
Analytics: moving from cost to value

The times when technical documentation teams needed to justify their worth are well and truly over. Or are they? Here is how TC can deliver information that could be essential to the entire enterprise.

Text by Fabrice Lacroix

Image: © TommL/istockphoto.com



Your tech doc team has written another manual, printed it out, put a nice cover on it, bound it, and placed it on your desk. Congratulations – this proves that they have done their job! You’ve paid for something, and there it is. In other words, this beautifully composed manual on your desk is evidence of one thing: money spent. Presumably, it has value, or you never would have asked for it. But how do you measure this value? Dr. Deming famously said, “It is wrong to suppose that if you can’t measure it, you can’t manage it – a costly myth” (<https://deming.org>). But if you have a choice, being able to measure is better. Indeed, we’d rather trust Lord Kelvin when he says: “If you cannot measure it, you cannot improve it.”

Without measuring the value of technical documentation, all management knows is how much it costs (all those paychecks and supplies), but the value is taken on faith. It must be worth something, right? Or we wouldn’t do it. Let’s unpack this, starting with this undeniable statement:

“How you measure depends on what you’re measuring.”

Physical books

If all you do is print physical books, you know how many books are sitting in crates. The number will be similar to how many customers you have and how many products you have sold. But, as you presumably already knew these numbers, not a lot of value has been added.

There are a lot of good things to be said about books: The interface is intuitive, they sit on our shelves reminding us of the good old days, and they smell nice. It’s sad to see them go. But they’re expensive to print... all those dead trees.

PDFs

With PDFs, we start dematerializing books. We convert our manuals into digital files, and post them on a website. This is a good move – you have eliminated the cost of printing and shipping... or, at least, transferred it to the customer, at his option. The customer can always read on the computer (unless he’s using the manual to repair the computer, in which case, the paper, toner, time, trouble, and staples are on him).

You can count how many PDFs are downloaded, but the rest remains unknown. Have they been read at all? Have they been shared on a corpo-

rate server with dozens of people? Which parts have been read? All of that is unknowable. All you get from download numbers is a figure that lets the manager feel that they know what’s going on and that the cost of producing the PDF and its content has not been wasted.

This is just not enough data, and not the right data. So, let us move on to...

HTML web pages

This is quite an improvement on PDFs. First, you know how many times a page has been opened, at what time and for how long until it is clicked away from. If this time is long enough, it has probably been read. And, because the manual probably doesn’t fit on a single web page, you have a somewhat finer grain of knowledge. Instead of just being aware that a PDF has been downloaded, you find out that the HTML page containing Chapter 5 of the manual has been loaded in someone’s browser. This is better! You can tell management that Chapter 5 has received some attention.

If you could tell where on the page the customer was looking, that would be even more informative. Sadly, you (probably) don’t own the browser, so you can’t get that kind of detailed feedback. If you broke the chapter into too many small web pages, you could learn more precisely what parts of the chapter have been read, but it would be a pain for your users: They would spend their time clicking Next and Back buttons to navigate your content. So you get to pick: good user experience or fine-grained metrics.

But what metrics are we talking about? Web pages have been around for a while, and people believe that some useful analytics have been collected using tools such as – most prominently – Google Analytics. It’s right there in the name! Not to mention that trusted brand. But having a lot of data doesn’t necessarily mean knowing something useful. We could call this “the data delusion” and here’s why:

Web analytics products were designed for marketing. It’s right there in the summary of their site: “Google Analytics lets you measure your advertising ROI as well as track your Flash,

