

Editomat 2: Continuing Adventures in Commercializing a Tcl Application

Clif Flynt
Noumena Corporation,
8888 Black Pine Ln,
Whitmore Lake, MI 48189,

<http://www.noucorp.com>
clif at noucorp dot com

September 25, 2017

Abstract

The Editomat product was first prototyped in 2014 and released to the public in 2016. A previous paper (Tcl Conference 2016) discussed the evolution of the prototype to a product. This paper describes the evolution from a product to a business. It discusses developing vs purchasing services, Search Engine Optimization, product promotion, and tracking results.

1 Introduction:

Creating a functional prototype is easy. You can use ugly buttons and not worry about font sizes, multi-platform support, unexpected user behavior, regression testing or many other issues.

To convert a working prototype (or most in-house projects) into a shrink-wrap product is harder. Now you *must* consider the user experience, functioning the same on all platforms, maintaining bug compatibility between releases, handling unexpected (and malicious) input, protecting your intellectual property and more.

In my experience, it takes roughly twenty times the effort to create a product than to create the working prototype.

Releasing a product does not create a business. To be a viable business, you need to *sell* your product. This moves into a new domain of marketing, advertising, websites, Search Engine Optimization, user interaction instructional videos. All this and much, much more.

Some of these activities lead to sales, some lead to wasted effort. Given the limitations of time and space, a business needs to spend time on productive activities and not waste time on others.

2 Define your Product:

Step one is to produce a product that people actually want badly enough to pay for.

The best applications are designed by people who understand the problem domain.

Editomat grew from my own needs and desires. In that respect, it grew from a survey of the user community. While it is easy to calculate mean, median and mode for a sample population of one, this might not represent the target population of customers.

It didn't.

Within the first six months of release, it became obvious that editing with the Tk text widget's Emacs bindings worked for me, but not for any other writer in the known universe.

The `twapi` (twapi.magicplat.com/) and `cawt` (www.posoftware.de/html/extCawt.html) extensions made it relatively painless to make Editomat run as an MS Word add on.

Extending this support to LibreOffice is a surprisingly tough problem. LibreOffice support has not been implemented.

For all Microsoft's faults, their documentation is better than most open source projects and there are more examples of using the API.

Twapi solved the problem of working with Word on Windows, but does not solve the problem of providing a Word interface for the Mac.

The web logs show that nearly half of the downloads are for Macs. This is a percentage of the target population that can't be ignored.

To further complicate matters, Word on the Mac cannot save a document in a text format that Tcl can easily read. Either Unicode characters are garbled, or paragraph information is lost.

RTF (Rich Text Format (or Really Truly F-d)) is a Microsoft format for document exchange. To its credit, it works. Less to its credit, the specification is a mass of special cases and exceptions. It is not as trivial to parse as HTML.

However Microsoft Word, LibreOffice, Scrivener and other text processing packages (even on a Mac) can export a document that conforms to the RTF standard. The popular writing tool Scrivener uses RTF for its internal format.

This analysis of needs led to an intense effort to read and write RTF files. My first approach was to use the GNU `rtflib` package. This is a LGPL 2 licensed package, so it can be used commercially without needing to release all your code.

However the package is C++ based, requires callbacks to handle display, and just looked messy.

RTF is a plain text based format, so it should be simple to parse it using plain Tcl.

This is only true for insignificant values of simple. RTF uses backslashes and braces in ways that Tcl does not. A bare-bones RTF reader required about five hundred lines of Tcl code. The reader is so small because Fiction uses minimal formatting. Much of the RTF specification is ignored.

I looked at RatFink for generating RTF output, and discarded it as more heavyweight than I needed, again because fiction has so few requirements. All Editomat needs is paragraph markers until I add support for italics and bold to Editomat's editing functionality.

The RTF writing procedure dumps a hardcoded prolog to declare the file to be basic RTF, and then a sequence of "new paragraph" commands and the text of a paragraph.

As of writing this paper, the RTF handling code has been released to the Bleeding Edge users.

