

Activity regulation for the participative creation of data on-line

Christophe Courtin, Stéphane Talbot, Mathieu Mangeot

▶ **To cite this version:**

Christophe Courtin, Stéphane Talbot, Mathieu Mangeot. Activity regulation for the participative creation of data on-line. International Conference on Machine and Web Intelligence, Oct 2010, Algiers, Algeria, France. pp.132-139. hal-00968711

HAL Id: hal-00968711
https://hal.science/hal-00968711

Submitted on 1 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Activity regulation for building dictionaries on an on-line collaborative platform

Christophe Courtin
Universite de Savoie
73376 LE BOURGET DU LAC CEDEX
Email: Christophe.Courtin@univ-savoie.fr

Stephane Talbot
Universite de Savoie
73376 LE BOURGET DU LAC CEDEX
Email: Stephane.Talbot@univ-savoie.fr

Mathieu Mangeot
Universite de Savoie
73376 LE BOURGET DU LAC CEDEX
Email: Mathieu.Mangeot@univ-savoie.fr

***Abstract*—This paper discusses automatic regulation in participative Web systems. We present a generic solution with an original trace-centered approach. We describe an experiment with a general trace-based system (TBS) called CARTE (Collection, activity Analysis and Regulation based on Traces Enriched) featuring a regulation mechanism and we couple this system with an on-line generic platform for managing lexical resources called Jibikipedia.**

I. Introduction

Web2.0 is characterized by the Internet users' participative creation. Large-scale projects have highlighted the importance of the mode of participation which is extended to all Internet users, who are often anonymous, which thus allows the creation of consequent corpora of data (e.g. Wikipedia®in 2001, AgoraVox®in 2005 [10]), where information validity is subordinated to the self-regulation principle. At the same time, this principle intrinsically demonstrates weaknesses that are due to the lack of systematic information control. The checking of sources' quality being a universal criterion, we have observed abuses such as self-proclamation as an expert in a domain (example of identity forgery with the following hoax: a main contributor was a 24-year-old student who pretended to be a professor of religion at a Faculty of political science [1]).

The consideration of these abuses led to the emergence of new projects (Citizendium®[1] and Knol™in 2007 [2]), where Internet users must be identified and where thematic roles (e.g. author, collaborator) must be allocated to them. Actancial roles must be associated to these thematic roles, i.e. a set of predefined authorized actions (e.g. publish, modify with validation levels). This hierarchical organization of the roles led to the emergence of various groups of participants, who were more or less expert, and had greater or lesser power. Role allotment is based on activity and on notation of articles produced.

In the Knol™project of Google™, the validation of articles is carried out by contributors given the role of "questionable" (an expert's role). It is worth noting that the workload relative to the task of the role allocation can become very heavy (or time-consuming) given the sizeable increase in the number of participants (e.g. 800 authors in 2007 at Citizendium®, 1 million visitors a month and 5 to 10,000 regular writers a month in 2008 at AgoraVox®[3]).

This article is divided into three main sections. The basics of our original approach to answer the regulation problem in the participative systems is presented in the first section. An experiment with an existing participative system is described and the regulation mechanism is detailled with concrete examples in the second section. The expected results are covered in the third section. Several avenues of research will be discussed before concluding.

II. Trace-based approach

The approach we propose would not be original if it was only a matter of making the same report about regulation. Most of the time, a regulation mechanism is specific to a system and takes into account events which are generated inside it only. The originality of our system consists in using an external regulation mechanism which works with a trace-centered approach. The main advantages of such an approach are the possibility of linking events stemming from several software applications to be regulated (e.g. a chat room and a collaborative text editor) and that of changing a tool for another one that has similar functionalities with few modifications of the regulation mechanism (at the lowest level of abstraction, but not at the higher levels).

The coupling of the on-line generic platform for managing lexical resources (Jibikipedia) and the system based on trace interaction (CARTE) enables one to take into account automatically and dynamically the increase in the Internet users' participation. The activity trace analysis in the Jibiki platform [8] by the observation station CARTE allows retroaction operations to be performed automatically in order to allot roles.

