

I2G

INTERGEO

Deliverable N°: D4.3

Review System First Implementation

The INTERGEO Consortium

Sept 2008

Version:

Main Authors:

Paul Libbrecht (DFKI GmbH)

Christian Mercat (Université Montpellier 2)

Sophie Soury-Lavergne (INRP)

Jana Trgalova (INRP)



**Project co-funded by the European Community
under the eContentplus Programme**

Project ref.no.	IST-507826
Project title	INTERGEO - Interoperable Interactive Geometry for Europe

Deliverable status	submitted
Contractual date of delivery	M11 – August 2008
Actual date of delivery	M12 – September 30 th 2008
Deliverable title	Review System First Implementation
Type	Report
Status & version	submitted
Number of pages	20
WP contributing to the deliverable	WP4
WP/Task responsible	Paul Libbrecht
Authors	Paul Libbrecht (DFKI GmbH) Christian Mercat (Université Montpellier 2) Sophie Soury-Lavergne (INRP) Jana Trgalova (INRP)
EC Project Officer	Spyridon Pilos
Keywords	quality, review, community, resource
License	This document is available under the license <i>Creative Commons Attributions Sharealike 2.5</i>

Contents

1	Introduction	6
1.1	Comparable Quality Approaches	6
1.2	Usage scenarios	8
1.2.1	Ranking	8
1.2.2	Improving	9
1.2.3	Reviewing	10
1.3	Experiments and first revision	11
2	The Review System	13
2.1	Technical Architecture	13
2.2	Development Methodology	14
2.3	Usage of I2geo Review System	14
3	Outlook	17
3.1	Implementation Status	17
3.2	Risk Management of the Quality Framework	17
4	Acknowledgements	19
5	Bibliography	20

Executive Summary

The quality assessment in INTERGEO has two main objectives: *ranking* the “good” resources first to increase confidence in the content, and *improving* the resources by identifying criteria that can be worked upon by its author.

The tool to achieve this goal is an online questionnaire, taken *a priori* by teachers interested in using the resource, and *a posteriori*, after the teaching event has taken place. This questionnaire was already validated by tests in the classroom and revisions.

This questionnaire must be easy to use in order to be taken by numerous teachers, and deep enough that it can actually be used to monitor the quality. We solved the issue by asking 8 main broad questions, that can be answered “as is” for a quick review, or opened up to reveal deeper questions. Depending on the involvement of the reviewer, from 8 to 52 questions can be answered. This information will help both the author to improve her resource and fellow teachers to choose resources appropriate to their needs and taste.

This multidimensional quality assessment is averaged into one quality ranking. The weight of each review depends on how thorough it is and on the confidence level of the user that contributed it: a seasoned teacher or a pedagogy expert is more trusted than a user that is whether new or misbehaving by systematically over or under-rating.

Its implementation is built on the Curriki platform. We redesigned completely the existing quality monitoring system to our framework yielding, we believe, a system that invites to the review activity as a community participation action.

1 Introduction

Interactive geometry, just as most forms of computer-supported learning resources, may have tremendous learning effects, and may just as well produce learning failures. The quality of a learning resource is an indicator that attempts to classify learning resources by rating it; although the quality is often summarized in a single dimension (e.g. from *poor* to *good*), its assessment is generally realized during a review process which encompasses many dimensions and is bound to a given educational context.

Because knowing about the quality of a resource can be a great help to identify that such a resource could be used in a specific educational activity, the INTERGEO project has devoted a whole work-package to this task. The work-package 6 has delivered the INTERGEO QA Standard [MSLT08] which provides details on the quality dimensions and also explains well the insertion of the quality assessment exercise within the larger scope of the lifecycle of a learning resource.

This is particularly appropriate for the INTERGEO platform, explained in [ELD08], which not only catalogues learning resources but allows them to be edited live, thus offering a resource-centered collaboration space.

In this report, we present the implementation of the review system which is the central tool of the quality assessment activity. This system is visible at <http://i2geo.net/>, the INTERGEO platform, integrated within the metadata enrichment workflow.

This deliverable first presents typical usage scenarios that present the review-system in action in realistic cases. It then describes the review system and its surrounding platform and concludes with an outlook.

