



DITA Production Metrics: Looking at Topic Ratios



Keith Schengili-Roberts, IXIASOFT

What I think is still a largely under-appreciated aspect of DITA is that it enables documentation teams to measure ongoing content production and quality. Much has been written about DITA metrics, and the majority of that content focuses on measuring the Return on Investment (ROI) for moving to DITA plus a CCMS. But once the ROI for your company's investment in DITA has been established, continuing measurements can be made that can help improve the quality of content that your documentation team produces.

So why would you want to measure product documentation output? One reason is to gain a better understanding of the content quality that is produced. A sample DITA production measure that I think worth examining are DITA topic ratios, which look at the relative numbers of topics either within a single map or across several maps. Why would this matter? It can provide an information architect or documentation manager with information as to whether the DITA content being produced is properly “typed”, and also ensures that the technical writers on the team are creating and structuring content that is aimed at the user's needs.

Here's an example to help illustrate my point: let's take a sample user guide and see what types of topics it uses. There is a nice sample guide written in DITA that was produced by Joe Gollner and Eliot Kimber to show off DITA best practices—particularly when it comes to constructing effective keys—that is also freely available online <<https://github.com/gnostyx/dita-demo-content-collection>>. This fictional manual is for a software product called Thunderbird, and the User Guide tells the user all about how to use the program to manage a cluster of servers it monitors. In this case, I am looking specifically at the basic User Guide, which consists of 23 topics. With such a small data set, all you have to do is open each topic in the manual and keep a count of which topic type is being used. When you graph the results in a pie chart, you get results as shown in Figure 1.

Given that this is a User Guide—supposedly aimed at telling users how to operate the software—there is *a great deal* of conceptual and reference content but not many tasks that tell the user how to actually use the Thunderbird software product. There's even a slice depicting generic topics that are in this manual, which begs the question: are the Thunderbird technical writers properly typing all of their content?

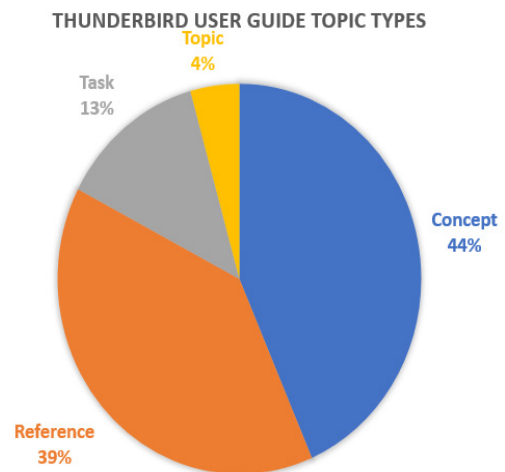


FIGURE 1: TOPIC RATIOS FOR THE THUNDERBIRD USER GUIDE

After viewing this result, if I were the technical documentation manager for the Thunderbird team, I would ask my writers whether they have added sufficient task topics to this manual to tell users how to operate the features of the product, as the evidence suggests strongly that this is not the case. As for the generic topic slice in the pie chart, further examination reveals that it represents a single generic topic for an FAQ, which answers questions like “How do I change my password?”, “How do I generate a custom query?”, and “How do I diagnose an issue with cluster activity?”. While I don't want to denigrate FAQs in general, I can't help but wonder whether these same questions should instead be answered either in additional tasks within the document, or be covered in separate troubleshooting topics.

As you can see from my analysis of the topic types comprising a single, sample document, I already get a perspective of a document's structure, which in this case indicates possible areas for improvement.

EXAMINING A REAL-WORLD EXAMPLE OF TOPIC RATIOS AT A SEMICONDUCTOR FIRM

When I was a technical documentation manager at AMD, I had access to several years' worth of data from the IXIASOFT DITA

CMS that we were using. This is where I first thought of the idea of analyzing topic ratios, which I started to do back in 2009¹.

The process for doing this within the IXIASOFT DITA CMS is straightforward: simply pick the number of topics created and modified for a given time period, ensure that “(Topic) Type” is selected in the search results, and then export the results. This information can then be imported into an Excel spreadsheet, and then do a count of each topic type using a simple formula: =COUNTIF(E2:E2000, “concept”), where column E is used for the results of the topic types, successively replacing “concept” for the other topic types, and then graphing the results.