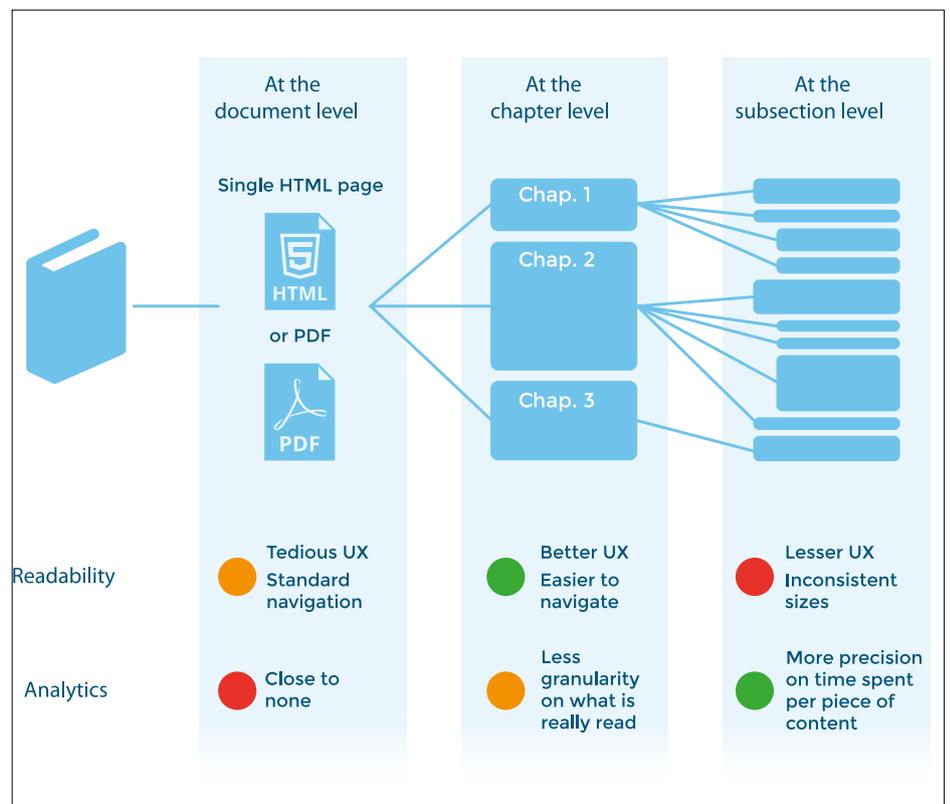


Figure 1: Long web pages provide lots of info on a ‘paper like’ experience, but less details. Smaller pages provide more fine-grained metrics, but a weaker UX. Source: Fluid Topics



Your fast track to success in any language

- ✓ Translation and localisation services in all languages
- ✓ Use of specialised translation memory systems and localisation tools
- ✓ Document, software and website localisation
- ✓ For almost 50 years experience in markets worldwide



www.e-kern.com
info@e-kern.com



KERN Global Language Services
 Kurfuerstenstrasse 1 · 60486 Frankfurt / Main
kern.frankfurt@e-kern.com · Fax: 74 99 98

+49 (0) 69 75 60 73-0

International:

Amsterdam · Berlin · Cologne · Eindhoven · Frankfurt
 Graz · Hamburg · Hong Kong · Linz · London · Lyon
 Marseille · Munich · New York · Paris · Rotterdam
 Salzburg · San Francisco · Utrecht · Vienna · Warsaw

video, and social networking sites and applications." Very useful for a company website of 20 to 200 pages. You get a report with the URLs, the number of times they have been loaded, how users got there, where they came from, and where they clicked to next. If you have a small number of pages, with a lot of viewers for each, and you want to focus your marketing dollars, this is very useful.

But how many tech doc web pages do you have? Hundreds? Thousands? Given a range of products, each with a set of manuals, in various languages and versions, the number of HTML pages could explode. Yet, many pages may only have a small number of readers. Your Google Analytics report will be a spreadsheet with thousands of items with impossible-to-decipher names with a wide range of numbers next to them. How can you make sense of these numbers without seeing the content at the same time? How can you make sense of a flow diagram showing the path users follow when thousands of small elements must be displayed? In addition, if two pages are related to the same task for the same product, but in two different versions, we need to see the metrics either separately or aggregated; but as Web analytic tools don't understand semantics, it's impossible to automatically create and navigate clustered numbers (by book, by version, by product).

Not to mention that changing the granularity of a HTML page (such as from topic to section) would render any comparison impossible.

What kind of action can you take with web analytics information? Not much. Metrics need context to be interpreted; we need to see the numbers and the content together in order to gain insight. Also, web analytics may tell you what has been looked at, but there's a vital bit of info they don't give you – what has not been read. Knowing what's happening at the web server level is not very useful. This renders marketing-style web analytics useless for technical communication.

So now we know that counting books isn't useful, counting downloads doesn't teach you anything, and that counting page loads is slightly more helpful, but not by much. We need to know what's happening at the user level. We need a different way to track and log.

What we need

If you really want data on how valuable your technical documentation is, you'd practically need to

be standing over the user's shoulder, observing what they're reading and for how long. But that's impossible, right? Well, maybe not.

We live in the era of the data-driven enterprise, where we can engage the flood of information produced by users. Social media, online retail, mass surveillance, census – all produce torrents of data. Algorithms are being developed that can process this data to enable insight. If we could make our technical documentation generate data about its use, and apply these advanced big data mining technologies, we could start to truly understand how our users are reading our documents, learn more about our tech doc and, most importantly, about how our users interface with our products.