The participative Web enables the production of very large data banks in a short period of time. This method of production of contents is very useful, and even indispensable, for under-resourced domains because of the lesser impact of the related systems.

The final research question that we are going to deal with in this article is: "How to take into account in the most automatic way the increase in the Internet users' participation?" The evolution of this research work consists in estimating the relevance of the regulation for the development of participative applications.

Unlike encyclopaedias, where contributors may describe very complex and detailed entries, it is more frustrating to contribute to dictionaries because of their restricted structure. Therefore, we consider the postulate that the participants' involvement in such a system is subordinated, among other factors, to the participants' consideration. From this, it follows that the participants' activity must be analyzed in order to calculate indicators about the participation's quality. Generally, these indicators are not intrinsically provided by the software applications themselves. Indeed, generated traces (e.g. log files) represent low level traces (e.g. "edit" event), which are not elaborate enough to recognize complex situations (e.g. the event "many contributions without correction in a limited amount of time").

III. Experiment

A. Description of the experiment

First and foremost, it is essential to begin by describing the observation objectives from which we built our experiment. Considering that some researchers would like to know if the activity regulation modifies the Internet users' perception of the IT tools, and if this new perception modifies their involvement in the activity, our main objective is to define how it is possible to take this participation into account automatically and dynamically.

In this experiment, the observers are researchers in computer science and in linguistics and also participants in the collaborative activity themselves. Therefore, observations have been made from activity traces having an abstraction level close to that of the observers. We designate as indicators the traces which may be directly interpreted by the observers. Generally, the software tools do not produce traces with a high enough abstraction level to build these indicators directly. In this experiment, we use two types of traces in a trace-based system (TBS) called CARTE (Collection, activity Analysis and Regulation based on Traces Enriched) [4] [6]:

1) interaction traces stemming from software tools which are used for the collaborative activities;
2) enriched traces, i.e. those originating from transformations of interaction traces or already enriched by means of use models of the software tools from which they arise.

In the following part, we will present the software tools of the experiment (see Figure 1) and then the indicators we wish to observe. Some of these indicators stem from the tools themselves and the others are generated by means of the observation station CARTE. This means that the indicators description must be defined by the observers before starting the collaborative activity. These indicators are integrated into the collaborative software tools and are activated via the retroaction mechanism of the observation station CARTE.

From our point of view as researchers in computer science, we have set up this experiment to test the mechanism of activity regulation in a context of participative production (Web 2.0). A first result will therefore be the expressiveness of the underlying models of the TBS (models of trace collection, use of tools and regulation of the activity) [7] [12]. Another result will consist in estimating the effects of the regulation [13] on the participants' perception of the collaborative software tools, which thus have augmented functionalities.

B. Description of the equiped platform

Our experiment takes place in a research project called MotMot [9] built on the Jibiki platform. In this project, there are two main challenges: to gather a community to produce dictionaries and to ensure the quality of the content.

In the first challenge, it is a matter of providing online generic tools to co-produce dictionaries. As far as the second challenge is concerned, famous projects such as Wikipedia®or AgoraVox®have tested the limits of self-regulation (e.g. hoax, identity forgery, relative errors, etc.). Conversely, the static definition of a committee of experts raises questions about the level of participation or about the variation of level of expertise over time. A dynamic definition would consist in setting different profiles in real time by considering several predefined criteria. This operation may become complicated (too many criteria) and time-consuming for a human team with a large set of participants.

In this context, we have developed a prototype called Jibikipedia in the Jibiki platform[11] implementing an automatic profiling process for participants. The participants receive thematic roles (contributor, reviewer, validator) in the editing interface of the Jibikipedia prototype according to their involvement and the quality of their production: a contribution consists in creating and editing an entry of a dictionary; a reviewing activity consists in verifying the exactness of an entry; a validation consists in storing a correct(ed) entry in a dictionary. In order to have a role attributed to them, the participants have to be registered in the Jibiki platform, i.e. they are not anonymous. Therefore, the various participants share the entries by modifying their status. During the contribution, the entry's status changes from "not finished" (editing) to "finished" (save). In the reviewing process, the entry may again be edited and saved to be corrected. At the end of the validation stage, the entry's status becomes "validated" and the entry will be effectively stored in the dictionary.

