Designing an Analytics Approach for Technical Content

Arthur Berger
IBM Cloud, IBM, RTP, NC, USA
arthur.berger@ibm.com

ABSTRACT

Working on an enterprise cloud product, my documentation team rethought our approach to content analytics. Despite a variety of tools and awareness of industry best practices, my team felt stuck using analytics only in annual or on-demand reports to management, instead of to produce value for our end users. We employed Design Thinking practices to guide a multifaceted user research project that led to changes in the way that we created documentation and automated quality content checks. Key takeaways include to involve the technical documentation team in identifying not only what metrics to collect, but also how to collect, report, and use the metrics in order to increase buy-in and the likelihood that data analytics about content leads to meaningful change within the content itself.

CCS CONCEPTS

• **Applied computing** → Document management and text processing; Document management; Document management and text processing; Document capture; Document analysis; • **General and reference** → Cross-computing tools and techniques; Measurement.

KEYWORDS

Data analytics, content strategy, design thinking,

ACM Reference Format:
Arthur Berger. 2020. Designing an Analytics Approach for Technical Content. In *Proceedings of the 38th ACM International Conference on Design of Communication (SIGDOC '20), October 03, 04, 2020, Denton, TX, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3380851.3416742

1 INTRODUCTION: LOTS OF DATA, LITTLE CHANGE

As a technical writer for IBM Cloud, a growing platform whose featured products include Artificial Intelligence and Big Data processing tools like Watson, my team and I are constantly being asked to use data to find insights, improve delivery processes, and personalize content to users. We are fortunate to have access to analytical tools that let us customize dashboards with information about our users and documentation topics, and to produce a variety of reports that capture key metrics. However, we felt that data reporting was just another chore on the list of things that a "good" technical writing team does, and not a particularly helpful chore towards keeping our content house in order. In this industry report, I present how our team used IBM Design Thinking techniques to rethink the end-to-end content analytics experience and suggest takeaways for practitioners on how to take ownership of their own content analytics.

Every Friday morning, an automated build that I set up generates content analytic reports and sends these to the documentation team. As I review the charts of page hits by documentation set, my eyes tend to glaze over a bit. Because we are a growing product, our documentation page hits generally trend up week over week. If the arrow goes down, I might ask a few cursory questions as to why: perhaps there was a holiday last week, or perhaps another automated build that we have to post new documentation links to user channels went down. If the arrow spikes up, I might check this against a marketing schedule, such as a conference, demo, or training session that might explain why people are unusually interested in the docs this week. I also take in our top performing pages, and keep a mental list that certain topics, like networking ingress annotations, are perennial favorites and deserve more attention in our sprint planning sessions. Otherwise, outside mid-year and year-end reports, I do not use content analytics very much, or even regularly check all the other metrics that our product gathers.

Many practitioners and researchers have identified what metrics are important to measure content's efficiency and effectiveness. A review of the content analytics articles that have been published in the Society for Technical Communication (STC) practitioner-focused magazine, *Intercom*, suggests a variety of content metrics and potential applications for these metrics, as shown in the following Table 1. Content metrics can be grouped into areas like overall usage, freshness, findability, support, demographics, behavior, and delivery of content, which might be helpful for organizing a report or thinking about applications. I summarize the metrics and examples based on the references from *Intercom*, but this list is by no means exhaustive, complete, or exclusive only to the sources referred to.

These metrics can naturally lead to certain applications and content changes. Often, there is not a one-to-one correlation between a metric and a specific change, but rather numerous metrics inform a change or an overarching goal, such as delivering the right content to the right user at the right time (Gilhooly, 2020). Specific applications for metrics suggested by the literature include the following.