1.1 Comparable Quality Approaches

Quality in e-learning is not a new field and many initiatives have been running about it. Most did not suit our approach of requesting peer-based quality evaluation. The approach presented in [MSLT08] is based on the results of the e-Quality project which seems to have pushed the furthest possible the integration of quality reviews with the lifecycle of the learning resources.

As explained in [ELD08], the INTERGEO platform is based on the Curriki platform. This platform comes with an existing quality monitoring system, the Curriki Review System [inc08] which is split in two facets:

- The basic *file-review* is an automated procedure which checks regularly the contents of the learning resources under various criteria and, if the threshold of minimal requirements is not passed, make the resources private notifying the users about it per email. Criteria such as *unchanged templates* and *emptiness* are easily checked in this way. The i2geo platform intends to re-use these routines. Users are then expected to either repair the necessary aspects or enter a dialogue with the editorial board.
- The review process of Curriki is based on a team of voluntary reviewers whose work is activated by authors or users nominating the resources as *ready for review*. The reviewing board then assigns it a mark along with a commentary.

The INTERGEO quality approach wishes to be different than this by stimulating multiple reviews made by users in various educational contexts. As explained in [MSLT08], it thus bases on a questionnaire that every user could be able to fill and relaying to the comment-forum of each group or resource the appreciation of this review. No particular enrollment process is wished before filling a review but an identity should be verified.

The MERLOT repository has a similar review-process presented at [mer08] which also inserts into the workflow of comments about the resource.

Most other systems base on community communications to support the quality development. For example the *talk pages* of the classical Wikipedia server or the Connexions' collaboration means [Pro08]. This makes it impossible to extract machine-processable results of the review process allowing, for example the search engine to bring higher ranked resources judged of a higher quality.

Several other initiatives rank just about any URL in a single-dimensional rating. This has appeared insufficient for the INTERGEO purposes where the lifecycle of resources intends to be visible.

Sésamath is a French association of secondary math teachers. One of their activities is to produce the highly successful repository of exercises **MathEnPoche**: Its more than 1700 exercises are used by 8 750 registered teachers and their more than 450 000 registered students. In order to produce these exercises, the Sésamath team, after having identified the need for a particular development, assigns the task to a small group of teachers. They define the activity as a pedagogical scenario. A team of developers then implements the activity as a piece of software. The activity is internally reviewed and possibly enhanced cyclically. Once deemed suitable, it is placed on a beta-test web-site where 375 dedicated teachers enlisted as testers report on each activity. Bugs, suggestions and comments are gathered on this test web-site forums. When a consensus is reached,

the activity is validated and makes its way towards the official release web-site. Regional mailing lists, gathering around 8 400 teachers, are as well a place where comments are collected and taken into account but once validated, the activities are not reviewed by the users, neither for searchability nor for improvements identification. This project, while clearly collaborative, introduce differences between the actors. Their roles are different and don't evolve much. Time will tell whether INTERGEO will need the introduction of such roles. These roles are described in the deliverable D6.1 [MSLT08], based on the e-Quality processes model [The04]: several hats for users such as "content manager", "pedagogical planner" and so on, but we have not yet felt the need to actually implement them, they remain symbolic and not attached to an explicit artifact.

1.2 Usage scenarios

Quality assessment in INTERGEO has several aims, leading to several different scenarios. The main objectives are *searchability* and *improvement*. That is, reviews should help to rank first the *good* resources, and help to *improve* the not so good criteria of a given resource. As was said previously, review has to be seen in a multidimensional way, a resource can be very good along some criteria and could need improvement along other criteria. The fine grained review framework that we propose helps the author with identifying which elements in her resource can be worked upon, leading to the next version. The teacher that reviews a resource should be able, on the other side of the review system, to send these signals to fellow teachers and to the author of the resource.

Here we list several usage scenarios that have to be addressed by the system.