Figure 2 shows a French entry in the MotMot dictionary on the Jibiki platform. This entry has been reviewed by a level 3 user, thus the entry level has gone up to 3 stars.

C. Description of the indicators

According to the way a contribution has been accepted, i.e. with or without correction, the contributor's level of expertise may be respectively downgraded or upgraded. For example, each participant will be allocated a set of stars, the number of which corresponds to the level of expertise. When the contributor obtains three stars, s/he gets a second role: reviewer. Ditto with a third role from reviewer to validator. Conversely, s/he may lose stars and hence roles, back down to contributor. From the participants' point of view, these indicators are elements of awareness which are integrated

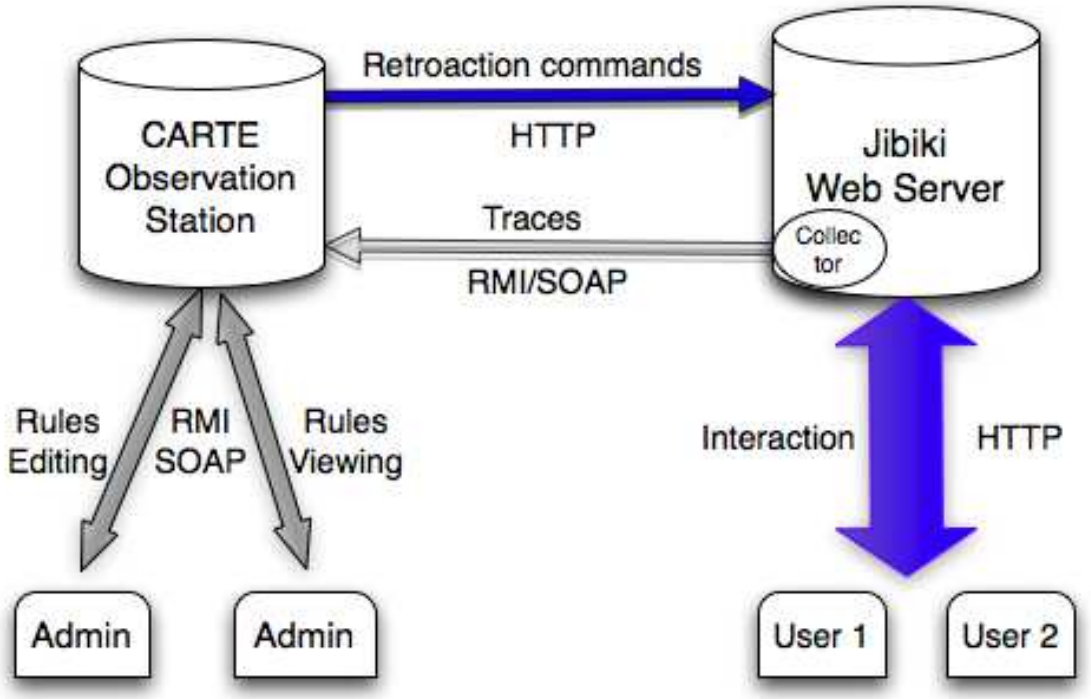


Fig. 1. Schema describing experiment



Fig. 2. 3-star rated entry on the Jibiki platform

directly into the graphic interface of the collaborative software tools.

We present below the list of the indicators implemented in the Jibikipedia system:

1) creation of n entries by X during m days: this indicator corresponds to the number of word creations that a contributor X has made in a limited period of time m (e.g. in days);
2) n proofreadings of A: this indicator means that the definition A has been read n times;
3) n proofreadings of A by X: this indicator means that the participant X has read the definition A n times, but s/he has not necessarily modified it;
4) n corrections of A: this indicator specifies that definition A has been modified n times;
5) n corrections of A by X: this indicator specifies that the participant X has modified the definition A n times;
6) creation of n entries in the domain D by X: this indicator specifies that the participant X has contributed n times in a particular domain D.