If only a single topic in a manual could send a message to us that says: "Hey, I'm being read now!" For this to happen, we need two things.

First, our content needs to be granular, so that we can track the consumption of each fragment individually. Structured documentation (such as DITA) helps with this, but apart from that, it is the job of the dynamic delivery platform to break the content into chunks, whatever its source format. It must take differently formatted inputs and make them all consistent and uniform so that you can present the documentation any way you like. Because everything that could be considered documentation – wikis, knowledge bases, user forums, trouble tickets, and catalogs, as well as good old manuals and guides – should be searchable through the same portal. It's all product information, and it all contributes to customer success. So it all should be tracked and analyzed with the same tool.

Second, we need a custom-designed reading and tracking technology (see Figure 2). Because you can't control the browser (you don't own it), have your customers view the documentation through a web-based reader that you do control. This sends these messages home when it displays a fragment for a long enough time that it could have been read. Tracking the display at the device level, instead of the download at the server level, creates more relevant data.

This way, the benefits of a dynamic delivery portal go in both directions. Your customers read your material benefiting from a contextual and tailored documentation, and you read about their access. You get detailed, unbiased data on everything they see, click on, and read, generated on the fly.

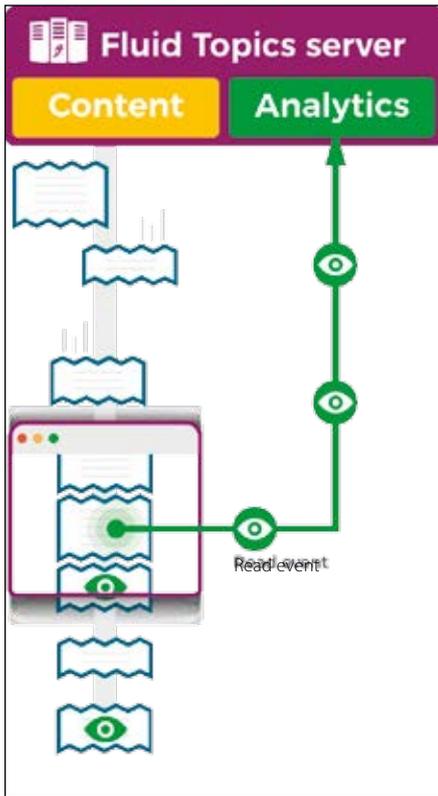


Figure 2: Displaying topics individually allows to track what is really being read and when.

Source: Fluid Topics

By accessing your documentation, your customers are telling you what's on their mind. It's time you listened.

Reading the numbers

Data visualization

Having done the above, your tech doc team can get a much better grip on which bits of documentation are getting attention and which are not. Using intelligent analysis algorithms and advanced visualization tools, you can get beyond just a number attached to each document.

You will be able to see the entirety of the documentation using graphical tools (rather than a list of numbers and indecipherable long URLs) and see where readers are focused and what grabs their attention. You can do that while seeing the content at the same time, which provides context.

You can aggregate these numbers to get more meaning. You may have an installation guide for each version of your software, but wouldn't it be more interesting to find out how much trouble us-

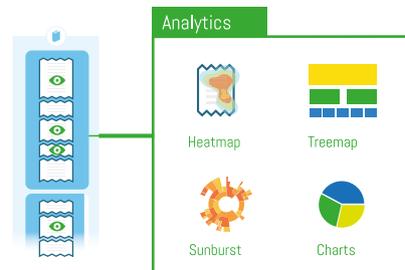


Figure 3: Using diagrams such as heat maps, sunbursts, tree maps, and others, you can gain immediate insight into what is driving your users.

Source: Fluid Topics

ers are having with installation as a whole, regardless of version? Metadata are not only useful for users at search time, but they can also be used to create clusters and axes for exploring the numbers. See content consumption per product line, type of content, etc. The possibilities are endless, and creating aggregates dynamically helps you understand and reveal the hidden data (see Figure 3).

Taking action

Maybe what the user needs to figure out is how to set the clock on the stove. If your technical writers look at the data and see that this is a popular topic, they could shoot an email over to customer support, and give them a heads-up that this is bothering users. Or they could forward reports to product design and let them know that their interface is confusing – and maybe they can do something about that in future models. Or they could contact sales and suggest that when they sell an item, they chat about how to set the clock.

The possibilities are suddenly blossoming because you have discovered the data hidden in your documentation consumption.

Your technical writers have gone from being a known cost with an unknown value to being a known cost with a known value. Your data-driven organization now has a data-driven tech doc group.

