30 LearnRank: Towards a *real* quality measure for learning

Erik Duval

Katholieke Universiteit Leuven, Belgium

This paper starts from the notion that quality is context dependent and proposes "LearnRank", a context-dependent ranking algorithm focused on learning applications, as a vision for how to *really* measure quality. This proposal is based on early experiences with the ARIADNE learning object repository and avoids the focus on characteristics that may be easier to measure, but that are less relevant to the usefulness and usability that determines the relevancy of a learning object. This paper does not develop a precise definition of LearnRank. Rather, it analyses in some detail the notion of context and how that influences LearnRank.

30.1 Introduction[115]

This chapter starts from the notion that "quality" is not so much a characteristic of a learning object[116], but rather a characteristic of how such object is used in a particular context. Many facets influence the quality of a learning object in a particular context: target audience, learning objective, time, location etc. This notion of context applies to quite traditional learning resources as well as to more technology based ones, as the following examples illustrate:

---

[115] Acknowledgements: This work is financially supported by the "Onderzoeksfonds K.U.Leuven//Research Fund K.U.Leuven" and the ProLearn Network of Excellence (http://www.prolearn-project.org/, itself supported by the European Commission). The work also benefited greatly from the discussions in the CEN/ISSS Workshop on Learning Technologies (http://www.cenorm.be/isss/Workshop/lt/) and within my research unit (http://ariadne.cs.kuleuven.ac.be/hmdb/jsp/Wiki?Research). Special thanks to Wayne Hodgins for the inspiring conversations!

[116] We will use the term "learning object" in a very loose sense here, referring to any kind of resource that is useful and relevant for learning. In this context, a learning object can be any kind of content, of any kind of granularity. It can also include other kinds of entities that are relevant for learning, like activities or people. See (Verbert et al., 2005) for a more in-depth analysis of a learning object content model with a much more precise definition of "learning object".

- A physics handbook for 16-year-old children will have a different "quality" for PhD. students in physics than for the audience it was originally intended for.
- A handbook that is very well suited for self-directed learning by PhD. students is not necessarily that useful as a teaching support in classroom settings.
- A physics handbook can be well suited as a general introduction in the overall field of physics, but may be less suited for a maintenance repair technician who needs to refresh a particular physics principle.

Not only is quality, by its very nature, context dependent, it is also intrinsically subjective. In that sense, the subject (for our purposes: the learner) is part of the context – or determines the context, that is more of a philosophical difference that is less relevant here.

30.2 Early Ariadne experiences

Our experiences over the last decade in the ARIADNE Foundation (Ariadne) illustrate how seemingly sensible strategies to deal with quality, typically derived from the non-digital world or methodologies developed for academic publishing, are naïve and misguided in the context of a large-scale repository.

We originally intended to impose a quality policy that would restrict the introduction of learning objects into our repository (Forte et al., 1999). After long and difficult discussions, we had to recognise that there was no consensus on precise and usable criteria that would determine the quality. One of the important reasons why such consensus could not be reached was the severe cultural difference in how such quality was viewed in for instance Latin and Nordic European countries.

We then compromised for quality review of the *metadata about* the learning objects, rather than reviewing the objects themselves. The idea was that, if for instance the metadata asserted that a handbook was meant for graduate students, then that was a piece of information that could be reviewed for its relevancy and accurateness. In order to implement this approach, we operated a scheme in which reviewers were assigned a domain of expertise (related to subject, but also, for instance, to language), and reviewed periodically new learning objects, or rather: the accuracy of their metadata. Initially, only the objects whose metadata were reviewed were made available to end users. However, this created a serious bottleneck, as it would sometimes take a long time before the new content would actually be reviewed[117].

In the end, we opted for a scheme wherein the *ranking* of a learning object in a list of search results could be influenced by, among others, the fact that its metadata had been reviewed. Reflecting on this evolution with the benefit of hindsight, it seems to me that there is a deeper trend here: rather than thinking about quality in a binary way, where resources either do or do not qualify to quality criteria, we

[117] (Zemsky and Massy, 2004) report a similar finding with the MERLOT repository, where less than 10% of the material is peer reviewed – although this concerns an actual review of the content itself, rather than the metadata only, the findings go in the same direction as our experience.

should reconsider this notion as one that is multi-facetted and that influences the ranking rather than the inclusion of an object in a particular context.

30.3 The problem with quality

The fact that quality is context dependent complicates its automated processing: it suggests that algorithms need to take into account a wide variety of characteristics for a realistic determination of quality. Worse, the subjective nature of quality suggests that such automated processing may be intrinsically impossible: how can an algorithm capture the highly personal preferences and characteristics that determine the experience of an individual?