These indicators are represented by sequences and may be combined again, together or with other trace elements, in order to make more complex requests. For example, in the experiment, we consider that if a contributor produces three entries in a month, which are validated without modification, then her/his level of expertise is upgraded by one star, and at the end of three stars, s/he becomes a reviewer (or proofreader). In the following part, we will present a rule description with

our CARTE analyser which enables one to recognize certain situations of regulation.

D. Description of the rules

In this part, we will explain how to build indicators from generated and/or enriched traces of interaction. The first stage consists in defining the trace-based specification of the observation. As is explained above, the observation concerns the effects of the regulation of the activity in participative software tools. Regulation is a mechanism which transforms a system from an initial situation (or precondition) to a final one (postcondition or goal). The initial situation is represented by means of activity traces stemming from tools instrumentation [5] and traces enriched by the CARTE system. In this system, these traces are called signals and sequences, presented in Figure 3. The final situation is obtained automatically by means of the action part (or retro-action) in the rules of the observation station analyser.

The first and the last indicators can be broken down using the basic signals "create", "edit" and "save". The other indicators stem from the CARTE analyser.

In the experiment we used three sets of rules. The first one is used to interpret the actions of the users. Indeed the events collected from the collaborative entry editor, the Jibiki, at the first stage are quite basic:

- A user has created a new entry.
- S/He has started to edit an entry.
- S/He has saved some modifications to an entry.
- S/He has changed the status of the entry (for example from "not finished" to "finished") and so on.

However the actions, which really interest us are a little bit more complex. For example the editing of an entry should be associated with a set of related events (the user loads the entry, starts editing it, then saves the modifications). The proposal of a new entry implies its initial creation, then a sequence of edits that conclude with the modification of the entry status (the entry is considered as proposed when its status is changed from "not finished" to "finished"). In the same way we must distinguish entries reviewed and validated without further modification from the others. The first set of rules is used to identify those relevant sequences of events : first the edits, then the entry proposals, which are reviewed and validated with or without further modifications.

The second set of rules (see Figure 7) is used to compute the indicators, in our case the star numbers awarded to each element : the entries and the users. Since this process depends on the number of entries contributed, modified and validated by each user in the previous periods, it implies using the sequences identified via the first set of rules. It also depends on itself: highly-rated users are supposed, without further evidence to the contrary, to produce high quality entries, and users who produce highly-rated entries should get good evaluations.

Finally the third and last group of rules is used to modify the statuses of users. Highly-rated users will be able to review, validate or invalidate the entries contributed by the other users. In contrast, new or poorly-rated users will have to work with the "expert" ones and improve their skill before being able to contribute new entries without supervision.

All the rules are built from the traces with an ad-hoc editor, see Figure 4.

IV. Results

We have highlighted above the difficulty in obtaining Internet users' participation in under-resourced domains because of the lesser impact of the related systems. We postulate that if there is a regulation mechanism which provides the possibility of adapting the users' profile to their own activity, we expect that these users will have the feeling that their activity will be more highly thought-of by the community of participants. We thus consider that regulation improves interactions with the computing environment and can have an influence on the users' participation.

However, a computer-based study alone does not allow us to evaluate regulation's effects on participation quality. Therefore, the main research objectives of this article consist in transforming information which derives from computing traces of collaborative activities in order for it to be used for the regulation mechanism, and in testing our trace-centered model's strength. The experiment has revealed the regulation model's capacity for specifying the authorized actions of the asynchronous collaborative editing session in progress. In order to present the results, we propose to describe the regulation mechanism and to present the indicators generated.

In this part, we present an instrumentation technique which takes place at the level of the collaborative software tools themselves and then close to their use model. The definition of the indicators is facilitated by the structuring and the explicitation of the collected elements and their interpretation requires the definition of a use model of the software tools which make up the participative work area. The collect instrumentation enables the gathering of trace elements in order for them to be analyzed by an observation station (see Figure 6). The latter is an external system which transforms the traces and re-injects actions (postcondition) in the Jibikipedia prototype, by means of retroaction instrumentation, in order to modify participants' roles for example (see Figure 1).