Possibilities

With data flowing out of your dynamic delivery portal, you can do many more things than you could before.

Once you are able to collect data on every topic your users read, when and for how long, you can create a

The value of the manual

The old-fashioned way to measure the value of something is to put a price on it and make it available on the market. If people pay for it, it has value. But if you give it away, how can you tell? Thank-you notes? Would you get complaints if you don't include it? But you have to include a manual, right?

Well, some companies don't. Apple started this trend in 1999, when it decided not to create a manual for Mac OS 9. David Pogue, prominent tech blogger and former New York Times tech columnist, jumped in and wrote *Mac OS 9: The Missing Manual*, which was published in 1999 and picked up by O'Reilly Media in 2000. David Pogue went on to write numerous books – many of them best-sellers – just to fill the gap between the known cost of technical documentation and its assumed value. He wrote many more books in the “Missing Manual” series, their slogan being, “The book that should have been in the box”. Indisputable sales numbers prove their value.

So, your customers know the value of technical documentation. If you give it away with your product, though, as in a user guide or reference manual, it is harder to measure. Especially because most software these days doesn't come in a box, and product information can be buried in an arcane file system.



Image: © Andrii Torianyk / 123rf.com

data portrait of each of them. Your software can start to group them in novel ways that people, full of biases and preconceived notions, could not.

Subjects of interest

Beyond visualizing the consumption of content fragments, you can begin to perceive what subjects interest your users, not just which bits of documentation they are looking at. There will be forum posts, trouble tickets, knowledge base en-

tries, and user guide topics, all having to do with the same subject, "How to restart the XY module," for instance. The combination of advanced text-mining and data-analysis algorithms permits you to see the bigger picture – not what documents your users are looking at, but what problems are bothering them.

Now we have information that is not just interesting to tech doc teams, but to engineering, sales, management, product design and, indeed, the

entire enterprise. If calibrating the power supply is a persistent pain point for your field techs, maybe your support team needs to get in touch with engineering to figure out how to offer better support; your product design team might want to reconfigure it in the next version; and your training team could devote extra time to that topic.

Patterns and personalization

Once you know what subjects attract your readers, you will start to notice clusters of users, all interested in the same subjects. This gives you unprecedented power to offer suggestions based on what users with a similar reading pattern have looked at. You can offer your users a kind of curated serendipity, by using similar users' searches to affect result ranking.

Search results can be tailored to the individual. If a user is always looking at expert maintenance content, what is the point in proposing basic installation documentation? Personalization is even more important when the delivery channel is not a web browser, but more constrained devices such as heads-up displays and chatbots.

Real-time support

Using the new power of being able to hear and see your customers in the act of looking at your documentation, you can offer targeted, smart, real-time support. Instead of giving your chat window a general rule such as "pop up after two minutes," you can give it a smart heuristic such as "pop up if the user is looking for subject X, with a relevant suggestion." Suddenly you're being helpful instead of covering your user's browser tab with a useless interruption.

And if your phone support agents can see what the user has been searching for, they could be more helpful and faster than ever before. They will no longer have to ask an expert if they've tried rebooting, or offer a novice a complicated download and install. They will know the data portrait of the user and adapt to it.

Predictive support

From looking at the past with data visualization to perceiving the present with real-time support, you can look at the future. If a certain search pattern has tended to result in a call to support, you can start to predict when such a call is likely, and reach out to the customer first. A quick and easy call now could prevent a difficult, lengthy call later.

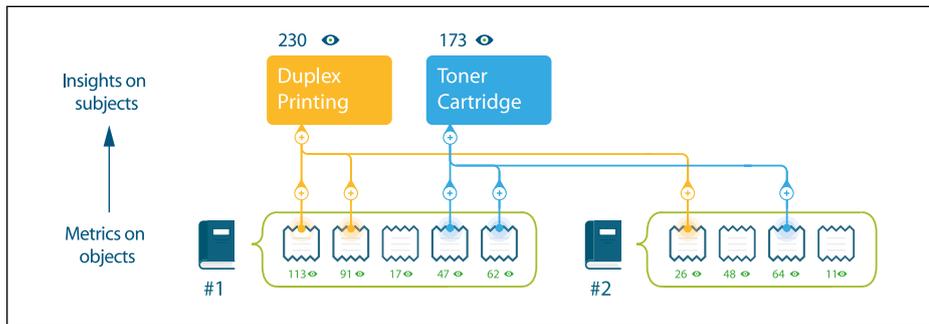


Figure 4: Numbers on distinct topics can be combined at a higher level to generate more insightful analytics. Source: Fluid Topics

