

Denominator-Problem and Last Published Schema (LPS)

Implementing a possible solution to the "Denominator Problem" for classes with more than one publish schema

Problem Statement

- We want to give users of Dashboard an accurate sense of how many people are answering questions, spending time on pages, filling in inputs while, in fact, it is often the case that new versions of the album are being published so the number of pages, knowledge checks and inputs are changing over time. So, how do you calculate the percentages against total published things that are changing over time and still give the user an accurate sense of reality?
- Some learners work at their own pace and complete topics out of order, etc. Some users in a class may have completed a given topic in one schema and others in another.
- We also want to focus the users' attention on the elements that are current. Say, for example, that the user is an author and the author has deleted a confusing KC. We want the user to see information about the KCs that remain, not the ones that are no longer live.

Overview of Possible Solution

LPS is short for Last Published Schema and it means that we pretend that all the work a learner did was done in the last published schema, the one that was in effect at the last moment the user was in that topic.

Terms and Definitions

A users score on a content division is given by:

- **kcScore** is $\text{kcCorrect} / \text{kcPublished}$
- **inScore** is $(\text{inAnswered} + \text{kcAnswered}) / (\text{inPublished} + \text{kcPublished})$
- **pvScore** is $\text{pvSeconds} / \text{pvPubSeconds}$

Where

- **kcCorrect** is the number of correct answers to kc questions in all topics of the content division.

- **kcAnswered** is the number of total answers to kc questions in all topics of the content division.
- **inAnswered** is the number of answers to input fields in all topics of the content division.
- **pvSeconds** is the number of seconds of browser time spent all topics in the content division.

- **kcPublished** is the number of kc questions published in all topics of the content division.
- **inPublished** is the number of inp questions published in all topics of the content division.

- **pvPubSeconds** is the number of seconds of estimated time to complete all topics of the content division.

Some Solution Details

Someone who has actually completed none of the content necessarily has a score of zero, and someone who has only part of the content will have a maximum score of less than 100%.

Many classes may be too short to expect the students to complete the whole album, students might join late or leave early; but those are not the problems we are trying to solve here, such classes and users will simply have perhaps unfairly low scores.

However this problem may be compounded by the album being revised while a class is using it. This happens whenever the date range of the class overlaps the date range of more than one publish schema for its album.

Originally we attempted to base the scores on material from all of the schemes seen by the class. This was considered unworkable because

If a kc or inp was present in one schema and then removed, learners working in the later schema would never have seen it, and yet their missing answer would be counted against them. Similarly, when kc or inp were added to the later schema, learners working in the earlier one would have never seen it.

Drilling down into LPS / the Current Direction

This solution rates a learner on each topic in terms of how well they did with the elements that exist in the last publish schema for which they worked on that topic,

We take the last publish schema in which a user worked on a topic to be the one in effect on the last day for which we have pageView record for that user and topic.

The learners score for the whole class is based on only the topics which exist in the last publish schema seen by the class, defined to be the one in effect on the last day of the class.

This means that when an album is re-published while a class is in session, any topics removed will no longer

contribute to its user scores.

It also means that when album is re-published while a class is in session and new topics are added, all users will lose percentage points if they do not complete them.

If a user has worked on given topic and does not return to it while the class is in session, his scores will be based on the last schema in which he worked.

However if the topic is updated while the user is working on it, or he comes back to the now revised topic after having previously completed it, He will be expected to complete any kc or inp that have been added, and will lose credit for any kc or inp that have been removed.

This might be the best we can do, and might be good enough for now, but it does have some problems.

What's to like about this perspective?

- Seems like it might well produce numbers that are informative
- I think we will be able to achieve good response times

Some problems and concerns with this direction

- If new topics are added towards the end of the class there might not be enough time to complete them.
- If new questions are added to some topics towards the end of the class only learners working in these topics at the time will be expected to complete them, but might not have enough time to do so.
- If a learner has completed a topic, and revisits it casually or accidentally after a revision, he ends up losing credit just because they revisited a topic after publication.
- If a topic is moved from one module/section to another, we must not consider it to be removed and then added again, otherwise users will be forced to redo it. This may be tricky to implement as our current schema identifies topics by their full path (including module and section).