It should be recalled that the use model enables the analysis of collected traces and is based on how the concerned software tools are supposed to be used. The analyser is composed of a set of rules of transformation (see Figure 4) with logical operators (AND, OR, NOT), a temporal relation (THEN) and priority relations (brackets). We present partially in Figure 5, with the XML format (DTD), the rule description of the experiment.

In Figure 7, we present an instance of a rule based on the previous DTD. This rule specifies the following indicator: "5 entries have been created by a contributor in 1 month".

V. Conclusion and perspectives

In this article, we have presented how to adapt the users' profile and then the functionalities of a Web-based participative system called Jibikipedia [8], by means of an external

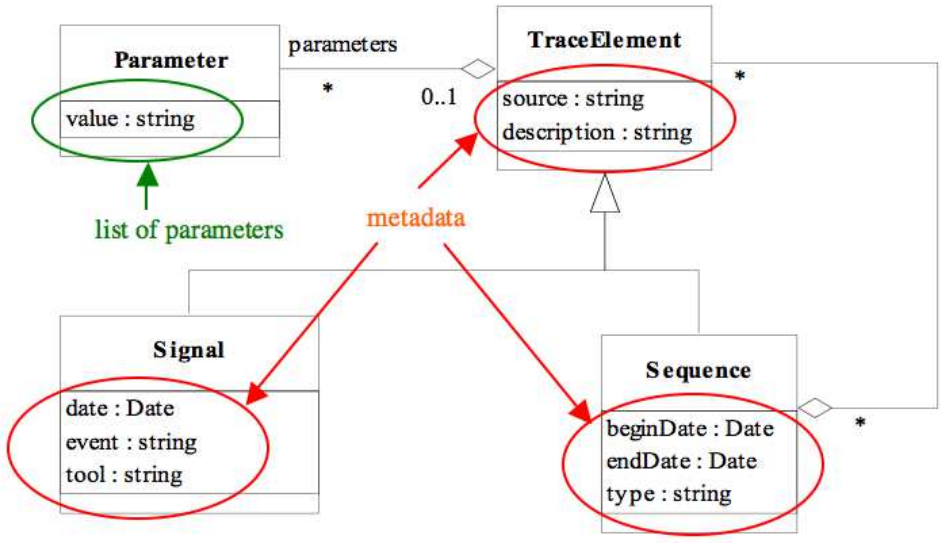


Fig. 3. trace format

trace-based system called CARTE [4]. We have postulated that this activity regulation could have an influence on the users' participation. Therefore, we have worked on how to operationalize our regulation mechanism in the project about multilingual dictionary asynchronous co-construction. Our main objectives have consisted in testing both the relevance of the coupling of the two systems and the corresponding models' strength. In particular, we have described an example to take into account in the most automatic way the increase in the Internet users' participation. An important first stage of this work has consisted in defining relevant indicators.

The perspectives of such research work consist in plugging our regulation mechanism into other participative systems and in measuring the effective impact of regulation on the Internet users' participation, for example, in multilingual dictionaries and especially for under-resourced domains.

REFERENCES

[1] Citizendium inaugure son alternative à wikipedia, March 2007. http://www.lemondeinformatique.fr/actualites/lire-citizendium-inaugure-son-alternative-a-wikipedia-22465.html.

[2] Encyclopedia knol (google), May 2009. http://knol.google.com/k.

[3] Jérôme Colombain. Agoravox devient une fondation, 21st May 2008. http://www.france-info.com/chroniques-nouveau-monde-2008-01-31-agoravox-devient-une-fondation-79828-81-109.html.

[4] Christophe Courtin. Carte: an observation station to regulate activity in a learning context. In *IADIS / CELDA*, pages 191–197, Freiburg, Germany, 13-15 October 2008.