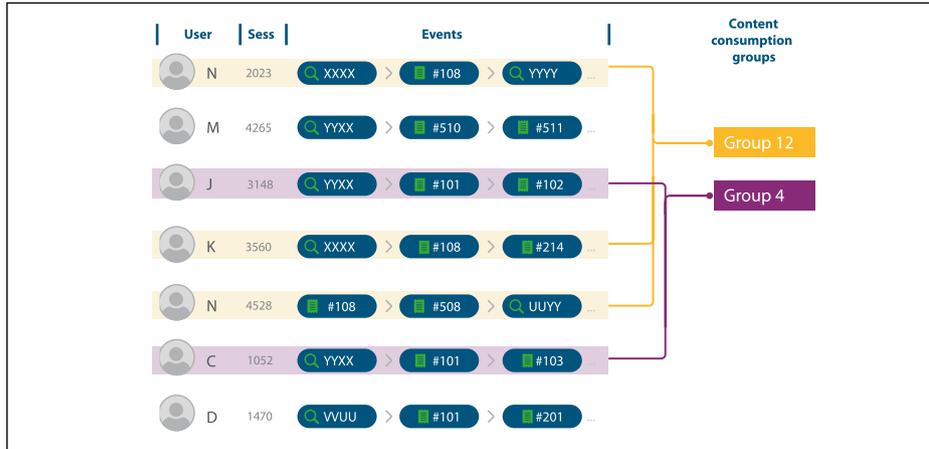


Figure 5: By mining content consumption, we can automatically detect and pattern groups of users. Source: Fluid Topics

User Group	Content				
	Book 1	Book 2	Book 3	Book 4	Art. 508
1			✓		
4	✓	✓		✓	✓
12					✓
274			✓		

Promote to user C content preferred by group 4

Figure 6: Search engine results and suggestions can be based on user characteristics. Source: Fluid Topics

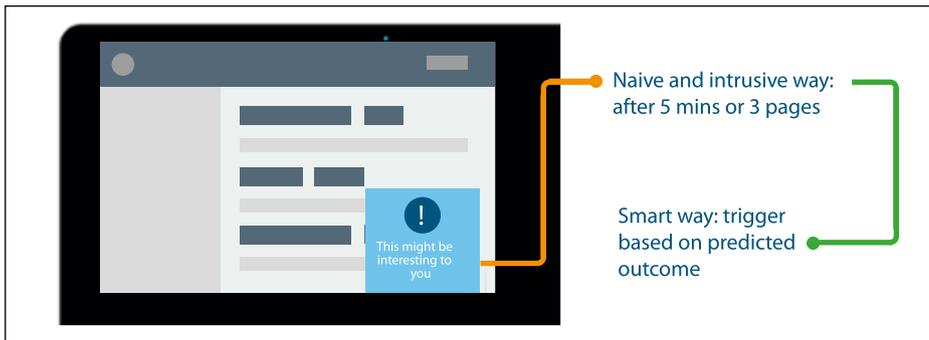


Figure 7: Popup windows could be displayed at the right moment by tracking user behavior and content consumption.
Source: Antidot

Conclusion

Your technical documentation started as something you sent out in the mail. All you knew is that it cost money to create and ship, and that you had to do it. Now it has transformed into a medium of communication between you and your users. And it's even better than a conversation, because it's totally honest. Your user is interacting with the

portal and has no reason to be anything except totally candid.

Do it right, and it will tell you what's bothering your users, and what you can do to help.

Resources:

[1] JohnHunter,TheDemingInstitute,August29,2013,
<https://deming.org>

ABOUT THE AUTHOR



Founder and CEO of Antidot and inventor of Fluid Topics, **Fabrice Lacroix** is both a serial entrepreneur and a technology pioneer. During the early years of the Internet boom, he launched Infonie, the first publicly traded French ISP and founded Zone-jeux.com, an online gaming platform, before tackling the vast challenge of information mapping and retrieval through advanced mathematical algorithms with Antidot.

@ lacroix@fluidtopics.com
 www.fluidtopics.com
 @fluidtopics

Intelligent Information for Smart Users

Get in touch with us to become a member and benefit from being involved in the further development of the standard. Share your opinions and your know-how with other experts and get involved in the decision-making process.

Be amongst the first to know which steps we will take next in the world of Industry 4.0!

Join the Consortium!

tekomp iIRDS
 Intelligent Information Request
 and Delivery Standard

More information:
iirds.tekom.de