- Automatic changes or release notes [1]
- Subscribable content [1]
- Expiry date and automatic unpublish [1, 7]
- Measuring organizational content strategy maturity [8]
- Strategic guidance for different outputs like documentation for machines, chatbots, or voice agents [4, 9]
- Techniques for understanding users [10]

SIGDOC '20, October 03, 04, 2020, Denton, TX, USA

ACM ISBN 978-1-4503-7525-2/20/10…$15.00
https://doi.org/10.1145/3380851.3416742

Table 1: Common content metrics

Area	Metric	Example	Reference
Usage	Best and worst performing topics	Release notes has 700 hits/week	Gilhooly, 2020 [1]; Paolini, 2014 [2]
Usage	Clicks in the interface	Delete button hit 15 times/day	Ilan, 2018 [3]
Freshness	Time since last update	28 days since Index page was updated	Bailie, 2019 [4]; Gilhooly, 2020 [1]
Freshness	Time since published	1 year, 3 months since Index page was published	Bailie, 2019 [4]; Gilhooly, 2020 [1]
Findability	Return users to a topic, over time	600 users (67%) return to the Ingress page week over week.	Gilhooly, 2020 [1]; Carlos et al., 2016 [5]
Findability	Average time on page	5.5 minutes spent on the Ingress page	Carlos et al., 2016 [5]; Paolini, 2014 [2]
Findability	Average pages per visit	7 pages per visit	Carlos et al., 2016 [5]
Findability	Bounce rate	30% (100 people visited, and 30 leave the site next)	Paolini, 2014 [2]
Findability	Site search	Top 5 search terms	Paolini, 2014 [2]
Support	Visits to doc site with no tickets	10,000 doc hits on 7/1/2020 with 2 tickets opened	Carlos et al., 2016 [5]
Support	Doc vs. forum vs. tickets	10,000 doc hits, 8 new forum threads, 800 forum views, 13 tickets this week	Carlos et al., 2016 [5]
Support	Unanswered forum questions	Average of 2 unanswered questions per week	Carlos et al., 2016 [5]
Demographic	Segmentation and demographics	Internal vs. external users, geographic areas, age, job titles	Paolini, 2014 [2]
Behavior	Click-thru journey, behavior, and flow through content	From the landing page, 30% visit the how-to section next, and 50% of those visitors create an object	Paolini, 2014 [2]
Delivery	On-time delivery	Doc updates in line with product and marketing deadlines	Trunzo & de Vries, 2012 [6]
Delivery	Fixes during alpha, beta, and general release periods	50 alpha fixes, 30 beta fixes, and 8 general fixes show continuous improvement	Trunzo & de Vries, 2012 [6]
Delivery	Alpha, beta, and general release participants	50 alpha participants, 80 beta, and 1000 general release, compared to previous alpha and beta data can show if docs help encourage adoption	Trunzo & de Vries, 2012 [6]

- Predictive analytics and learning behavior of users [3]
- Identification of behavior patterns of groups of users as well as individual users [3]
- Relating content contributions to business objectives of macro and micro conversions [11]
- Better interactions with customer support [5]
- Ascertaining whether users are finding content quickly or the first time [2, 5]
- Monthly reports to management [5, 7]
- Celebratory successes and voice of the customer [5]
- User persona validation [2]
- Integration of behavioral data to change content shown, such as geolocation for language [2]
- Information architecture [12]
- Questions for user research and technical investigation [12]
- Places in documentation or product lifecycle where users experience problems [6]
- Input into how documentation impacts key performance indicators (KPIs) for the product, like time-to-value, onboarding, and efficiency [6]

Despite collecting many of these recommended metrics and trying some of these applications, my team felt overwhelmed by such a to-do list, with lots of data and little change in our content. While our analytics dashboards do not have all the metrics suggested by the literature, is the answer to pursue more charts with more data, when we do not feel like the data we have has led to meaningful change? Or perhaps the solution is to be more creative or more adamant in working data analysis into our regular team practices. But who has the time, especially in today's environment of budget cuts, hiring freezes, and just-in-time delivery? Our dashboards were useful in end of year reports when management asked us for proof points to brag about how great the docs are, and so that was the primary way that we used the dashboards.