[5] Christophe Courtin and Stéphane Talbot. an architecture to record traces in instrumented collaborative learning environments. In *IADIS / CELDA*, pages 301–308, Porto, Portugal, 14-16 December 2005.

[6] Christophe Courtin and Stéphane Talbot. Automatic analysis assistant for studies of computer-supported human interactions. In *EC-TEL*, pages 572–583, Nice, France, 29th September - 2nd October 2009.

[7] Julien Laflaquière, Lotfi Settouti, Yannick Prié, and Alain Mille. A trace-based system framework for experience management and engineering. In Springer, editor, *Second International Workshop on Experience Management and Engineering*, pages 1171–1178, Bournemouth, UK, 9-11 October 2006.

[8] Mathieu Mangeot and Antoine Chalvin. Dictionary building with the jibiki platform: the gdef case. In *LREC 2006*, pages 1666–1669, Genoa, Italy, 21-26 May 2006.

[9] Mathieu Mangeot and Hong-Thai Nguyen. Building lexical resources: towards programmable contributive platforms. In *IEEE-RIVF 2009*, volume 1/1, pages 84–92, DaNang, Vietnam, 14-16 July 2009.

[10] Miléna Nemec-Poncik. Wikipédia : à consommer avec discernement, November 2007. http://www.lemondeinformatique.fr/actualites/lire-wikipedia-a-consommer-avec-discernement-24498.html.

[11] Gilles Sérasset and Mathieu Mangeot. Papillon lexical database project: Monolingual dictionaries and interlingual links. In *NLPRS-2001*, pages 119–125, Tokyo, 27-30 November 2001.

[12] Lotfi Settouti, Yannick Prié, Pierre-Antoine Champin, Jean-Charles Marty, and Alain Mille. A trace-based systems framework : Models, languages and semantics. http://hal.archives-ouvertes.fr/inria-00363260/fr/, LIRIS, Lyon, France, 29th May 2009.

[13] Stéphane Talbot and Philippe Pernelle. Helping in collaborative activity regulation: modeling regulation scenarii. In *IHM 2003: Proceedings of the 15th French-speaking conference on human-computer interaction on 15eme Conference Francophone sur l'Interaction Homme-Machine*, pages 158–165, New York, NY, USA, 25-28 November 2003. ACM.