At the time, there were two main audiences and types of documents that were produced:

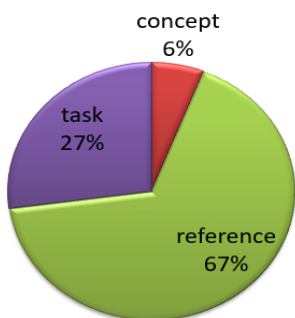
- ◆ end-user manuals that instructed users how to install and use the various hardware- and software-based features of their video cards
- ◆ electrical engineering documents designed to tell engineers about the various settings for board designs

In terms of overall volume, there were more electrical engineering documents (and therefore topics) produced, and there was little overlap between the two sets of document types. I experimented with various metrics that I could pull out of the DITA CMS and graphing them, and topic ratios were one of the areas I decided to look at. I set yearly parameters on the topic outputs, and then charted the results for each of the three years of solid data I had available. Figure 2 shows the results.

As you can see, in the first full year of DITA-based topic production there was a significant number of tasks that were produced as opposed to what was developed in the following years. Delving further into the data, I found that the majority of task topics in that first year were contained within the user guides. This makes sense, since the electrical engineering documents consisted primarily of reference topics along with a few concept

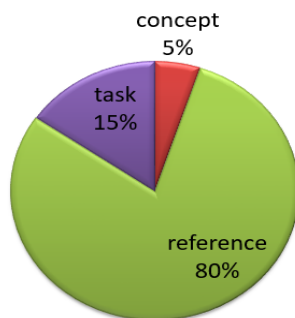
¹ This information was previously published in a presentation I did back in 2011, but this article is the first time I have explained the rationale extensively for topic ratios in print.

Topic Ratios (2007)



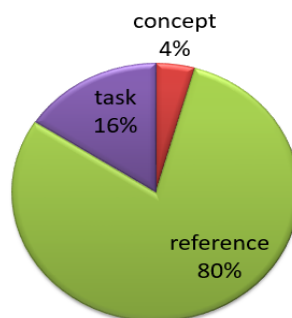
Total topics: 1703

Topic Ratios (2008)



Total topics: 2145

Topic Ratios (2009)



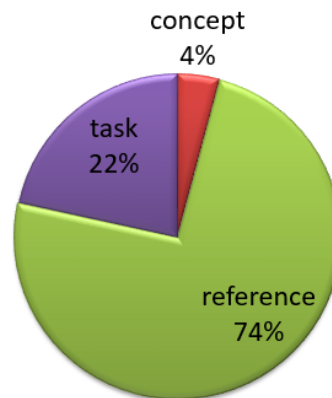
Total topics: 2716

FIGURE 2: THREE YEARS OF PRODUCTION TOPIC RATIOS FROM A SEMICONDUCTOR COMPANY

topics. Over the next couple years the total production output of content increased, and the ratio of end-user manuals and electrical engineering manuals was roughly the same—so why the disparity in the number of task topics that were subsequently used in those publications? For the next year, I made it a goal for the technical writers to ensure that they were focused on the needs of the user for the end-user manuals, and to ensure that any procedures were fully described.

I already knew well before year-end that the technical writers on the team had taken this to heart, and the results can be seen in Figure 3.

Topic Ratios (2010)



Total topics: 4520

FIGURE 3: THE FOURTH YEAR OF TOPIC RATIOS PRODUCED BY THE SAME COMPANY, AFTER INSTRUCTING WRITERS TO CREATE MORE TASK TOPICS

There was a significant increase in the number of task topics that were created for 2010. You will also notice that the total number of topics produced and published went up significantly too, representing an increase primarily in the number of electrical engineering documents produced. In other words, the number of tasks topics created for the end-user documentation is actually higher than is represented here.

AN EXAMPLE OF TOPIC RATIOS FOR A SOFTWARE FIRM

When I joined IXIASOFT, one of the areas I was interested in exploring further were DITA production metrics. Early in 2016 I was asked to do some presentations on DITA production metrics both at conferences and for clients, and I was asked whether I could use IXIASOFT's own DITA production data for these purposes.

Naturally one of the metrics I wanted to investigate was DITA topic ratios. I figured that the results were likely to show a different set of topic ratios, one that focuses squarely on end-user documentation, as that is the focus for IXIASOFT documentation and its two main products (the DITA CMS and TEXTML Server). There is no equivalent to the electrical engineering documents that were also produced at AMD, therefore what I expected to see was a much higher ratio of task topics, followed by concept topics describing the features of the software, and a much smaller number of reference topics. And that's pretty much what I found when I initially looked at the first couple of years' worth of topics I examined (as seen in Figure 4).