I believe that the complication described above leads to an overly strong focus in the "learning quality world" on less relevant quality aspects that *can* be measured and processed, like process characteristics or simple learning object properties. Whereas it is clear that these aspects are relevant to some extent, they avoid the real problem of trying to measure quality *in context*.

I believe that a different approach to quality is needed altogether and will refer to it here as a "LearnRank": the basic idea is that we can very well process automatically *real* quality aspects, very much in the same way that for instance Google can serve quality web pages, or Amazon can serve quality books and other products, or Tivo can serve quality television content[118].

30.4 LearnRank

30.4.1 Context revisited

The notion of context, and how it determines quality, is worth considering in more detail. There has been substantial research on the subject of context in the field of ubiquitous computing – see for instance (Coutaz et al., 2005).

However, here we briefly analyse some of the characteristics of quality and how they relate to learning.

- *Learning goal:* What the learner wants to learn (Miller et al., 1996)
  This is quite obvious: a student in medicine will not learn about the functioning of the human heart with material on economical statistics… In a slightly more subtle way, that same material on economical statistics may be relevant to someone studying the evolution of disruptive technologies. However, if the learner is not aware of the relevancy of the material, then it will still fail to support his learning in an efficient and effective way.

[118] This theme was also explored in (Rehak, 2004) and (Duval, Hodgins, 2004).

- *Learning motivation:* Why the learner wants to learn (Murphy, Alexander, 2000)
  An inherently motivated student will need less explicit rationale for why she is engaged in a particular activity, whereas a student going through mandatory material in a formal learning session may need to be reminded of why statistics is relevant in the context of learning about for instance business administration. Of course, if the relevancy of the material needs to be made explicit, then that seems to suggest that there is a problem with the instructional design in the first place.
- *Learning setting:* Activities surrounding the learning object (Dalziel, 2003)
  The sort of activities that Learning Design (IMS) or Learning Activity Management System (LAMS) define around learning objects (or within the objects, that distinction is not important here) also influence quality: as an example, a provocative historical document may be very appropriate in a discussion context, and much less so in an unsupported self study mode.
- *Time:* Available time, as well as time of day
  When a mechanic needs to learn how to repair a piece of equipment necessary for an ongoing emergency operation, the time-to-learn should be minimised at all cost. However, when the same mechanic is learning about the same kind of repair procedure during a master class, a more in-depth learning process, which also considers the underlying principles of physics may be more appropriate.
  Time may well have an effect in a different way, as many of us learn differently according to the time of day: there is at least anecdotal evidence that suggests that many learners are better at more abstract learning in the morning, and more background learning in the evenings.
- *Space:* Geographical as well as architecture
  Especially in "mobile learning" applications, the influence of space on quality of learning objects is obvious: connecting the learning experience to the physical environment can reinforce its effect tremendously.
  Physical spaces in general, and architecture in particular, also influence the efficiency and effectiveness of learning. Ambient learning through educational "roomware" could have a deep impact on the quality of learning resources.
- *Culture and language:* Not only geographically (Niles, 1995)
  Especially in a multi-cultural and –lingual context like Europe, the effect of language and culture is obvious: a learning object in a language that the learner doesn't master will not have the same quality as an otherwise similar object in her mother tongue.
  What is often neglected, is that this observation applies not only to geographically determined differences in culture, but also to the difference between other kinds of cultures, such as the academic versus the corporate culture, or a learning context for engineers versus scientists versus medical learners, etc.
- *Educational level:* Age and learning background
  It is rather obvious that toddlers learn differently from elderly learners, and that age has an impact on how for instance colours and sound can either reinforce or detract from the learning experience.

Similarly, academic researchers and "knowledge workers" probably have a different set of expectations and requirements for learning objects from repair mechanics or nurses.

- *Accessibility:* Design for all (Stephanidis, 2001)
  Finally, it is important to note that some objects may require specific auditory, visual, motor or other skills from the learners. Designing resources with that in mind, so that alternative interaction modes are possible, in a "design for all" approach, often benefits everybody, and not just the "disabled" learner.
  As an example, most of us have problems learning from highly visual material while driving a car, as our eyes tend to be engaged with the task of driving. Making the relevant resources also available in a non-visual way (for instance: in an auditory way) not only benefits the learner with a visual disability, but also all the driving learners!

30.4.2 PageRank revisited

As long as the web was relatively small, the main selling point of search engines was how much material they indexed and how efficient they were in processing a query.

However, as the web moved past the "tipping point" and the number of results to a typical query rose from tens to hundreds to many thousands and even millions, ranking the results in a meaningful way became more and more important. Nowadays, the success of search engines like Google and Yahoo depends to a large extent on their ability to rank results in response to a query. Indeed, Google's success is often attributed to its PageRank algorithm, used to rank search results. Contrary to common belief, this algorithm is only to a minimal level based on keywords, the number of times they occur and the location in the document where they occur.