3 Follow the bouncing ball:

Even with a solid product definition, there will be surprises as the world under your product slides around.

Almost a year after Editomat's installation program was tested and released, the Mac OS/X Sierra release added a new security feature—running a downloaded application in a safe, read-only sandbox.

According to the web, this broke many installation programs. It also broke Editomat's registration process that inserts the registration key into the executable.

This was finally reported by a user. Once known, it was relatively easy to modify the registration proc to handle the Sierra special case.

An artistic friend reworked the square washing-machine buttons I designed for Editomat into pretty circular buttons. It was fairly easy to create "round" buttons on PC and Linux platforms by setting the button's borderwidth to 0.

The Mac port wasn't so simple. The Macintosh world wants every application to look like a Mac App, which may not be the way I want it to look.

In particular, Tk's Mac buttons have square (or rounded) borders.

In order to use circular buttons with no square outline, I borrowed (and reworked) Steve Landers' `gbutton` (Tcl Conference 2003) concept of creating custom buttons by adding bindings to a Canvas widget.

A welcome side effect of the home-made buttons is that it made it easier for `tktest` to perform the regression tests.

4 In-House vs Out-House:

A business should define the expertise it offers. Anything outside that area of expertise should be purchased, not built.

This is a tough decision for a systems engineer/software developer. Building websites, optimizing search engines and creating instructional videos all use available skills and expertise.

The issue is that In-House solutions take time that should be devoted to product development, and an external expert probably does a better job than an in-house person with barely adequate skills.

On the other hand

- Hiring an Out-House expert requires fully defining the requirements, instead of thrashing around solutions until you understand the problem.
- Out-House solutions might not match what you need.
- It may take longer to customize a *standard* solution than it takes to write a custom application from scratch.
- In-House time is “free” while Out-House solutions cost money up front.
- It often requires an expert to distinguish a capable service provider from a novice claiming expertise.
- It can also require more time and effort to beat results out of an overscheduled expert than it would to do the job In-House.

Editomat uses (too) many In-House solutions. The website design gets poor reviews, and I expect the YouTube instructional videos will also get panned. When there’s a revenue flow, I intend to upgrade both.

5 Getting Stuck in the Web:

People go to the web to find products. Your product needs a website.

The more professional looking the website, the better. People *will* judge a book by its cover.

For minimal effort, you can get a mostly-pre-built website from hosting services like GoDaddy, HostGator, DreamHost or others. These services usually come with a domain name registration and may even include security certificates.

These sites commonly include a template for building the pages and may have support for shopping carts, etc.

If you want more control and are able to do your own web-server maintenance, you can rent a host from companies like Linode, Rackspace and others. The provider will maintain the hardware and may update the operating system, but the bulk of the maintenance will be your problem.

These services don’t commonly include a domain name registration. Therefore, you’ll need to register your domain name with a registrar like Network Solutions, GoDaddy, Dyn or Gandi. You should also purchase a security certificate from Comodo, Network Solutions, DigiCert, etc.

For tests, (or very shoestring budgets) Let’s Encrypt (<https://letsencrypt.org/>) offers free certificates.

Finally, you can buy your own hardware, register your domain, rent a fixed IP address, and build your site from scratch.

Because Noumena already had hardware, network connectivity, web and email server configured, it was simplest to piggyback Editomat onto the Noumena hardware.

The website is served with `tc1httpd` using the Tcl-On-Track package described at the 2014 Tcl Conference.

6 Face(book) the Music:

Like it or not, a Facebook business page is as critical to being a legitimate business as having a telephone.