1.2.1 Ranking

- Catherine browses the resources on the INTERGEO platform by clicking the paragraph she wants to illustrate in the online version of her textbook. This selects for her some competencies attached to the paragraph and a fuzzy search is made among the resources in order to find suitable ones, in effect affecting a score to each resource. Her user's preferences are taken into account and the first scores are weighted against these preferences, weighting higher resources that match her preferred language or the level of the classes she has declared. Then a third weighting is performed according to the quality of the resources, as a single number. A list of the first resources is presented to her.
- Catherine explores the resource that ranked first in her search. Her first impression, playing with the resource online for a few minutes, is good. She reads the comments left by the previous teachers who reviewed it. The 3 comments written in her one language are presented, but a counter informs her that she could read 4 other comments in Spanish. It comforts

her to know that their opinion was taken into account by the review system on every item of the questionnaire.

- But the comments she reads seem, although enthusiastic, very technical to her; so she chooses to explore the review deeper and opens up the results of the questionnaire. The main 8 items have good marks, except one, “I know how to implement this activity” about pedagogical planning. So she opens this item up in order to see more precisely what was badly marked. The four statements that were summarized in this item are not too good, especially the statement “A schedule of the activity is proposed”, which is particularly bad. Since she is only a beginner in interactive geometry, she backs off and decides to go with the second resource, which has marks that are not as good but which was deemed straightforward to implement in the classroom, according to the reviewers.
- Because she understands that these quality criteria are very important for her, she decides to weight once and for all the pedagogical planning items higher in her preferences so that the projection from the multidimensional quality to the ranking weight takes into account her own criteria. Her ranking is personalized to what is important for *her*.
- Roberto is an expert in interactive geometry. His passion is to neatly implement tricky figures and improve existing ones. But he doesn't want to write text that much. He tweaks the search engine in order to look for resources with *bad* marks in interactivity but good ones on scientific and pedagogical content in order to spot resources that could benefit from his expertise.

1.2.2 Improving

- Günter is a good teacher. Everybody around him praises him for his talent. What a shock when the first reviews came in! He took it very personally at first not to get an A+ as he expected. But he went back to the reviews to understand the marks he got. First, a comment spotted some typos, which he fixed quickly. Then he realized that he got lots of good marks and only a few dirty spots spoiled the overall rank. A comment suggested two other competencies than the only one he chose, and it was true, it suited the subject better. Okay, he was a bit harsh with his students, which were above average so he was not exactly following the curriculum, therefore he should reposition the level as well. And yes, it was clear for him, with his great expertise, what to say to get the students to start with a few hints, and there was no harm giving away some of the little pedagogical tricks that made him famous in his school. After carefully reading the review and taking action where it was needed, he grew over some time an online fan club that was waiting for his releases and warmly discussing them.

- Boris belongs in an association which collaboratively creates and uses math content for courses. Some of this content uses interactive geometry. They use the review system in order to store and exchange ideas around ways to improve their content before actually using it. They found it more practical than multi-purpose bug-tracking systems found in project management systems. When he has to work on an exercise, the first things he looks for are the bad points to improve. It happens that some teachers use their figures and it is ok by him as long as they give feed-back to pin-point the improvement venues.
- Olivier has nurtured a resource for several versions, he has seen the quality reviews going up and the number of users accumulating over the revisions. He takes pride in the improvement of the confidence level put in it. He remembers when he was wondering, as the first reviews of the first version were coming back, whether he should stick with improving this resource, and taking with it the burden of the first negative reviews, or redesign a completely new one, with the risk to have no quality reviews and therefore a low confidence level. He looks back at this decision as the right one: when he asked his first “customers” to come back and reevaluate their questionnaire, most of them were pleased by the interest Olivier was putting in their opinion. A few of them actually went on board and are now enriching the resource as co-authors, and began new collaboration ideas their discussions lead to.

1.2.3 Reviewing

- Alice tried out a resource suggested by an online friend. She gave a first quick review after using it for a while. While preparing her course, she actually thought about specific aspects that she really liked: it was easy to induce the students to make a conjecture when this point was dragged around. She went back online to the review where she had crudely reviewed the resource. She found the review standing as it was previously, opened up the relevant item “Interactive geometry adds value to the learning experience” and fills in the more specific questions, giving a particularly good mark to the item “Dragging around, you can illustrate, identify or conjecture invariant properties”. She leaves a comment describing how the feature could be implemented. She is comforted in her effort by knowing that more specific reviews are taken into account with a higher weight than lighter reviews.
- A week after the course where she taught using the activity, Alice went back online to give back an *a posteriori* review. She ticked the checkbox stating that the activity has taken place. She knows that it will be taken into account more than her previous *a priori* review. Things went smoothly except that the schedule was not that on tracks, she drifted a bit from what was described. So she lowered a notch her mark on that point,

leaving a comment that it was in fact longer for her. But the students showed that they got the idea in the classroom the day after. So she went to that item and put a good mark there.