Much more innovative and crucial is the use in PageRank of incoming links to determine the relevancy of a particular document. As explained in (Boldi et al., 2005):

> "One suggestive way to describe the idea behind PageRank is as follows: consider a random surfer that starts from a random page, and at every time chooses the next page by clicking on one of the links in the current page (selected uniformly at random among the links present in the page). As a first approximation, we could define the rank of a page as the fraction of time that the surfer spent on that page on the average. Clearly, important pages (i.e., pages that happen to be linked by many other pages, or by few important ones) will be visited more often, which justifies the definition."

In this way, PageRank exploits the human activities of all web authors who decide to link to pages that they consider relevant. Note that this sort of algorithm does not require any librarian type of effort, but rather is integrated into the very act of

authoring material in the first place: this is important, as "librarian metadata don't scale" (Duval, Hodgins, 2004b), (Weibel, 2005)!

30.4.3 Towards the development of LearnRank

If we apply the basic idea behind PageRank to learning, then the "LearnRank" of a resource should indicate how useful people have found this object for their learning. And, as with PageRank, we would need to be able to determine this without asking the learner, author or librarians to provide additional metadata about the object in question.

Objects that have been used in many contexts, or, more importantly, in many contexts that are relevant to a specific learner, should have a higher LearnRank for that learner. The underlying assumption here is that learning objects with a higher quality for a particular context will end up being reused more often in that context than lower quality ones. As LearnRank becomes more and more established, that assumption will become a self fulfilling prophecy, as the higher quality material will be higher ranked, and therefore more used, and therefore this effect will strengthen itself in a quality feedback loop.

In an ideal world, we would actually bootstrap and steer this process through empirical data on the learning effect that specific objects have actually caused (or helped to realise) in specific contexts: if we were able to track that a particular simulation helped graduate students understand with a high degree of efficiency and effectiveness certain effects of the laws of thermodynamics, then, surely, the LearnRank of this object could be quite high when a similar learner is trying to master similar material in a similar context.

However, even though we are optimistic that progress is being made on gathering that kind of empirical data on learning effect (especially in a corporate or military context, where, admittedly, the task of measuring the relevant indicator is often simpler), there are alternatives available already now to help bootstrap the LearnRank algorithm.

- Imagine that 10% of professors in Human-Computer Interaction in the French language for undergraduate students start using a particular tool with their students. Is that not a strong indication that this tool has a high "quality" in that context? What about 20%? Or 80%?
- Suppose that we track (as we can!) the correlation between the objects that learners work with and their performance on a post-test that assesses whether they have actually mastered a specific law of thermodynamics. Would that correlation not give a good indication of "quality"?

Of course, there is much work to be done on elaborating and evaluating the fine details of the LearnRank idea, but the principle should be clear from what has been presented above. Now that the technical standards are finally in place to enable the development of an open infrastructure for learning, we can finally achieve the scale that is necessary to get to the numbers where the idiosyncrasies of one learner or teacher or context will no longer skew everybody's results.

30.4.4 If content is king, then context is queen

Moreover, just like Google and other search engines rely on search terms (and more and more on context, such as location, past search histories, etc.!) to constrain the web graph to the portion that is relevant to the user at that moment, learning applications can constrain the search space to only those resources that are relevant to the learner.

Most interestingly, in a learning context, we can actually go much further than web search engines, because we can exploit much richer metadata about the user, his context and the learning object:

- Learning objects can be described with much richer (learning object) metadata than is typically the case for arbitrary web pages. Moreover, these metadata can automatically be generated, taking into account the learning context (Cardinals et al., 2005). As a consequence, we can restrict the search space of the user to *really* relevant learning objects, with the appropriate technical characteristics, suitable for his budget, accommodating his learning approach, in a language that he can learn in, etc.
- As learning objects are typically deployed in Learning Management Systems that provide explicit context, LearnRank can also rely on this information to take into account the characteristics mentioned in section 3.1.
- Using attention metadata, we can track what a user actually does with a learning object, beyond simple downloading or accessing it (Najjar, 2005). The potential here is huge as it allows us to create a "usage trail" of a learning object, which will eventually enable us to deduce metrics that indicate how good an enabler for learning a learning object is in a particular context.

30.5 Conclusion

Now that we have the open standards in place to build a large scale infrastructure for learning, we can start focusing on quality, much in the same way that web search engines shifted focus from being exhaustive to providing relevant results.

With this paper, I'd like to call for more focus on the development of good LearnRank measures, rather than on only indirectly and partially relevant indicators of quality for learning.