes; there are many more
  - and you can make your own (in fact all these have a subtle mod that my `mdbeamer` produces)

TATA CONSULTANCY SERVICES

beamer kicks ass

- it produces PDF
- some fantastic PDF actually
- look at the slide navigation on this one and the other two
- try clicking around to go to other parts of the PDF
- I defy anyone to come up with this kind of navigation in MS or OpenOffice!
  - I'm only showing three themes; there are many more
  - and you can make your own (in fact all these have a subtle mod that my `mdbeamer` produces)

TATA CONSULTANCY SERVICES

beamer kicks ass

- it produces PDF
- some fantastic PDF actually
- look at the slide navigation on this one and the other two
- try clicking around to go to other parts of the PDF
- I defy anyone to come up with this kind of navigation in MS or OpenOffice!
  - I'm only showing three themes; there are many more
  - and you can make your own (in fact all these have a subtle mod that my `mdbeamer` produces)

TATA CONSULTANCY SERVICES

so all I need is...

...some way to convert HTML to LaTeX-beamer syntax

TATA CONSULTANCY SERVICES

Outline

1. the problem
   - office software sucks
2. the pieces start to fall in
   - text to HTML
   - LaTeX to PDF
3. the last piece
   - HTML to LaTeX
4. bonus: images
   - aka: text to graphics
5. the end

TATA CONSULTANCY SERVICES

is "mdbeamer.pl"

- a 150-line perl program I wrote to convert HTML to beamer
- very simple, but handles all the markups I care about
- some parts of it feel a little kludgy because of the HTML in between

TATA CONSULTANCY SERVICES

Outline

1. the problem
   - office software sucks
2. the pieces start to fall in
   - text to HTML
   - LaTeX to PDF
3. the last piece
   - HTML to LaTeX
4. bonus: images
   - aka: text to graphics
5. the end

TATA CONSULTANCY SERVICES

detour: graphviz

- I'd recently discovered graphviz
- excellent for drawing simple diagrams
- for example, this code

```
digraph {
    node[fontsize=24]
    a -> b -> c -> d
    b -> p -> q -> x
    p -> y
}
```

TATA CONSULTANCY SERVICES

detour: graphviz

- produces this:

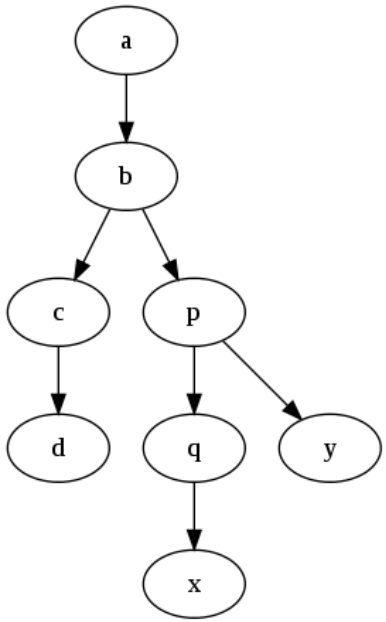

Figure: a git tree?

and something like this...

```
digraph G {
    subgraph clusterCS {
        label="Chennai server\n\ \ \ "

        cs2 [shape=box, label="bare repo\non server", style=filled, fillcolor=green]
        node [style=invis]
        edge [style=invis]
        cs1 -> cs2
    }
    subgraph clusterCL {
        label="Commits on\nChennai Lead PC"

        node [shape=box, style=rounded, style=filled, fillcolor=lightblue]
        c1 [label = "Commit #1\n.gitignore"]
        c2 [label = "Commit #2\nSource\nFiles"]
        c1 -> c2
    }
    cs2 -> c2 [lhead=clusterCL, ltail=clusterCS, label = "push", dir=back, color=red, constra
}
```

TATA CONSULTANCY SERVICES

produces this

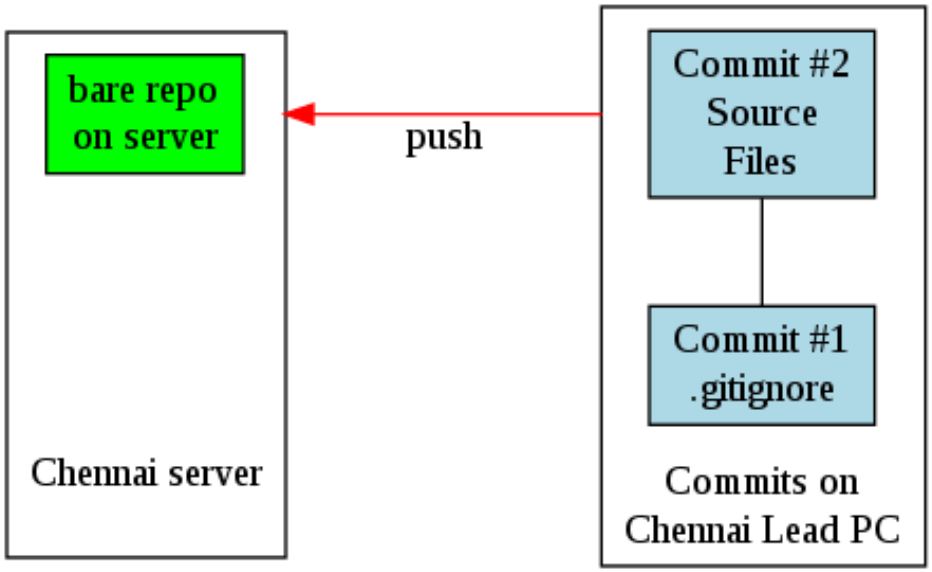

Figure: a more complicated figure

TATA CONSULTANCY SERVICES

so really the last piece

- well, I'm taking the HTML and converting it to LaTeX anyway
- so, devise a simple syntax to embed graphviz code directly into the text
- and make my `mdbeamer` program
  - extract that code
  - call graphviz
  - produce the image

And.... we're done. I can do pretty much everything in text now!
And ***everything*** stays in one simple text file!

TATA CONSULTANCY SERVICES