As you can see, the topic ratios when we are looking only at end-user product documentation are overall skewed heavily towards task topics, followed by concept topics, and then a much smaller percentage of reference topics. There's also a thin yellow slice of generic topics, and when I originally showed this to the IXIASOFT documentation team lead, she vowed to track down those topics and eliminate them for the following year.

There is something else here worth pointing out: there were far fewer topics created or modified in 2015 than in the previous year. When I looked at the numbers more closely, I found that there were more new topics created in 2014 than in 2015, and while the number of published documents was similar across both years, much content had been tweaked (or modified) in 2015.

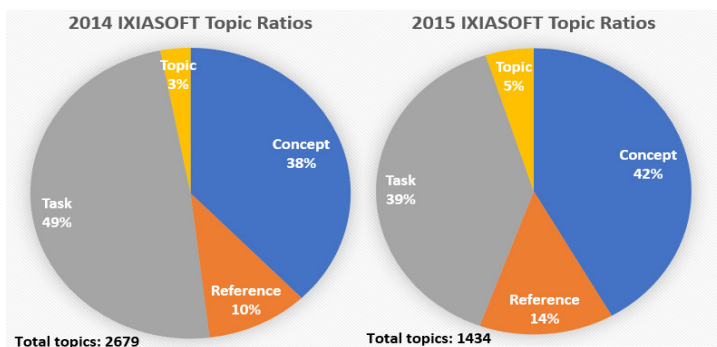


FIGURE 4: IXIASOFT'S END-USER DOCUMENTATION TOPIC RATIOS FOR TWO YEARS

These statistics are good, as this suggests content reuse rates are optimal.

Looking at the pie charts, I was surprised that there were not more task topics, since the primary goal of a user manual is telling the user how to use the product. When I looked at one of our more typical documents—the manual for the IXIASOFT DITA CMS—I found that just over 60 percent of its topics were tasks. The one thing I suggested to the head of our technical documentation team in 2016 was to focus on the needs of our users by ensuring that all possible procedures for our products be explained.

Shortly after 2017 began, I crunched the numbers for the previous year. The results in Figure 5 were what I had hoped for. Sure enough, more task topics had been created or modified. Delving further, significantly more task topics had been created over the course of 2016, suggesting that we are providing better coverage for all of the procedures involved in using our software products.

Could we be going overboard and documenting too many procedures and perhaps not providing enough conceptual or reference material? I don't believe this is the case, as I suspect the ratio we are seeing for 2016 may be close to the ideal for any firm producing user manuals, but now that we know what the overall trends of our topic production are like, we can keep an eye on this. Even better, we now have an idea as to what to ask our users when it comes to the quality of our content: Would you like to see more conceptual content on using our product? Are we missing any procedures for specific product features? Do you want to see more information on basic settings and configurations (which would cover most of our reference topics)? Using this technique provides us not only with information on what types of topics we are producing, but also provides us with guidance

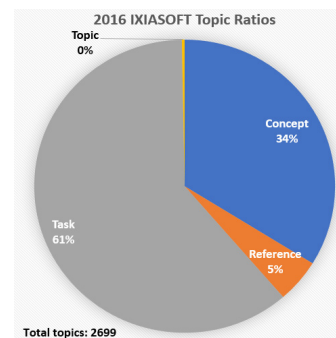


FIGURE 5: THE LATEST END-USER DOCUMENTATION TOPIC RATIOS FROM IXIASOFT, INCLUDING SEPARATE CREATED AND MODIFIED ONLY CHARTS

on the types of questions we can ask our users in order to better improve our documentation.

Notice also that the number of generic topic types were greatly reduced. This fact made the head of our documentation team happy. From an end-user perspective, properly typed content ought to be more tightly focused on meeting the needs of the user.

The one thing that this information also tells me is that we are missing troubleshooting topics. We have begun to author topics in DITA 1.3, and yet we do not have a single troubleshooting topic in the documentation set. When I did a search for the word “trouble” or “troubleshooting” in our existing topics several good candidate topics come up—this is something I am hoping to see more of over the course of this coming year.

SUMMING UP

Measuring topic ratios is one of the ways a technical documentation manager or information architect can help determine whether the documentation being produced meets the needs of its users. It is relatively easy to measure topic ratios, and the data it produces provides an accurate view of how your topics are being targeted at your readers. It can also provide direction on how to guide your writing staff and ensure that the content your group produces is focused squarely on users. The examples I have provided here from two different industry sectors should give you a benchmark for comparing them against your own results. 