1.3 Experiments and first revision

After the first release of the questionnaire [MSLT08], that was based on didactical expertise and pedagogical analysis, the methodology was tested against average secondary math teachers. There was no experimentation in the classroom. It took place in São Paulo, Brazil, in a private university training math teachers. It was conducted by Ana Paula Jahn, under the supervision of Jana Trgalova and Sophie Soury-Lavergne from the INRP [JTSL08]. The teachers had an average of six years teaching experience but most had a low expertise in interactive geometry, they were trained with Cabri II Geometer for six hours. Apart from the country of choice, this group represents our main target audience. They worked on a selected activity as we expect a teacher would:

1. First, they solved with a paper and pen the mathematical problems that arose in the activity.
2. Then they analyzed the activity without any constraints, on their own as professional teachers do.
3. The *a priori* questionnaire was then filled.
4. They discussed possible implementation of the resource in their classes, elements that ease its appropriation, as well as elements they would like to share once the resource experimented in their class.

As a whole, they found the questionnaire well built. Of course they understood more deeply the relevance of the questions once the activity was fully anticipated. The *a priori* questionnaire arose their interest in little points of didactic that they would have clearly missed without the help of an accurate questionnaire: for example they were opened to interpret errors by the students not as a misconception of the activity but on the contrary, since the student herself can invalidate the false conjectures that she may come up with, as an active way of building solid knowledge based on a complex mental representation of the situation. The *a priori* questionnaire was pointing out that “The drawing is clear and sharp” or that “Different configurations are easily produced”, which is great, better than paper and pen, but the *a posteriori* questionnaire pointed out much more relevant pedagogical insights like “I am helped in identifying resolution strategies, whether correct or invalid, that the students may choose”, which were overlooked before the activity took place. Which means that the questionnaire drove them in the way they conducted the activity and helped them change the way they were looking at interactive geometry, from a nice gadget towards a tool to improve teaching.

The collaborative part was as well essential: They felt an urge to share their experiences and report on their findings, on how they would deal with the misunderstandings the students can run into, on suggestions they had in order to improve the activity, whether in the instructions given to the student or advice given to the teacher on how to conduct the activity. This sharing activity, writing and reading questionnaire results, appears to guide the pedagogical implementation.

The experiment is a little biased in the respect of that last point because the teachers were physically gathered and shared several hours in common. They were eager to share, they were personally involved in the process. But if only a fraction of their enthusiasm to collaborate in order to improve the activity and their teaching transfuses to strictly online communications, the INTERGEO framework will prove a success.

We discussed in D6.1 [MSLT08] measures to let the teachers feel part of a community that cherishes their opinion and takes it into account. Being physically together is obviously a measure that goes into that direction, but online tools can help as well: of course forums to discuss related questions and experiences, but as well counters that point out how many contributions were made, how “alive” the activity is, and the fact that every user contributes to this life by nurturing it with his feedback. We should not shy away people who like their privacy and we should as well be careful of the fact that reviews can be interpreted as criticisms, that are felt personally by the authors. Therefore the balance between involvement and indifference will have to be fine tuned.

As a conclusion, the experiment induced some little reformulations of a few statements but validated the overall structure. It pointed out especially the fact that the *a priori* questionnaire, far from being a chore to be dealt with lightly, is really an eye opener for the teachers. They came back from their class understanding the relevance of what they took at first as tedious hair-splitting.

2 The Review System

The INTERGEO platform is a web-based service running the Curriki platform tuned for interactive geometry needs as well as for the INTERGEO search engines requirements: cross-curriculum search taking in account quality reviews. The platform described in [ELD08] allows users from all over Europe to contribute resources, to metadata annotate them, to discuss about them, and, as described in here, to review them.

