html2PraatMan
Version 1.0

Charalampos Karypidis
ch.karypidis@gmail.com
http://addictiveknowledge.blogspot.com/

October 16, 2013

Contents

1   Introduction

This tool allows for a batch conversion of all .html files in a given directory into
.man pages (documentation files for the Praat software). It has been written
and tested in Python 2.7 under Windows XP. It is available for free at `http:
//addictiveknowledge.blogspot.com/`

2   Why bother with this tool?

I can see two main reasons for using html2PraatMan:

- it is easier to write HTML code, especially with an HTML editor;

- html2PraatMan allows for batch processing of multiple .html files at a
  time.

3   .html tags and .man commands

The syntactic structure of .man pages is described on Praat's web site. Although
you should read that resource before going on with the present guide, we will
revisit most aspects of the .man file format here.

The complete .html and .man source codes of the example presented in this
guide can be found in sections 6 and 7, respectively. Section 8 present a list of
the HTML tags you may use and their corresponding commands in the .man
files.

3.1   Header

HTML

```
<address>
Charalampos  KARYPIDIS
20130930
0
</address>
```

Praat

```
ManPagesTextFile
"Example"  "Charalampos  KARYPIDIS"  20130930  0
```

The first line of a .man page should only contain the string `ManPagesTextFile`
and will be generated automatically.

The second line contains four bits of information. The title of the page should be identical to the .html filename (without the extension) and will be extracted automatically. The rest of the information should be entered manually using the <address> tag in the following way:

1. First line: the author's name

2. Second line: the date in the "yyyymmdd" format

3. Third line: the recording time

The <address> tag may be entered anywhere in the <body> section. You should, however, respect the order in which the information is presented within it: author, date, recording time. If the recording time is omitted, the value will be equal to 0.

3.2 Main text

3.2.1 Strings

HTML

```
<cite>Lorem ipsum dolor sit amet, consectetur
    adipisicing elit, sed do eiusmod tempor incididunt
     ut labore et dolore magna aliqua.</cite>
<h1>Lorem ipsum</h1>
<p>Ut enim <b>ad minim veniam</b>, <i>quis nostrud</i
    > exercitation <kbd>ullamco laboris</kbd> nisi ut
    <sub>aliquip</sub> ex ea<sup>commodo</sup>
    consequat.</p>
<blockquote>This is a paragraph with a blank left
    margin.</blockquote>
```

Praat

```
<intro> "Lorem ipsum dolor sit amet, consectetur
    adipisicing elit, sed do eiusmod tempor incididunt
     ut labore et dolore magna aliqua."
<entry> "Lorem ipsum"
<normal> "Ut enim ##ad minim veniam#, %%quis nostrud%
    exercitation $$ullamco laboris$ nisi ut
    __aliquip_ ex ea^^commodo^ consequat."
<definition> "This is a paragraph with a blank left
    margin."
```

The <cite> tag is used for the introduction part of the .man page. Section titles are enclosed in <h1> whereas paragraphs are enclosed in <p>. When paragraphs with a blank left margin are needed, <blockquote> should be used.

<b> is used for text in bold, <i> for italics, <kbd> for text in monospace and <sub> for subscripts.

3.2.2 Links

HTML

```
<p>Visit <a href="welkom.man">another .man page</a></
    p>
<p>Click <a href="script.praat" alt="bli|bla|blu">
    here</a> to draw a spectrogram.</p>
<p>Click <a href="sounds\i.wav">here</a> to hear a
    synthetic [i] sound.</p>
```

Praat

```
<normal> "Visit @@welkom|another .man page@"
<normal> "Click @@\SC""script.praat"" ""bli"" ""bla""
    ""blu"" |here@ to draw a spectrogram."
<normal> "Click @@\FIsounds\i.wav |here@ to hear a
    synthetic [i] sound."
```

Links are introduced with the <a> tag. The <a> string will be used as the clickable text whereas the "href" value will be the file target. If the target is another .man page, it should be located within the same directory. If the target is a praat script, any arguments should be entered using the "alt" attribute and separated by "|".

When the target is a praat script or an audio file, both relative and absolute file paths are possible.

3.2.3 Pictures as embedded scripts

HTML

```
<code width="6" height="3">
Create Sound from formula... sineWithNoise Mono 0.0
    1.0 44100 1/2*sin(2*pi*377*x)+randomGauss(0,0.1)
To Spectrogram... 0.005 5000 0.002 20 Gaussian
Paint... 0 0 0 0 100.0 yes 50.0 6.0 0.0 yes
plus Sound sineWithNoise
Remove
```

```
</code>
```

Praat

```
<script> 6 3 "
Create Sound from formula... sineWithNoise Mono 0.0
    1.0 44100 1/2*sin(2*pi*377*x)+randomGauss(0,0.1)
To Spectrogram... 0.005 5000 0.002 20 Gaussian
Paint... 0 0 0 0 100.0 yes 50.0 6.0 0.0 yes
plus Sound sineWithNoise
Remove
"
```

Embedded praat scripts should be enclosed in a <code> tag. Width and
height values should be entered with the "width" and "height" attributes. The
actual code should start on the second line.