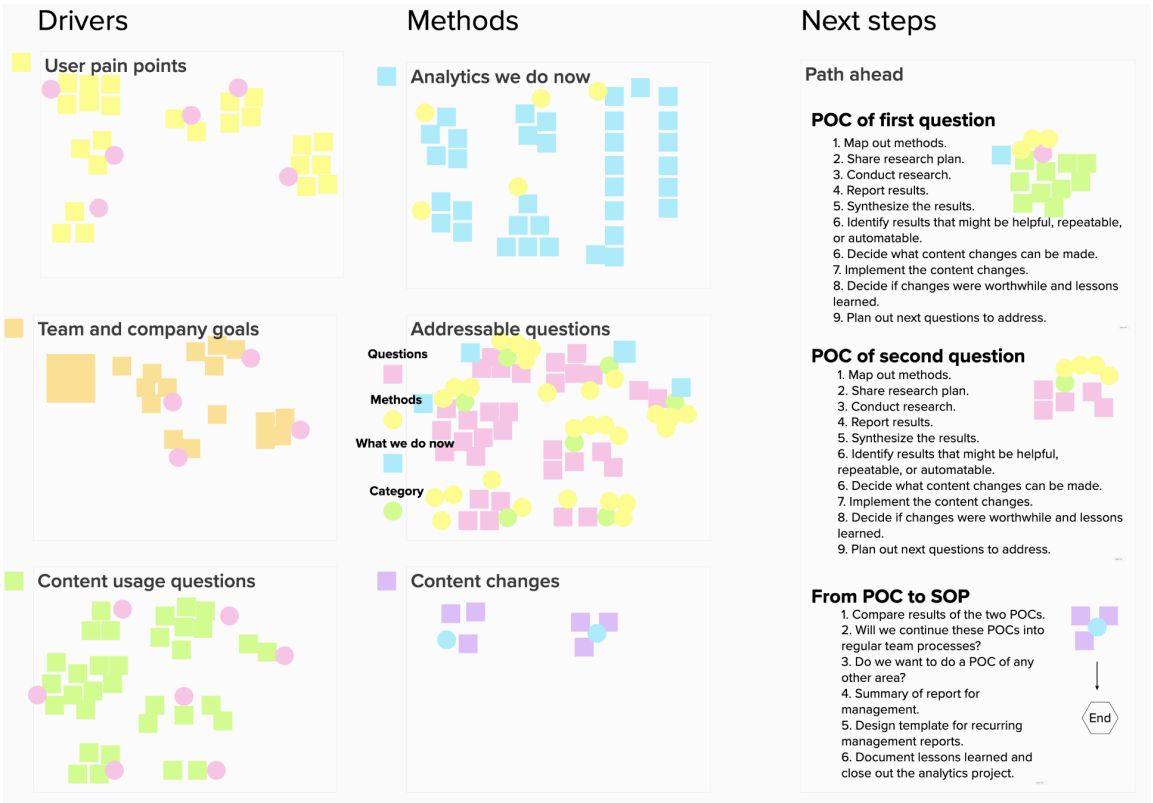

Figure 1: Design Thinking whiteboard to reimagine our content analytics.

However, we read and heard so much about the importance of content analytics that we decided there had to be a better way. Our team tried IBM Design Thinking [13] processes to reimagine how we identify, collect, interpret, and use content analytics to bring about not so much a measurable or incremental improvement in the documentation (we already do this), as instead a felt change in our approach, beliefs, and proof of the value of content analytics for the users of the documentation.

2 METHODS: DESIGN THINKING AND ITERATIVE EXPERIMENTATION

I led a series of Design Thinking sessions over the course of several weeks with all the members of our documentation squad. IBM Design Thinking is an enterprise-focused framework that is guided by a set of principles, driven by a loop metaphor of iterative ideation and development, and aligned by a set of keys that help teams implement the framework in a way that stays rooted to the needs of the users. It includes a toolkit of activities such as empathy maps, as-is scenarios, storyboards, and priority grids. More information can be found at https://www.ibm.com/design/thinking/.