In this section we present the technical foundation of the review system and its features.

2.1 Technical Architecture

The INTERGEO review system is based on the XWiki object structure explained in [Mas07]: the reviews are documents with an object that is a record of field representing the answers provided by the reviewers. They can be listed, viewed, and edited; only the originator of the review can edit a review.

The records contain the fields described in the questionnaire of [MSLT08] with a summary textual answer for each section. It also contains the overall ranking which is deduced from the individual questions using a weighted-mean. Finally it contains the educational context for which this review was done (independent of the possibility of having been experimented in classroom or not).

The reviews are stored within a separate space making them independent of the resource thus enabling different rights management on the reviews and the resource themselves. We expect to find a normal amount of a few dozens of reviews for an average resource.

The edition of the reviews is a simple HTML form with sections for each domain of interest. Each section is made of several questions, each expecting an answer from 1 (disagree) to 6 (fully agree). Each section can either receive one global judgement or, if unfolded, obtain a judgement for each point in which case a mean is applied automatically to obtain the section's judgement. This process is achieved in the client.

The 8 sections, together, need to be summarized to provide a global ranking used in the display of the overall quality. This is where the *personal weighting*

appears: the user's profile has a tab for quality preferences where the users can indicate the weights (unimportant, medium, important) they give to each section of the reviews. These weights, converted to numbers (0.33, 0.6, and 1.0) allow the overall ranking to be computed and stored. The index does not store one overall-ranking per resource and per user. Instead, the choice of users is matched with a weighting profile and the overall ranking is stored in the review-summary for each weighting profile. The search index does the same and the queries are boosted differently depending on the overall ranking for the user's profile.

Because computing all the means of the reviews for the purposes of displaying them may be computationally intensive, in particular within list of resources such as the search engine, a *review-summary* object is linked to each resource and is updated every time the resources are changed.

As of today the review-summary tracks the mean of the overall-ranking of each reviews, the number of reviews and the latest review dates. This update strategy will allow us to adapt the review-summary computation methods to: more social oriented preferences (as described in section 3.2), simple history-based decrement (where more up-to-date reviews are more weighting), or to one that relies on the space of educational contexts.

2.2 Development Methodology

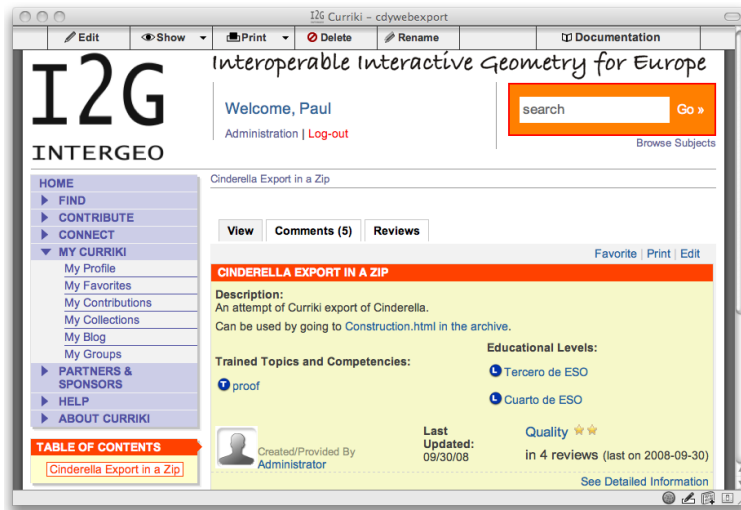
The quality review system of i2geo is an XWiki application as described in [Mas07]. That is, it is a set of XWiki documents which can be scripts or classes and contain attached objects. The code is written in Velocity for the rendering part to HTML, in Groovy for the summarization part, as properties file for the translations and in XML for the class descriptions.

The code is not entirely written by hand, a part of it is deduced from a model of the questionnaire, itself written using a carefully authored javascript which generates the HTML form and the XML describing the values of each class. Insertion into XWiki is done using the web-browser for the velocity parts and with the xar import facility of XWiki for the classes. The resulting application can be exported as an application archive to other Curriki having exported the necessary macros.