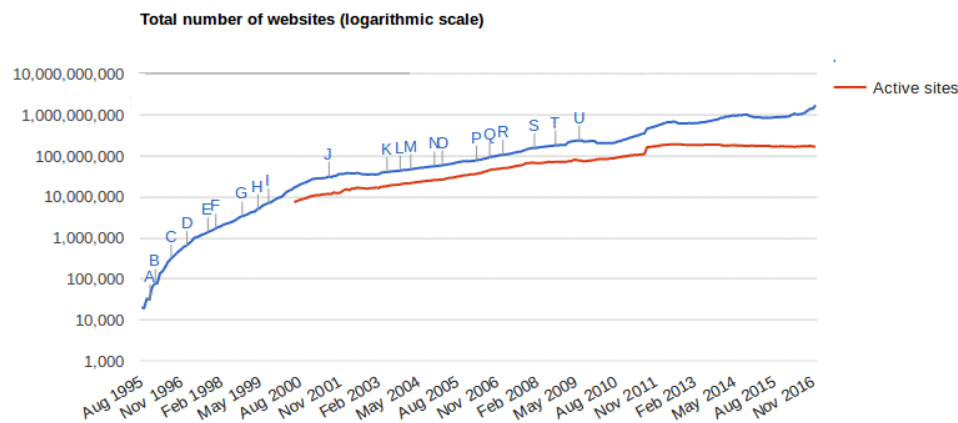
Fortunately, it's cheaper and easier to set up a Facebook page than convincing Ma Bell to hook up a land-line telephone.

- Visit `www.facebook.com` - login as yourself
- Browse to `www.facebook.com/pages/create`
- Fill in page info. Perhaps create photos and figure out text before you start. Facebook is relatively forgiving about image formats, but not informative about what formats they accept.
- Reference the page as `@pageName` when you post about it.
- Anyone can do this, legitimate business or not.

7 SEO is a Four Letter Word:

In the mid 1990s, there were about 100,000 websites. It was easy to be ranked in the top dozen and have your site show up on the first page of a Google search.

As the next graph shows, that's changed. Note that the Y axis is log scale.



With over 100,000,000 active sites, the search engine's task of finding the best sites to report has become more than just 1,000 times harder.

As a data point, in the early 2000's Noumena Corporation was the second site listed when you searched for Tcl Training. After neglecting the search engines, it dropped to the fifth results page (and website hits approached zero). After reworking the web pages and registering with Bing, Google Business, Yext, and Yelp, Noumena has moved to the second page of hits. Not good, but better.

The two most important pieces of Search Engine Optimization are

- Your website
 - Important information comes "above the fold"
 - Make it easy for search engines
 - * Use keywords - more specific is better.
 - * Use structured data.

- * Use descriptive page names.
 - * Use Dash and Underbar carefully in page names.
 - * Include sitemap, robots.txt, favicon.ico,
- Websites that point to your web pages.
 - Register with search engines: bing, yelp, yext, etc.
 - Ask legitimate sites to reference your page. Perhaps join BBB or Chamber of Commerce.
 - Find bloggers to review your product and page.
 - Get references/endorsements.
 - Don't pay for links.

Some of these points are obvious: Put important information first. Of course you want the most important information to be the first words people see.

Others are less obvious: Page names with underscores are not treated the same as page names with hyphens. Google translates `Tcl-Training` as two words: *Tcl Training*, while it parses `Tcl.Training` into one word: *TclTraining*.

Search Engine Optimization has become so complex that it's a career specialization for many IT professionals.

That said, an amateur can play the game. These sites explain the basics.

moz.com/beginners-guide-to-seo/growing-popularity-and-links
www.google.com/webmasters/docs/search-engine-optimization-starter-guide.pdf
support.google.com/webmasters/answer/35769
www.woorank.com/
tools.neilpatel.com/
www.rankingcoach.com/
<http://www.ecreativeim.com/blog/category/seo-basics/>

8 You must be registered:

Less than half of the small businesses in the US have a website. While many businesses are just now constructing their first website, there is a growing industry in creating fake business websites to sunder the unwary customer and their money.

The search engines, Google, Bing, Yelp, etc. put serious effort into separating the legitimate wheat from the scam chaff.

One technique they use is to see what sites reference your pages. The more quality back-links, the stronger your business. The keyword is *quality*. Back-links from `scams-r-us` will get you a negative score.


This is important to the Tcl community because it's so easy to get names in the `.tk` domain that back-links from a `.tk` site are considered weak. We may want to consider a `wiki.tcl.org` mirror to the `wiki.tcl.tk` site.