We used a collaborative whiteboarding tool called Mural to brainstorm, organize, vote, and agree upon our ideas, as shown in Figure 1. To identify our as-is scenario, I came up with a few particular drivers that I wanted to ground our discussion in, including the pain points of our users, team and company goals, and our own assumptions and questions about content usage. Then, we identified the analytic methods that we currently have. For each of these areas, I explained the reasoning behind the area; then, we took time silently to generate ideas as separate sticky notes on the whiteboard; next, we organized them collaboratively into similar themes; and finally, we discussed each theme and item in it.

The next area that we addressed followed a similar process of discussion, but focused more around the to-be scenario of what questions we thought analytics could answer for us. These questions were then mapped against our as-is methods to help identify gaps where we might not have metrics to support a question. Finally, we voted on the priority of each thematic group of questions to help decide which one we wanted to focus on first for a proof of concept. The navigation and findability theme had the most votes as well as most unique votes.

For the proof of concept, I created a spreadsheet where I generated a list of methods that we had identified from our current work practices and the literature, tailored to the navigation and findability area that we wanted to focus on. Figure 2 shows an example of this spreadsheet. I mapped the methods against the research questions that we hoped to answer, to help visualize which method's results might address the most areas of concern that we had. Then, I categorized each method by how much effort it might take to complete, based on previous experience. Afterwards, I had one-on-one sessions with each team member to refine the spreadsheet and to

Research plan for Navigation and Findability Question

Research questions the method could help address

Method	Effort	Person	Find content	Getting lost	Framework	Product types	Versions	IA vs. goals	Keywords	Do next	Not found
Navigation usability test	High	Art	x	x	x	x	x	x			x
Search usability test	High	Rose	x	x		x	x		x		x
Survey with users	Med	Libby				x	x	x		x	x
Survey with SMEs	Low	Opal				x	x	x		x	x
Forum and tickets	Med	Ethan		x				x	x		x
Google keywords	Med	Lindsey		x		x	x		x		x

Figure 2: Example research method plan for proof of concept. The plans maps research methods, the amount of effort the method might take, the person responsible for researching, and which research questions the methods might address.

see which methods they would like to investigate further. After a certain amount of time, we agreed to meet together as a team, playback our results, and brainstorm what content changes the results from the methods could help us make going forward.

3 RESULTS: CONTENT ANALYTICS EMBEDDED INTO TEAM PRACTICES

Our findings from the methods in the first proof of concept area that we tried led to a variety of changes in how our team embeds content analytics into our processes. First, we found that much of the value in our research was not only in the results but in the time that it forced us to take to experiment how to get the results, think about the results, and decide on how the results matter to our user, team, and product context. This helped us understand why previously having dashboards or periodic, generated reports never seemed to matter much to our team, because we find greater value in engaging and making than in receiving a report.

Another focus of our proof of concept was in using the practices that worked for us and the metrics that we found to create our own scripts that would automatically check our documentation and recommend content changes. For example, based on our methods to identify keywords, top topics, and Google search result hits, we developed a script to scan our content and measure things such as headings, short descriptions, and reading levels. We tried as much as possible not to reinvent the wheel, but rather to pull in open source packages and use standard formulae to measure these things. However, we wanted to adopt this into our existing team automation and content quality checks in a way that we could control, customize, and include specific recommendations for how we could change the content.

Finally, we discovered that this type of multifaceted research takes a significant amount of time, upskilling, and resources to implement in ways that we felt mattered, even if it left us with further questions and uncertainties. While we had always recognized time as a barrier, we had previously not done such an in-depth combination of user research and analytical methods focused around design-driven question. This project gave us a better baseline for understanding what research questions we might tackle in the future, how we might tackle them and transform the results into usable content changes, and how many resources we might need to communicate to our management to do so.

4 DISCUSSION: FUTURE PLANS AND POTENTIAL IMPACT TO OTHER TEAMS

Moving forward, my team plans to expand our proof of concept from one documentation set into the entire product family of documentation, refining our automation builds in the process. Other practitioners can follow a similar Design Thinking process to find out ways to grapple with the questions that they have and the metrics that might help to lead to content changes unique to their situation. In particular, practitioners can use as-is scenarios and facilitated brainstorming sessions to identify a variety of drivers, assumptions, and questions behind their content before generating a list of questions that they want to research and what the to-be scenario is that these questions can help address. A gap analysis can be performed between the data and tools available vs. those that are needed. Questions and methods can be prioritized, and then a a short-term proof of concept can be piloted out. By enlisting the entire documentation team's help and playing back regularly the team progress, a team can arrive upon agreement together on what the content changes, future changes, and management report will look like.

