

<DITA>



DITA Metrics: Similarities and Savings for Conrefs & Translation

By Mark A. Lewis



Intro

This paper shows an interesting similarity between the reuse mechanism of the DITA content reference feature and a translation memory system (TMS). We then discuss this similarity in terms of cost and show that the savings that can be achieved using DITA's content reuse feature are similar to the savings possible using a TMS.

Let's start with a look at how the reuse mechanism of a TMS is similar to that of DITA. During the translation process, content in the form of DITA topic files is imported into a TMS. In the new content, the translation software finds all sentences and determines if they already exist in the translation memory. Do they exist? If not, the new sentences are translated and the translated versions of the sentences are stored in the TMS. When new content is imported, if an incoming sentence is found to already exist in the TMS, then it is not translated. Instead, the already translated version of the sentence is used (reused) in the translated version of the incoming topic. When no more sentences can be matched to the TMS and replaced with their translated version, then the topic is sent to the translators.

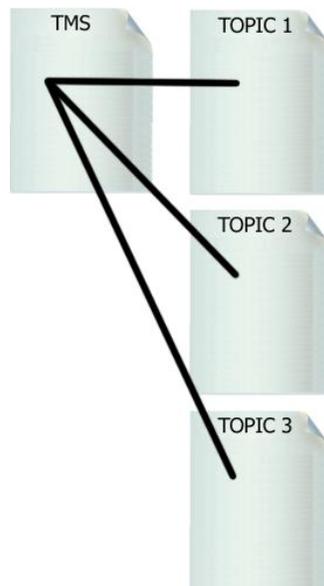


Figure 1: TMS reusing already translated content in the translated version of new topics

DITA Metrics: Similarities and Savings for Conrefs and TMS

The content reuse mechanism in DITA is similar. In DITA, this feature is known as *content referencing* or *conref'ing*. When writing a new topic, if you determine that the content you need already exists in the content repository, you can reference and reuse that the existing content.

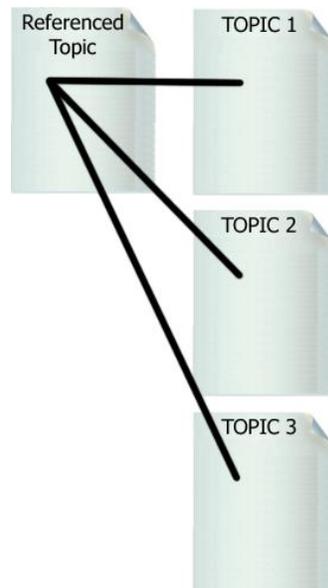


Figure 2: DITA topics reusing already written content

Given this, we see the parallelism between already translated content being reused and already written content being reused. Identical content does not have to be re-translated or re-written. Even though a TMS reuses content at the sentence-level, while DITA allows reuse at the content-element level, all the sentences in that content element would be translated and reused. So the idea of similarity in reuse holds true.

Similar Savings

If the content reuse mechanism between the two is similar, then the cost savings should be similar. In the second half of this paper, let's look at this similarity in terms of a cost model so we can show that the savings trend is also similar. Because DITA is topic-based and a TMS works with topics, we'll use a topic-based cost model. We'll use the model developed in "DITA Metrics: Reuse Strategy and Savings Trend With Warehouse Topics." Here's a quick recap, but see the original white paper for details.

Traditional cost metrics focus on the cost of a page¹. Since pages are similar to topics, we can create a cost model that represents the cost to create a DITA topic. For each element type in a topic, we determine the average number of occurrences of that element type and then assign labor hours. We can do this for each topic type (task, concept, etc.) and create a cost model that allows us to predict the cost of topics in a documentation product.

¹ For more information on traditional cost metrics, see Lasecke, Joyce. "Stop Guesstimating, Start Estimating. In *Intercom*, February 2006 issue, Society for Technical Communication.

DITA Metrics: Similarities and Savings for Conrefs and TMS

Consider a content referencing and reuse strategy that stores reusable content in a *warehouse* topic. The content in the *warehouse* topic is referenced by topics that are similar. Reusable content can be conveniently “warehoused” in its own topic as illustrated in Figure 3

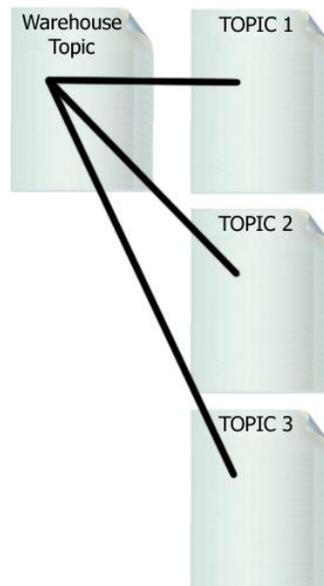


Figure 3: Content in a DITA warehouse topic can be reused in newly written topics

This model allows us to easily measure reuse using a cost model and metrics that are topic-based.