The search engines give preference to sites that have registered with them. As part of the registration, they validate each business and prove it is legitimate.

Here are the steps to register with the most important sites.

- Claim a Google business page.
 - Visit <https://business.google.com/manage/#/list>
 - Fill in the form as shown below.


Business name
Editomat


Country / Region
United States 


Street address
8888 Black Pine Ln Suite 3

[add line](#)

City
Whitmore Lake

State Michigan  **ZIP code** 48189

Main business phone
 734-449-8779

Category
Software Company 

Website
www.editomat.com

I deliver goods and services to my customers at their location
—[Important information](#)

Yes No

[Continue](#)

Figuring out the correct Category for your business is much like playing the old Advent games. It won't give you a list of known words, it just makes suggestions after you type in a few letters. There are no categories for Writing Aids, Line Editing Tools, Author tools, etc. The closest I found was the generic Software Company.

Note that when Google finds two businesses at the same address, it assumes the first business has closed and a new business has opened.

The workaround is to tell Google that your studio apartment is composed of multiple suites. This will convince them that the new business has a new address and has not replaced a defunct business.

Google will validate your business by sending a postcard with a special code to your official address.

How would you like to get your verification code? [Learn more](#)



By mail at

8888 Black Pine Ln Suite 3
Whitmore Lake, MI 48189
United States

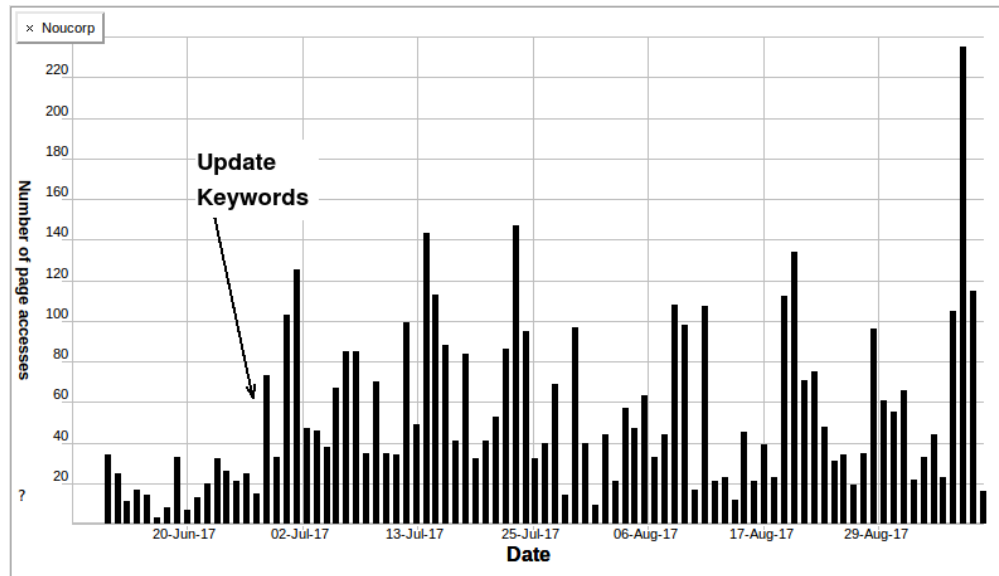
Mail

[Verify later](#)

- Claim your Yahoo (Yext) business listing
 - Visit www.yext.com/pl/yahoo-claims/index.html
 - Yext will require a bank statement, utility bill, or something to verify that the business is legit.
- Claim a yelp business page
 - visit <https://biz.yelp.com/>
 - Yelp ask questions like DUNS and FEIN numbers
- Claim a bing business page
 - visit <https://www.bingplaces.com/>
 - Validates your identity with a special HTML meta-tag.
- Also set up g+, Linked In, Twitter, Pinterest, etc
 - As defined on their pages. These are lower-value back-links.

9 It's in the details:

Elements as simple keyword choice can affect your site's rank. This graph shows how custom, targetted keywords for each page instead of using a single set of generic keywords for all pages on a site affect page hits.