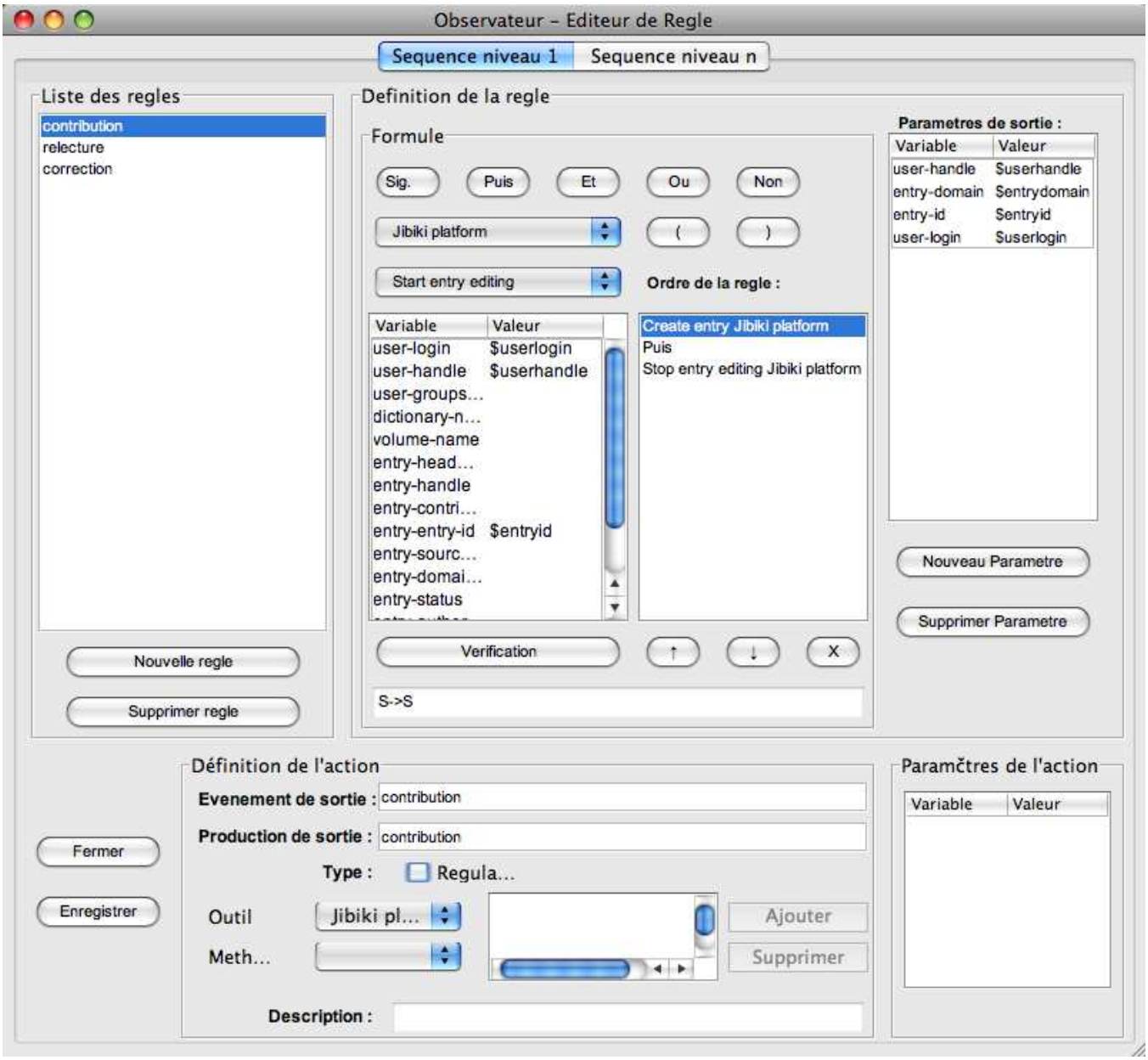


Fig. 4. Screenshot of the Rule Editor

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" version="1.0">
   <xs:element name="rules" type="listOfRules"/>
   <xs:complexType name="listOfRules"/>
      <xs:sequence>
         <xs:element minOccurs="0" name="rule" type="rule"/>
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="rule">
      <xs:sequence>
         <xs:element minOccurs="0" name="description" type="xs:string"/>
         <xs:element minOccurs="0" name="conditions" type="listOfConditions"/>
         <xs:element minOccurs="0" name="retroActions" type="listOfRetroactions"/>
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="listOfConditions">
      <xs:sequence>
         <xs:element minOccurs="0" name="condition" type="condition"/>
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="condition">
      <xs:sequence>
         <xs:element maxOccurs="1" minOccurs="1" name="action" type="action"/>
         <xs:element maxOccurs="1" minOccurs="1" name="numberOfResults" type="xs:int"/>