DITA Metrics: Similarities and Savings for Conrefs and TMS

Figure 4 shows the savings trend that emerges when content referencing and *warehouse* topics are used to document *similar* products. It shows how much it costs to document each additional *similar* product.

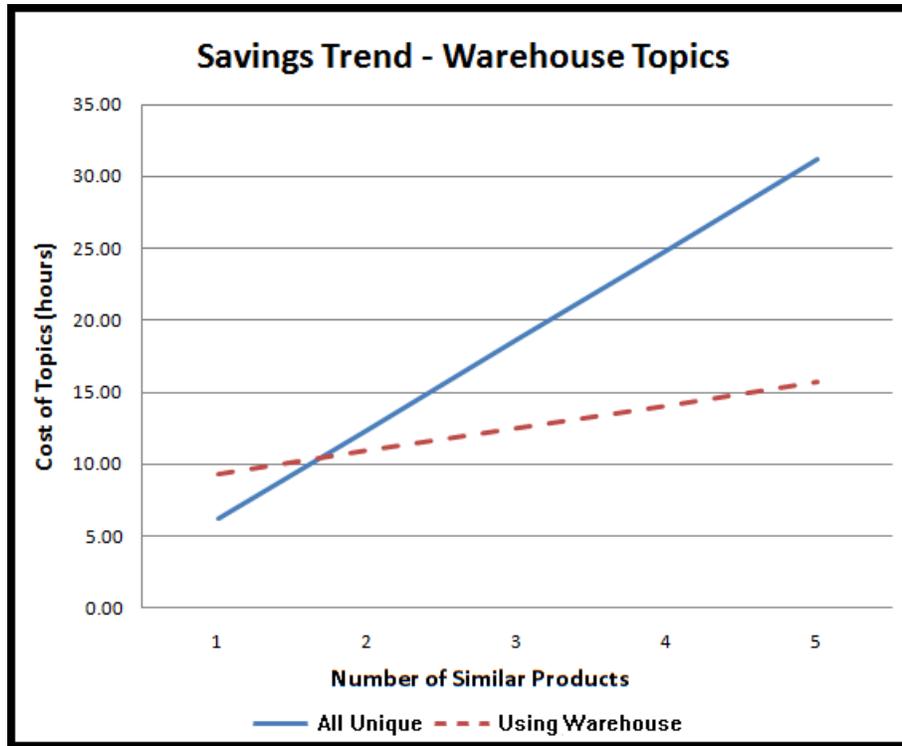


Figure 4: The savings trend for DITA content reuse

Figure 5 reviews the similarity in content reuse mechanisms.

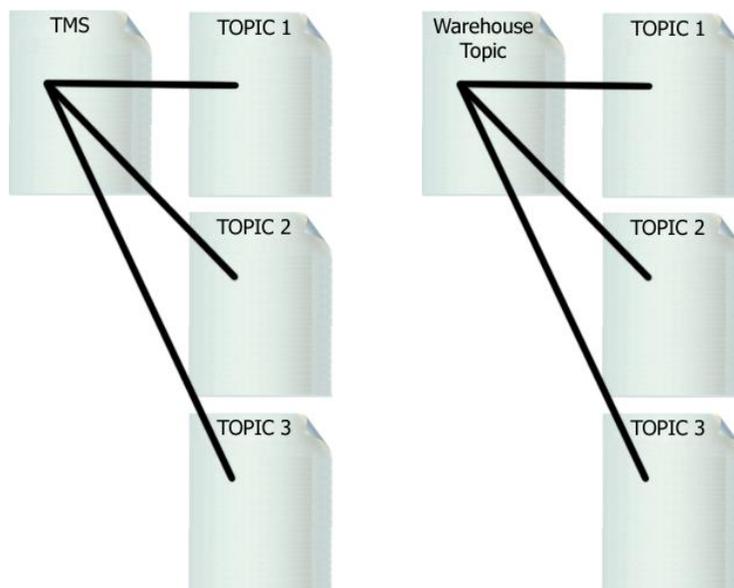


Figure 5: Content reuse is similar

DITA Metrics: Similarities and Savings for Conrefs and TMS

If the content reuse mechanism is similar, then the topic-based cost model and savings trend is similar. So the savings trend for DITA in Figure 4 could be converted to show the savings trend for a TMS, Figure 6.