The tclhttpd engine generates pages on the fly, as well as displaying static pages.

The Tcl-On-Track (Tcl Conference 2014) tools for creating tclhttpd websites makes heavy use of dynamic pages. The toolkit defines the meta information for all pages in a single file. Putting all the meta information in one location makes it easy to view and compare each page's meta-data.

```
array set meta {
/ {
  title {Training, Mentoring and Custom Software Development}
  msvalidate.01 ``45D1AB698F9179418CE78CD0EE1642E2''
  keywords {Tcl/Tk Training, Linux Training, Tcl, Tk, Consulting}
  description {Clif Flynt, Tcl/Tk Training, Linux Training}
  author {Clif Flynt}
  robots {index, follow}
  itemprop.name {Noumena Corporation, Linux & Tcl}
  itemprop.description {Clif Flynt, Tcl/Tk Training}
  itemprop.image {imgs/N6N-box3-4.gif}
}
Tcl {
  title {Tcl Support from Noumena Corporation}
  keywords {Tcl/Tk, Programming, resources, script, qa, tcl, tk}
  description {Tcl/Tk Tools, Scripts, Papers}
  author {Clif Flynt}
  robots {index, follow}
  itemprop.name {Tcl Support from Noumena Corporation}
  itemprop.description {Clif Flynt, Tcl/Tk}
  itemprop.image {imgs/N6N-box3-4.gif}
}
Training {
  title {Onsite Training from Noumena Corporation}
  keywords {Tcl/Tk Training, Linux Training, SQL, UML}
  description {Tcl/Tk Training, Linux Training}
  author {Clif Flynt}
  robots {index, follow}
}
```

```
itemprop.name {Noumena Corporation, Linux & Tcl}
itemprop.description {Clif Flynt}
itemprop.image {imgs/N6N-box3-4.gif}
}...
```

Tcl-On-Track uses the most specific meta array definition for a URL. Thus `www.noucorp.com/Training/Tcl` would receive the *Training* tags while `www.noucorp.com/Tcl/Whitepapers` gets the *Tcl* meta tags.

10 Promote Once, Measure Twice:

Nobody will buy a product they've never heard of. There are numerous ways to promote a product to the prospective purchasers.

- Paid advertisements
- Website promotion
- Public speaking events
- Online forums
- Blog entries
- Facebook postings

Some of these techniques will work for one product, and some will work for another. If you can't measure the effectiveness of a technique, you won't know what to do more of and what's wasted effort.

During the past year, Editomat has been

- presented at five events attended by writers.
- mentioned in an online forum.
- described in a writing blog.
- discussed at writing groups.
- added to Facebook
- registered with search engines.

These activities have pushed Editomat purchases into double digits—barely enough to count as a beta-tester population.

To track the effectiveness of the various techniques for acquiring customers, you need to measure and annotate when events occur.

The two pieces of critical information are website visits and sales.

The quantity of website visits defines how well optimized your search engine placements, keywords and other advertising are working.

Sales tells you whether people visiting the website are buying your product, or taking one look and running away as fast as possible.

Most web servers maintain a log in a "standard" format. The default location for `tclhttpd` log files is `/tmp`. This directory is likely to be cleaned when you reboot. This can (and should) be redefined in `bin/tclhttpd.rc`.

Online payment system like Paypal or Square, provide downloadable reports in a different "standard" format.

As has been frequently noted, the best thing about standards is that there are so many of them. Every web server uses a slightly different log file format and Paypal changes their report format almost annually.

There are pre-built applications to read, analyze and generate reports from log files and spreadsheets. Unfortunately, they fall into two categories: those that are application specific (and limited), and those that can be customized with a proprietary scripting language.

The Editomat website uses Tclhttpd. The log files aren't exactly the same as Apache or nginx, which rules out using any analysis package that does not support customization.

Rather than use an analysis package that needs to be customized with a proprietary scripting language, it's simpler to use a robust, well documented language like Tcl.