A secondary takeaway that might help practitioners, academics, and students alike is the idea of collaborating together across a variety of research practices, instead of just thinking of page hits or a particular metric. Not only can multiple metrics be used to support a content outcome, but multiple research methods can also be used to get such data: usability tests, keyword searches, competitive analyses, literature reviews, expert interviews, product-level metrics, customer surveys, tickets and forums, and more. Instead of looking at each of these methods individually and exhaustively, Design Thinking outcomes can help guide people concentrate a variety of these methods and their results towards a specific, actionable problem.

ACKNOWLEDGMENTS

Thank you to my coworkers and management at IBM Cloud.

REFERENCES

[1] Megan Gilhooly. 2020. Product answers: Engineering content for freshness. *Intercom* (Jan. 2020). https://www.stc.org/intercom/2020/01/product-answers-engineering-content-for-freshness/

[2] Massimo Paolini. 2014. Bleeding edge analytics for technical communication. *Intercom* (Apr. 2014). https://www.stc.org/intercom/2014/04/bleeding-edge-analytics-for-technical-communication/

[3] Oded Ilan. 2018. Analytics and personalized content: The doorway to AI in training. *Intercom* (Feb. 2018). https://www.stc.org/intercom/2018/02/analytics-and-personalized-content-the-doorway-to-ai-in-training/

[4] Rachel Anne Bailie. 2019. The role of content operations in delivering content for machine use. *Intercom* (Aug. 2019). https://www.stc.org/intercom/2019/08/the-role-of-content-operations-in-delivering-content-for-machine-use/

[5] Jenn Carlos, Travis Kelly, & Vivanne Costa. 2016. Measuring content like a marketer: Getting started with web analytics. *Intercom* (Mar. 2016). https://www.stc.org/intercom/2016/03/measuring-content-like-a-marketer-getting-started-with-web-analytics/

[6] Stephanie Trunzo and Jozef de Vries. 2012. Analytics aren't only for MBAs: Measuring information architecture value. *Intercom* (Feb. 2012). https://www.stc.org/intercom/2012/02/analytics-arent-only-for-mbas-measuring-information-architecture-value/

[7] Chris Whitehead. 2016. Getting started with small data: How trends in analytics can add value to the role of the technical communicator. *Intercom* (Mar. 2016). https://www.stc.org/intercom/2016/03/getting-started-with-small-data-how-trends-in-analytics-can-add-value-to-the-role-of-the-technical-communicator/

[8] Carrie Hane, Dina Lewis, & Hilary Marsh. 2019. Measuring and advancing your content strategy maturity. *Intercom* (May 2019). https://www.stc.org/intercom/2019/05/measuring-and-advancing-your-content-strategy-maturity/

[9] Marianne Macgregor. 2019. "Conversationalizing" user assistance. *Intercom* (Aug. 2019). https://www.stc.org/intercom/2019/08/conversationalizing-user-assistance/

[10] Kumar Dhanagopal. 2018. Understanding our users: A comparative analysis of the techniques we can use. *Intercom* (Nov. 2018). https://www.stc.org/intercom/2018/11/understanding-our-users-a-comparative-analysis-of-the-techniques-we-can-use/

[11] Sara Feldman. 2016. Why should technical writers embrace content analytics? *Intercom* (Sep. 2016). https://www.stc.org/intercom/2016/09/why-should-technical-writers-embrace-content-analytics/

[12] Laura Palmer. 2012. What's data got to do with it? Information architecture and analytics. *Intercom* (Feb. 2012) https://www.stc.org/intercom/2012/02/whats-data-got-to-do-with-it-information-architecture-and-analytics-2/

[13] IBM Design. (2020, June 6) *Enterprise design thinking*. https://www.ibm.com/design/thinking/