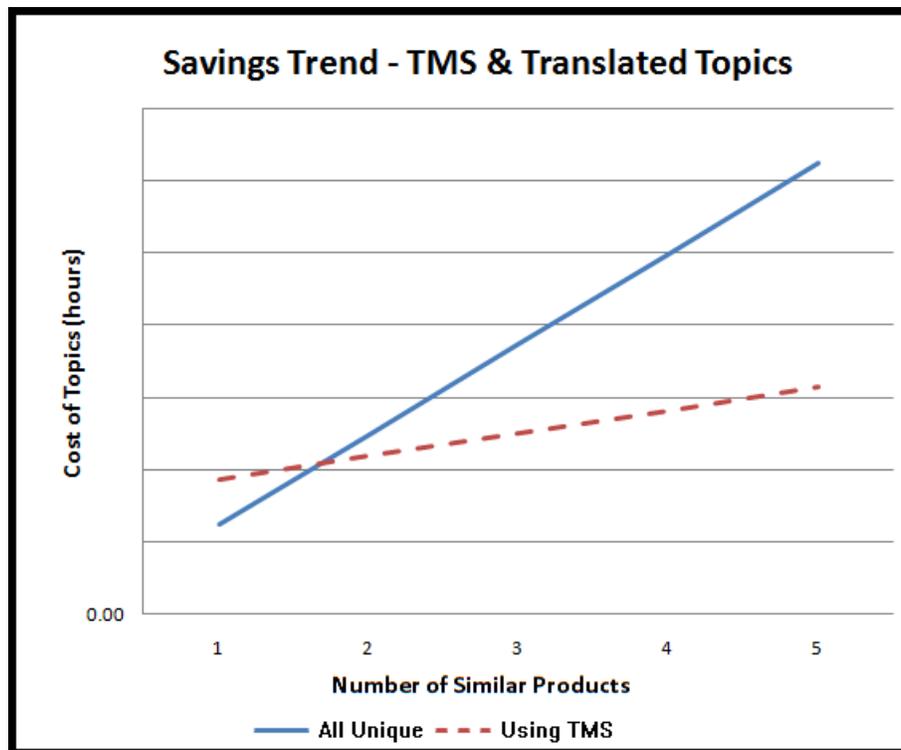


Figure 6: The savings trend for a TMS and translated content reuse

Savings From Combining DITA and a TMS

The savings are similar, but they are separate systems. Does combining DITA and a TMS save even more than what we've discussed? Only the source content of conrefs needs to be translated, not the conrefs themselves. Therefore, content can be more quickly imported into a DITA-aware TMS because the TMS should be able recognize conrefs and not process them for sentence matching. This does not reduce human effort. But, it does reduce content import time, so it is still an extra benefit worth considering.

Conclusion

We've shown how the reuse mechanism of the DITA content reference feature is similar to that of a TMS. Already written content is reused and already translated content is reused. Identical content does not have to be re-written or re-translated. The similarity can be extrapolated to a similarity in cost models and in savings trends. This provides a simple topic-based view of the savings possible with a TMS that you can use to help justify the move to a TMS and to DITA.

References

Lasecke, Joyce. “Stop Guesstimating, Start Estimating.” In *Intercom*, February 2006 issue, Society for Technical Communication.

About the author

Mark Lewis has received Society for Technical Communication (STC) awards for Distinguished Chapter Service and the Florida Technical Communications Competition. Mark is the DITA Product Manager for Usability and a product evangelist for Quark. He has presented on technical writing, DITA, and object oriented design topics at DocTrain, STC, DITA North America, and other national conferences. Mark is a member of the Organization for the Advancement of Structured Information Standards (OASIS) DITA technical committee. He and John Hunt are co-chairs of the OASIS DITA for the Web subcommittee. ROI is a hot topic in technical writing. Mark has authored several white papers on DITA Metrics that prove the savings and high content reuse percentages possible with DITA’s structured, topic-based architecture. His DITA metrics model was a JoAnn Hackos - Rare Bird Award 2009 Competitor. Mark manages the DITA Metrics LinkedIn group.

Send questions or feedback to mlewis@quark.com or hyperwriters@hotmail.com.

Other papers in the DITA Metrics series:

DITA Metrics: Cost Metrics – Part 1

DITA Metrics: Savings Trend With Reusable Master Topics

DITA Metrics: Reuse Strategy and Savings Trend With Warehouse Topics

DITA Metrics: Case Study of Quark XML Author Documentation

All papers available for download at:

The DITA XML Org Community

<http://dita.xml.org/search/node/dita+metrics>

and through the

DITA Metrics LinkedIn group

<http://www.linkedin.com/groups?mostPopular=&gid=2916249>

Keywords and Phrases

DITA metrics DITA reuse math DITA economics DITA cost metrics DITA reuse metrics DITA percent reuse DITA percentage reuse DITA statistics DITA formulas DITA ROI DITA case study DITA business case
single source publishing metrics
single source publishing statistics
Reuse calculate reuse calculation
DITA Cost Calculator
DITA Cost Calculations
DITA Reuse Calculator
DITA Reuse Calculations
Cost trends savings trends
XML average cost topic