The tclhttpd supports virtual domains, thus a single server hosts `www.editomat.com`, `www.noucorp.com` and `www.cflynt.com`.

The default tclhttpd log does not specify which virtual server has returned a page.

Changing the log definition in `tclhttpd.SITENAME.rc` causes tclhttpd to save each virtual host information in a separate log file. This line puts the logs into a folder which won't be cleaned upon reboot and makes it obvious which site's activity is being logged.

```
Config LogFile /var/tclhttpdLogs/Editomat_log
```

Over the years I've developed several quick and dirty tools based on shell commands like `grep` and `wc`. These proved inadequate for the level of analysis I needed for this project.

Tcl is the obvious choice for extracting and collecting data from the log files. Despite the aberrant text inserted by malicious clients (JavaScript exploits, database exploits, application specific exploits), it's relatively easy to extract information from the log files.

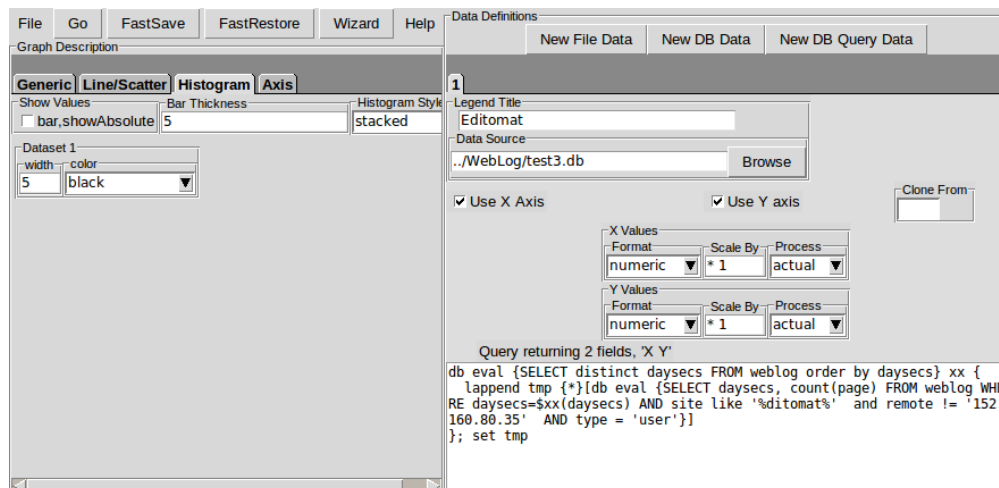
It was soon obvious that writing a custom data collection procedure for every type of analysis I wanted to do was not a long term solution.

I wrote a Tcl script to extract the data from the Tclhttpd log file and populate an SQLite database. This provided an opportunity to pre-process the data, converting the time/date string to seconds, creating a single "seconds" entry for each day, and applying heuristics to assign an originator tag (user, robot, vulnerability probe) to each record.

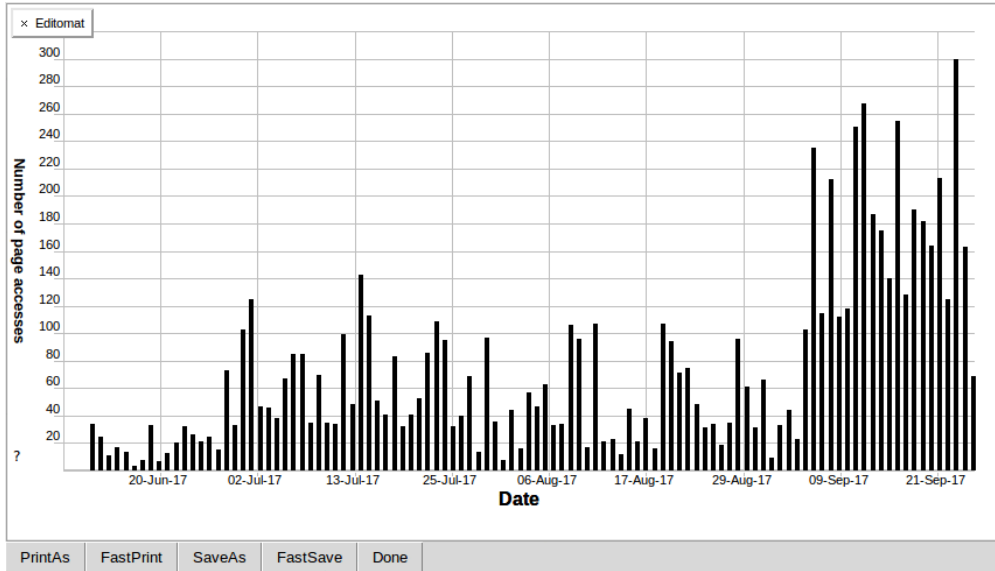
Once the data is in an SQLite database, it's easy to do analysis with a general purpose data mining tool. The tool I used grew out of my graphing library. It is available at