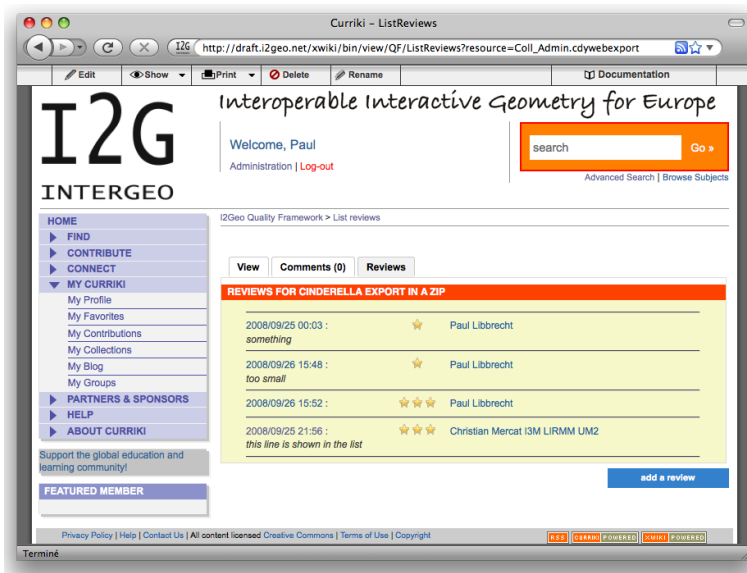
This development method has the big advantage of offering flexibility in the review form and its underlying model while still producing a runnable user-interface and storage system.

2.3 Usage of I2geo Review System

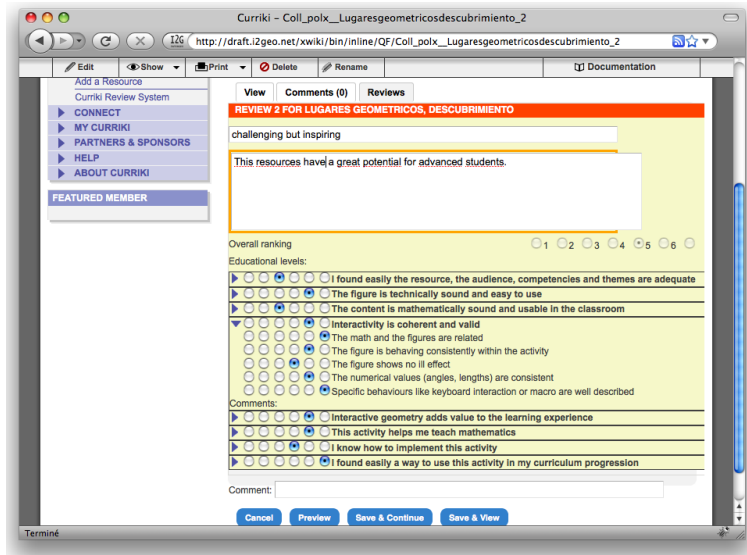
The i2geo platform displays the review summary within the metadata presentation for each resource as in the following screenshot:



The curious user may well be intrigued by more details about the reviews and its comments. She can therefore switch to the reviews tab so as to see the list of all reviews in inverse chronological order.



If the user is logged-in, she can add another review to this list which presents her the review form with all sections folded. Our user can open some sections and provide details for them.



3 Outlook

This deliverable has presented the review system of INTERGEO, a system meant for collaborative reviewing practice. The methodologies, use-cases, and software tools have been presented.

3.1 Implementation Status

The current review system is about to be deployed on <http://i2geo.net/>. Screenshots above have been taken on the developer-preview server.

As of today, reviews can be filed, are summarised, the rendering with foldable sections is working. The update routines are prototypical and the summarisation is unweighted. The personal profile weights is not implemented yet, as is the user-level-updating, and the display and input of educational levels. We expect the development to be concluded by the end of October with refinements expected to such aspects as the user-level-updating strategies to be done longer termed.

3.2 Risk Management of the Quality Framework

There are several different risks that quality review management faces.