         <xs:element maxOccurs="1" minOccurs="1" name="numberOfResultsSemantic"  type="xs:string"/>
         <xs:element maxOccurs="1" minOccurs="1" name="subCondition" type="listOfConditions"/>
         <xs:element maxOccurs="1" minOccurs="1" name="parameters" type="listOfParameters"/>
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="action">
      <xs:sequence>
      ...
      </xs:sequence>
   </xs:complexType>
   <xs:complexType name="listOfParameters">
      <xs:sequence>
      ...
      </xs:sequence>
   </xs:complexType>
   ...
```

Fig. 5. Rules description

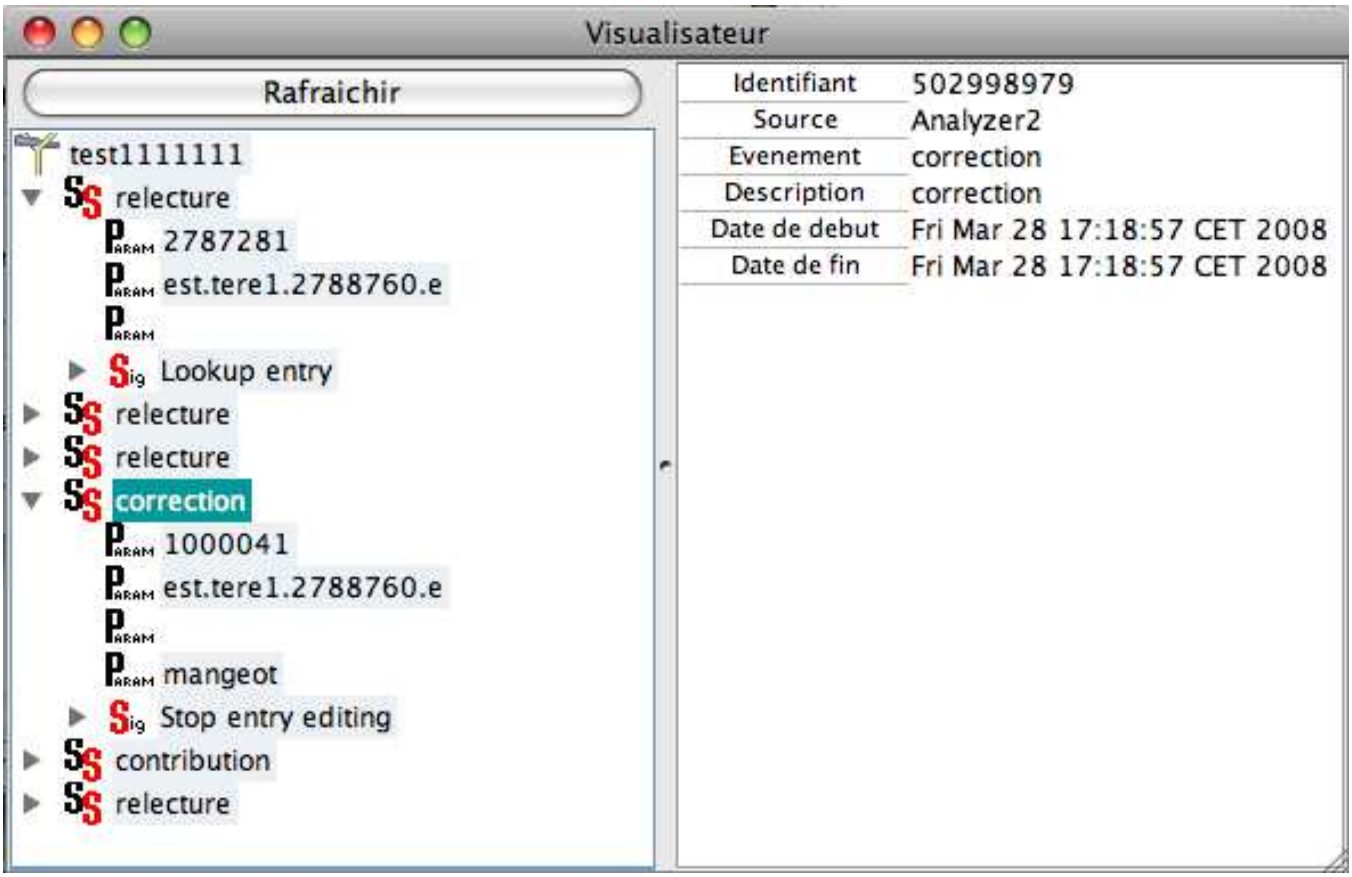


Fig. 6. Visualizer of the collected traces

```
<rules>
   <rule>
      <description> the first rule </description>
      <conditions>
         <condition>
            <action> contribution </action>
            <numbersOfResults> 5 </numbersOfResults>
            <parameters>
               <parameter>
                  <id> month </id>
                  <value> 1 </value>
               </parameter>
            </parameters>
         </condition>
      </conditions>
   </rule>
<rules>
```

Fig. 7. Example of rule