<http://www.cwflynt.com>

The application allows data to be read from flat ASCII files, similar to using `grep` and `cut`, as single value per row data, or as a complex looping query as shown here where each day's activity is selected and summarized.



This analysis produces output resembling this.



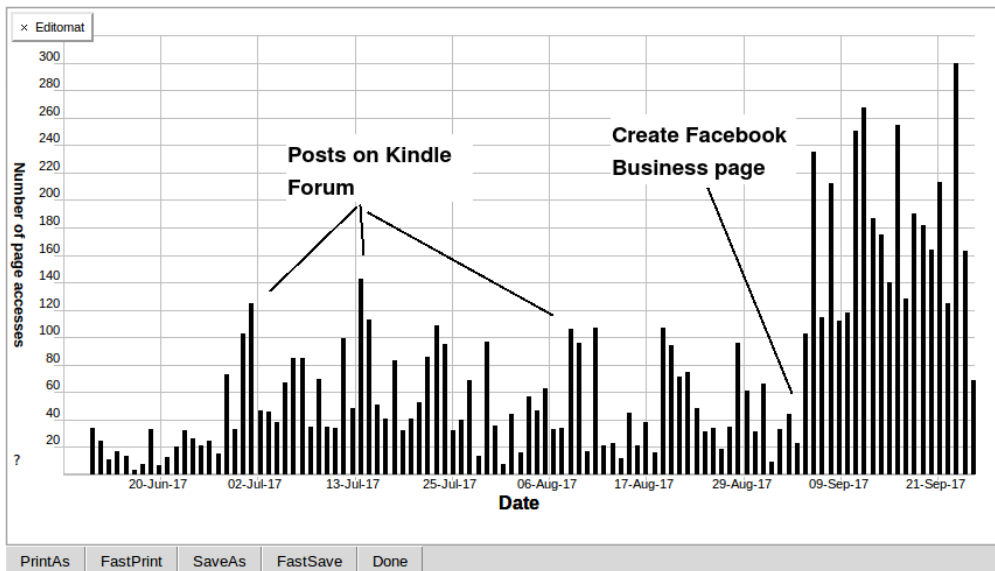
Which leads to the question of what caused the spikes on July 9 and August 22 and what was the big change on September 4.

Along with the automated logs generated by web servers and payment systems. I keep an log of my actions so I can check when events occurred and what the result was.

My personal activity log shows that the August peaks happened after friends mentioned Editomat on a writing oriented bulletin board.

The September surge followed creating a Facebook Business page.

To date, discussions on the kindle forums have been the most effective marketing tool.



