

bLaTheR, a jQuery Plugin and Network Service

George Kelly Flanagin

with contributions from Scott E. Lewis

November 8, 2011

Contents

1 Why Bother with bLaTheR?	1
2 The script	2
3 Using bLaTheR	3
3.1 Basics	3
3.2 Parameters	3
3.3 Predefined source material	4
4 bLaTheR's Method of Operation	5
5 Troubleshooting bLaTheR	7

1 Why Bother with bLaTheR?

[Lorem Ipsum](#) has been used since the 1960s for placeholder text. It was intended to lessen the distraction of meaningful, or even *readable* text so that one could focus on the visual design elements of the printable material. For some purposes it is adequate. However, Latin-like text has statistical properties that are substantially different from English, Spanish, and all non-Latin languages.

Quite often, if we are studying how text flows, hyphenates, kerns, and justifies, we need text that is statistically identical to the text that will eventually populate the document we are trying to design. For example,

- ✧ English has a great many short words, and the seventy most common words in English constitute 40% of both the written and spoken language.

- ✧ Finnish has the property that words are often quite long because of a number of doubled letters.
- ✧ Many languages have punctuation characters that are endogenous.

bLaTheR is a service that creates pseudo-random text from a database of sources. To a sufficient degree of accuracy for the purpose intended, bLaTheR's text will have the same statistical properties as the text used as the source material.

Examples where bLaTheR is particularly useful:

- ✧ Building a web page to display corporate documents with a great many acronyms and long words that are "made up." It is not usually a good idea to reveal the company's secrets in a mock-up web page.
- ✧ Constructing a multi-lingual web site, and determining how the page will look in two different languages.
- ✧ Demonstrations of the effects of the choice of typeface, leading, and the like while showing the same random text twice on the page.
- ✧ bLaTheR is fun. Remember [Jabberwocky](#)?

2 The script

Acquisition of the script is simple. Put the following line somewhere on the page.

```
<script src="http://blather.georgeflanagin.com/blather.js"></script>
```

For best performance, **scripts go near the bottom of the page, but before the `</body>` tag**. The script will determine if the necessary parts of **jQuery**¹ are already present, and it will include them if they are not.

IMPORTANT: The bLaTheR script itself is less than one network packet in size, so loading it should be a quick operation.

¹"Necessary parts" in the case of bLaTheR are the jQuery core and the AJAX functions.

3 Using bLaTheR

3.1 Basics

The creation of the bLaTheR-ized text is consistent with other **jQuery** invocations. Most will look something like the following:

```
$("#selector").blather(jQuery-obj);
```

These invocations must go after the inclusion of the **script**, and after the declaration of the element that each call modifies/populates.

3.2 Parameters

At this time (November 8, 2011), the jQuery-obj has the following properties:

depth is an integer in the range of ($4 \leq \text{depth} \leq 8$) that controls the degree of resemblance of the bLaTheR generated text to the original. For most original text, 5 or 6 is a degree of resemblance that is sufficiently useful.

The default value is 5.

source is a string that tells blather what kind of text to use as its starting point for generating your nonsense text.

All **source** strings match the regular expression

```
^[a-z0-9\.\-]+[*]?$
```

Typically, this will be something like

```
en.bill.sonnets // matches sonnets of William Shakespeare
en.bill.*       // matches all works of William Shakespeare
en              // Some English
es              // Some Spanish
es.*            // All examples of Spanish in the database
loremipsum      // "Authentic" Lorem Ipsum
```

The default value is 'en'.

size is an integer number of bytes requested in the range of ($1 \leq \text{size} \leq 10000$).

IMPORTANT: Users should note that the text is UTF-8 encoded, thus for non-English languages the number of visible characters will be less than or equal to the number of bytes returned.

The default value is 2000.

When executed, the **jQuery** call to **blather** will return **size** characters of **source** and populate the container(s) referred to by **selector**.

A request like

```
$('p').blather({"source":"en", "size":250});
```

will replace each `<p>` element with *the same 250 characters* of pseudo-English text. However, if one has correct HTML, then a request such as:

```
$('#prod-desc').blather({"source":"en", "size":520});
```

Will populate only the `<div>` with the **id** of **#prod-desc** because there will only be one such `<div>`.

3.3 Predefined source material

The simplest use is to supply an ISO 639-2 language code such as **'en'**, **'es'**, **'fr'**, *etc.*, as the source. **'loremipsum'** is also available as a source, and it will return traditional Lorem Ipsum text.

The following is a list of “well-known” sources that should be in the database for any version of bLaTheR after November 8, 2011. The parenthetical expressions describe the key feature (if any) of the text.