The most important one is the adoption of the quality paradigm by the community. Math teachers are quite individualistic. For example almost none will use a resource “as is”, they will tweak it to their ways before using it, which is good. But some think that being a good teacher is a birth gift and can not be taught or improved upon. For these one, the concept of quality review is taken, whether from the reviewer or the from the author points of view, as a judgment on the personal merits of the author of the resources. So they will try to hurt the feeling of the author when reviewing, and they will be personally hurt by the bad reviews and will stop contributing resources. Some people are simply not ready for collaborative work and we hope that the INTERGEO project will help the concept of quality improvement diffuse in the community by giving a concrete example on how it helps people improve their teaching.

Another one is cheating or defacing: users reviewing in an exaggerate way, whether overrating or disparaging.

In order to deal with such abuses we propose, as future work, to imitate Slash-Dot.org's karma system explained in [Ben06, chapter 3] and [Mal07], to assign to each user a weight, small at the beginning, and evolving with the user's activity. Positive behaviour then increases this weight while negative behaviour diminishes it. The quality of a resource is then the weighted average of all the different reviews, the weight of the user taken into account. Therefore a highly trusted seasoned teacher will weight more than several newbies.

Positive behaviours:

- steady use of the platform, frequent connections and resource browsing,
- comments on forums (which are not reported as abusive),
- balanced reviews,
- having one's resource or review being copied,
- contributing well ranked resources.

Negative behaviours:

- comments on forums which are reported as abusive to the administrators (possible banishment),
- reviews that are systematically biased towards very bad or very good values,
- reviews that are made too fast, just minutes after discovering the resource and without having browsed all the associated material,
- contributing systematically badly ranked resources,
- breaking the user's chart like using several users for a single person.

This last item should be carefully monitored by trying to point inconsistent activities from the same ip address.

We expect that a lot of people will take pride in their online achievements, whether as the author of a famed resource, or as a reviewer with high rank. It is important that the consequence of each action is rendered visible to the user: she should see that her work is taken into account and slowly changes the world, one review, one resource, one forum post at a time.

Of course, not all users will be subject to this seasoning: special users, like didactic researchers, pedagogical experts, groups, will begin with a high weight from the start because their opinion and judgment are highly valued.

We trust that, consequently, very few people will try to malignantly interfere with our project, but counter measures such as the one described above, and new ones if need arises, should keep the project on course.

4 Acknowledgements

The authors would like to thank the e-Quality project [eQc06] for their impulse on the quality approach of the INTERGEO project. Although not fully implemented in this report, their vision has influenced this realization deeply.

5 Bibliography

References

- [Ben06] Yochai Benkler. *The Wealth of Networks*. Yale University Press, 2006. See http://www.benkler.org/wealth_of_networks/.
- [ELD08] Santiago Egido, Paul Libbrecht, and Michael Dietrich. Basic user manual. Deliverable D4.2, The Intergeo Consortium, August 2008. Accessible at http://www.inter2geo.eu/files/I2G_Basic-User-Manual.pdf.
- [eQc06] e-Quality consortium. <http://www.e-quality-eu.org/> *General Quality Process Charter*, 2006.
- [inc08] Curriki inc. The curriki review system. Available at <http://www.curriki.org/>, September 2008.
- [JTSL08] A. P. Jahn, Jana Trgalova, and Sophie Soury-Lavergne. Analyse de ressources pédagogiques et amélioration de leur qualité : le cas de la géométrie dynamique. In *2e Simpósio Internacional de Pesquisa em Educação Matemática (SIPEMAT)*, Recife (Brazil), 2008.
- [Mal07] Rob Malda. Slashdot's meta moderation. Web-page, September 2007.
- [Mas07] Vincent Massol. Xwiki: a platform for collaborative apps, December 2007. Available at <http://www.theserverside.com/tt/articles/article.tss?l=XWiki>.
- [mer08] Merlot peer review process, September 2008.
- [MSLT08] Christian Mercat, Sophie Soury-Lavergne, and Jana Trgalova. Deliverable d6.1: Quality assessment. Technical report, Intergeo Project, Available at http://www.inter2geo.eu/files/D6.1_060508.pdf, March 2008.
- [Pro08] Rice University The Connexions Project. Ways of collaborating in connexions, September 2008.
- [The04] The e-Quality consortium. <http://www.e-quality-eu.org/>, 2004.