11 Lights! Camera! Action!:

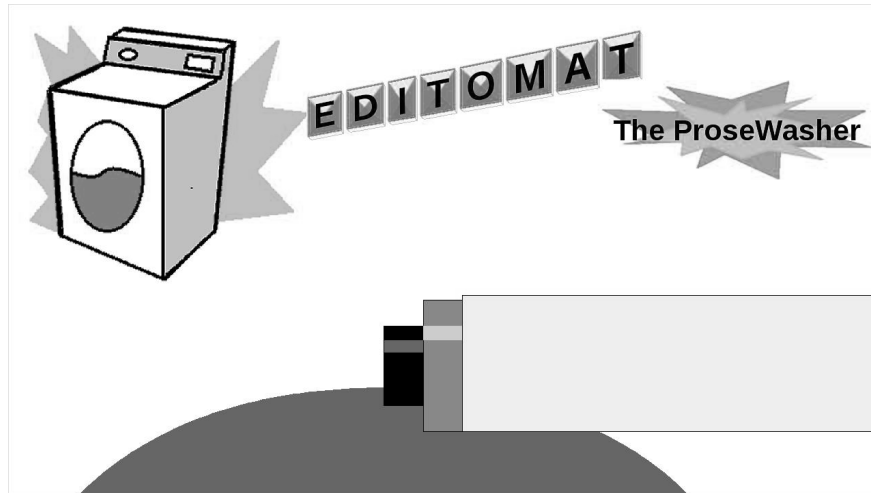
Along with a website, it helps to have instructional videos on Youtube.

I created my videos with KDenlive on an Ubuntu Media distribution.

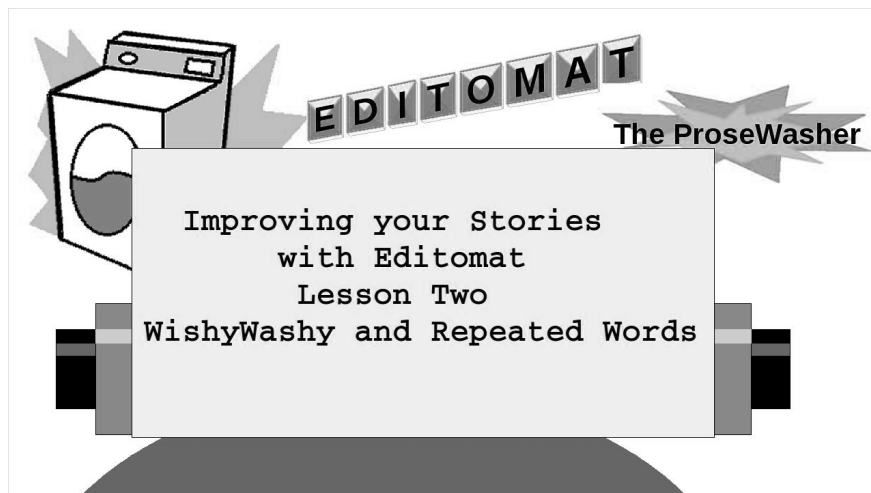
I recorded the audio with Audacity on a Mac, and tweaked the tracks with a Tcl script which invokes sox to reduce noise, normalize volume, slow my delivery and drop my voice a half-step.

KDenlive can create opening credits, but will not create animated credits.

The Tk canvas is a great drawing surface. It can be extended with the Img package to convert the canvas elements to a Tk image, and those images can be written to disk files. With these tools, it's relatively simple to create a set of animation cells that progress letter-by-letter from this screen



To this one



The image files are named with a pattern `lsort` can understand. The final step is to assemble the images into a video stream for KDenlive to merge into the video.

```
if {$argv eq ''} {
```

```

puts ``tclsh makeFlick.tcl 'pattern''
exit
}
set pattern [lindex $argv end]

# output file named by current timestamp
set id [string range $pattern 0 4],[clock format \
[clock seconds] -format ``%m%d'']

# Collect image files
set files [lsort -dict [glob $pattern]]
set of [open files w]
puts $of [join $files \n]

close $of

# Assemble the MPG clip.
exec mencoder mf://@files -mf fps=30 -ovc lavc \
-lavcopts vcodec=msmpeg4v2 -noskip -o $id.mpg

```

12 Conclusion:

This year has seen a few new features added to Editomat, notably support for dialog analysis, importing and exporting RTF files, and covering more edge cases. More effort has gone into extending the regression test suite, enhancing the website and reaching out to the user community.

Tcl's ability to process text, interact with databases, and do general programming tasks makes it a useful tool for dealing with many aspects of creating a business.

The canvas is a fine tool for simple animations.