419 Assorted Nigerian 419 Scam letters.

audio Reviews, filled with hyperbole, of audio products. (Relatively long sentences, with many adverbial phrases.)

bill.keller Emails sent by convicted felon Bill Keller pleading for money.

bush The *State of the Union* speech by President George W. Bush, 2006.

en.bill Text from the sonnets of William Shakespeare. (Early modern English poetry.)

latin.virgil.eclogues.4 *Eclougue IV* of Publius Vergilius Maro. (Real Latin instead of Lorem Ipsum.)

monster Skill descriptions from postings seen on Monster.com. (True blather.)

obama The *State of the Union* speech by President Barack Obama, 2010.

rowling A little text from J. K. Rowling's book *Harry Potter and the Order of the Phoenix*. (Rich in punctuation and dialog.)

yahoo Yahoo!, Inc.'s message about the state of the company. (Is there anything more bLaTheR-able than corporate emails?)

4 bLaTheR's Method of Operation

1. The jQuery plugin invokes a script on the server.
2. The script opens a socket connection to a [RESTful server](#) that routes messages to bLaTheR, and provides process handling and threading.
3. bLaTheR uses the **source** argument as a partial key to retrieve the matching n-distributed [PATRICIA \(or radix\) trie](#) from the database.
4. If the particular (source+depth)-tuple has not been previously requested, bLaTheR loads the matching text from an [SQLite database](#). The text is analyzed, and the resulting PATRICIA trie is written to the database, and the text-load routine is tried again.
5. The **depth** parameter of the request ensures that the result is [n-distributed](#), meaning that the distribution of **n-graphs**² in the bLaTheR generated text will be the same as the distribution of the **n-graphs** in the source material.
6. A starting point is chosen to be a random node of the trie that starts with **[A-Z]** and is followed by either a space or **[a-z]**. This is solely an æsthetic choice to make the text appear more natural.
7. The generated text is built by traversing the PATRICIA trie using the last **n-1** characters of the most recent node as the index to find the next node. A choice of the next character is based on the frequency of that **n-tuple** in the text.

Example:

Suppose we are working with a **depth** of 5, and the current node of the trie is **est_e** (where the written underscore is actually a space). The last 4 characters are **st_e**. We find the **st_e** node in the trie, and then choose the 5th character based on the frequencies associated with these **5-tuples** in the text.

²"**n-graph**" meaning a sequence of **n** characters. For example, "**Geo**" and "**e F**" are both **tri-graphs** in the shred **George Flanagin**.

In other words, we are examining the text for words that end in “st” followed by words that start with “e?”, where the ? represents all the possibilities that were found in the text.

8. When the **size** number of bytes is reached, traversal stops.
9. If the last character is already a period, a question mark, or an exclamation point, nothing further is done. If it is not, then the last space character is replaced with a period, and the characters after that are replaced with spaces.
10. bLaTheR does check to make sure that the generated text has some spaces, and it ensures that the text is never reduced to less than 90% of the requested size.
11. The result is packaged in a JSON representation and returned to the calling script.
12. The jQuery plugin has supplied “**json**” as the type of the reply, and extracts the appropriate value from the result.

5 Troubleshooting bLaTheR

Is the service available? The service is available on one of two CNAMEs. Generally speaking, they will each resolve to the same IP address.

The IANA has given port **33333** to Digital Gaslight, Inc., and this is the port used for the bLaTheR service. You may do these things to clarify the availability:

1. Attempt to **ping** the server with the following command:

```
ping blather.georgeflanagin.com
ping blather.technify.me
```

If **ping** responds, the server is plugged in and the network card has power.

2. See if you can get to the server's port.

```
traceroute -T -p 33333 blather.georgeflanagin.com
traceroute -T -p 33333 blather.technify.me
```

In this particular use of **traceroute**, the **-T** specifies a TCP connection overriding the default of ICMP, and **-p** specifies the port number that we require to be open.

traceroute is a powerful tool and will identify problems such as port **33333**'s closure somewhere between you and the service, or **routes** to nowhere.

Does the source exist? The most common problem in the request itself is that the source may be spelt incorrectly. A list of publicly available values is shown above in Section 3.3.

Not getting back what you were expecting? This is likely due to a misspelling of one of the other parameters, and a defaulted parameter being used in its stead.

Still having trouble? A useful diagnostic message can be placed in an execution of **telnet**:

```
telnet berlioz.digitalgaslight.com 33333
{ context="Ishtar"; client="Diagnostic"; method="echo"; }$$$
```

A reply of approximately 3200 bytes will be written to your terminal. Copy the reply to an email, and send it to: **me@georgeflanagin.com** putting **blather diagnostic** in the subject line.



George Kelly Flanagin

1901 Woodbine Road Richmond, VA 23225-3227 USA

Phone: +1.804.399.2699 from 9am to 5pm, US Eastern Time

Email: me@georgeflanagin.com **Skype:** georgeflanagin

Key: pgp.mit.edu EDF61649