3.2.4   Lists

HTML

```
<p>This is a list with no leading bullets</p>
<ul class="noBullet">
<li>I</li>
<li>love</li>
<li>Praat</li>
</ul>
<p>Yet another list, with bullets this time.</p>
<ul>
<li>Praat</li>
<li>saves</li>
<li>time</li>
</ul>
```

Praat

```
<normal> "This is a list with no leading bullets"
<list_item> "I"
<list_item> "love"
<list_item> "Praat"
<normal> "Yet another list, with bullets this time."
<list_item> "\bu Praat"
<list_item> "\bu saves"
```

```
<list_item> "\bu time"
```

Use the <ul> tag to start a list and <li> to introduce each list item. By default, items will be preceded by a bullet sign. If you do not need bullets, pass a "class="noBullet"" attribute to <ul>.

4 How to use html2PraatMan

To properly use html2PraatMan, you will need to:

1. install Python 2.7 (I have not tried using earlier versions);

2. install the Beautiful Soup module library for Python;

3. put all your .html files in the same directory;

4. make sure those files are syntactically correct or else the html parsing might fail;

5. drop the html2PraatMan.py in that directory and double-click it;

6. enjoy the result!

5 Known issues

The following issues will be addressed and should be corrected in version 2.0:

- Within <p>, <h1>, <li> <code> or <blockquote>, tag nesting is not allowed. For instance, this:

  `<p>Something something <b><i>something</i></b>.</p>`

  will raise an exception. This, however:

  `<p>Something something <b>something</b>, <i>more something</i>.</p>`

  is permissible.

- If you need to enter double-quotes within a tag (<p>, <h1>, ...), you will have to double them ("").

- Phonetic symbols are not supported. You should enter them manually.

6 Appendix A: example.html

HTML

```
<html>
<body>
<cite>Lorem ipsum dolor sit amet, consectetur
    adipisicing elit, sed do eiusmod tempor incididunt
     ut labore et dolore magna aliqua.</cite>
<h1>Lorem ipsum</h1>
<p>Ut enim <b>ad minim veniam</b>, <i>quis nostrud</i
    > exercitation <kbd>ullamco laboris</kbd> nisi ut
    <sub>aliquip</sub> ex ea<sup>commodo</sup>
    consequat.</p>
<blockquote>This is a paragraph with a blank left
    margin.</blockquote>

<p>Visit <a href="welkom.man">this .man page</a></p>
<p>Click <a href="script.praat" alt="bli|bla|blu">
    here</a> to draw a spectrogram.</p>
<p>Click <a href="sounds/i.wav">here</a> to hear a
    synthetic [i] sound.</p>

<address>
Charalampos KARYPIDIS
20130930
0
</address>

<h1>Lorem ipsum</h1>

<code width="6" height="3">
Create Sound from formula... sineWithNoise Mono 0.0
    1.0 44100 1/2*sin(2*pi*377*x)+randomGauss(0,0.1)
To Spectrogram... 0.005 5000 0.002 20 Gaussian
Paint... 0 0 0 0 100.0 yes 50.0 6.0 0.0 yes
plus Sound sineWithNoise
Remove
</code>

<p>This is a list with no leading bullets</p>
<ul class="noBullet">
<li>I</li>
<li>love</li>
<li>Praat</li>
```

```
</ul>
<p>Yet another list, with bullets this time.</p>
<ul>
<li>Praat</li>
<li>saves</li>
<li>time</li>
</ul>

</body>
</html>
```

7   Appendix B: example.man

Praat

```
ManPagesTextFile
"Example" "Charalampos KARYPIDIS" 20130930 0
<intro> "Lorem ipsum dolor sit amet, consectetur
    adipisicing elit, sed do eiusmod tempor incididunt
     ut labore et dolore magna aliqua."
<entry> "Lorem ipsum"
<normal> "Ut enim ##ad minim veniam#, %%quis nostrud%
     exercitation $$ullamco laboris$ nisi ut
     __aliquip_ ex ea^^commodo^ consequat."
<definition> "This is a paragraph with a blank left
    margin."
<normal> "Visit @@welkom|this .man page@"
<normal> "Click @@\SC""script.praat"" ""bli"" ""bla""
     ""blu"" |here@ to draw a spectrogram."
<normal> "Click @@\FIsounds/i.wav |here@ to hear a
    synthetic [i] sound."
<entry> "Lorem ipsum"
<script> 6 3 "
Create Sound from formula... sineWithNoise Mono 0.0
    1.0 44100 1/2*sin(2*pi*377*x)+randomGauss(0,0.1)
To Spectrogram... 0.005 5000 0.002 20 Gaussian
Paint... 0 0 0 0 100.0 yes 50.0 6.0 0.0 yes
plus Sound sineWithNoise
Remove
"
<normal> "This is a list with no leading bullets"
<list_item> "I"
<list_item> "love"
<list_item> "Praat"
<normal> "Yet another list, with bullets this time."
<list_item> "\bu Praat"
<list_item> "\bu saves"
<list_item> "\bu time"
```

8 Appendix C: HTML-Praat tag equivalence

Table 1: The "*" denotes the string enclosed in the corresponding HTML tag.

HTML tag	Praat element
\<a\>	@@*@
\<address\>	2nd line of .man file
\<b\>	##*#
\<blockquote\>	\<definition\>
\<cite\>	\<intro\>
\<code\>	\<script\>
\<h1\>	\<entry\>
\<i\>	%%*%
\<kbd\>	$$*$
\<p\>	\<normal\>
\<sub\>	__*_
\<sup\>	^^*^
\<ul li\>	\<list_item\>