Towards a Framework for a validated content management on the collaborative Web: Blogs case

A. ABTOY, N. AKNIN, B. SBIHI, A. EL MOUSSAOUI and K.E. EL KADIRI

Faculty of Sciences, Abdelmalek Essaadi University
Tetuan, Morocco

Abstract

Since 2004, the web 2.0 has been based on the contributions and the interaction of internet users to easily produce and publish the content. This type of philosophy which has been approved by the web community encourages users to participate and exchange, has resulted in a huge mass of diverse information. However, this philosophy based on the collaboration of users brings many problems such as the information overload, the lack of information quality's control, the search results accuracy and possibilities of sorting and classifying by the level of relevancy. This paper is a part of a work research conducted in our laboratory that focuses on the quality of information on the collaborative web and its related issues. We propose a content management framework that will handle the content during its lifetime on the web. Our approach reuses the static validation mechanism of content and adds the concept of dynamic validation mechanism performed by internet user's community. In this context, we will model a validated blog's system management based on our approach using UML standard.

Keywords: *Collaborative web; content management; blogs; quality of content.*

1. Introduction

The term Web 2.0 was introduced in October 2004 by Tim O'Reilly [1], since that date, web 2.0 has been considered as the second phase of the web's evolution. The new philosophy of this web has revolutionized the use of online services comparing to the previous ones. The fundamental concept of web 2.0 is based on the involvement of users in the process of production. The internet users are considered as active co-producers rather than passive consumers which increases the mass of data produced and stored on the Web [2].

According to Tim O'Reilly [1], Web 2.0 brings new features to the Web thanks to the seven concepts which are:

- The Web is a services platform.
- The power of collective intelligence should be exploited efficiently.
- Data is the next Intel inside.
- End of the software release cycle.
- Lightweight programming models
- Software above the level of a single device.
- Rich user experiences.

Web 2.0 is both a usage and a technology paradigm .Web 2.0 gains its success and popularity through its applications and tools which are based on the seven principles. These applications are characterized by a rich responsive user interfaces, dynamic contents and services interoperability. The main tools of collaborative web are represented in the following table:

Table 1: Web 2.0 tools and their utility [3]

Tool	***Utility***
Blog	Publish and share content
Wiki	Collective Intelligence
Social network	Create online communities
RSS	Information awareness
Tag	Customize and improve research
Podcast and Videocast	Share audio and video contents

Even though web 2.0 brings so many improvements to the internet world; the current Web has several limitations which represent obstacles to its development [4]:

- Information quality: The current web does not insure or guaranty the quality of content produced and published on the web.
- Information obesity: The huge amount of content produced makes the search for trustworthy information a very challenging task; also the content management becomes more and more difficult.
- Low rate of participants: Even with the simplicity of publishing, editing and participating on the Web 2.0, the number of active internet users remains very low compared to those non-productive.
- Information lifetime: The massive production of content makes the content lifetime very short. In the case of blogs, content is stored and browsed in a reverse chronological order which shows content by publication date instead of relevancy.

- Security and copyright: The content published on the web is not secure by access rights; also Web 2.0 cannot offer copyright to content that needs it.
- Lack of semantics: the search for relevant information requires semantics to give more sense to data and to avoid outdated erroneous search results.

The massive production of content in the web has made its management very difficult and complex [5]. The problem of the information quality of content represents a real challenge for the current Web [6]. This subject has been ignored in the web development and there are a few researches addressing this issue in a sufficient precise way in the current Web. In this paper we propose an approach of a Framework that supports the quality-relevancy aspect of the validated information on the web.

In the following paragraph, we will define the content validation mechanisms, and then we will detail the two types of this mechanism: static and dynamic. Then we will propose some principles and concepts to integrate with the content management Framework on the Web to resolve and fix its limitations. In the next paragraph, we present a validated blog management's system based on our approach using a UML diagrams to demonstrate the use of the validation mechanisms. Finally we will give an overall conclusion which will present a set of perspectives.

2. Content Validation

Our team presented a series of a work research that proposes a new version of the Web that solves the limitation and the problems of the current one. The global idea is to publish valid and pertinent information on the net with a measurable quality that can be controlled. This Web (called Web 2.2) [7] is based on the improvement that will be done at both levels: users and information.

- User's level: The community of internet users will be organized in four types of user's categories; these ones are formed by new types: users, producers, validators and experts. Each one of them offers both new functions and privileges within the web. The table below details the different categories of the users:

Table 2: The categories of users and their roles on the web [7]

Actor	***Role***
User	Read content
Producer	Read and produce content
Validator	Validate the produced content
Expert	Monitor and verify the published content

- Information's level: This web proposes a subdivision of the information in several categories; each category represents a content quality in order to simplify the content classification and management. The table below summarizes the different categories of the levels of information's quality:

Table 3: The subdivision of information into categories on the web [7]

Symbol	***quality***	***meaning***
B	Good	Validated and relevant
M	Medium	Validated with an average degree
F	Poor	Validated subject that needs correction
E	Erroneous	Not validated
C	Opinion or comment	

2.1 Static Validation

Static validation is mainly based on both users and information quality classification. Basically, this validation process is performed by authoritarian internet users who are members of the validation committee on the web. The producer is the user who participates in the production of the information/content. Depending on participation rate, users will have access to the validated content and its related services and applications. This philosophy encourages users to change their attitude of passive consumption and contribute in the collective intelligence. The validator decides whether the content published by the producers is relevant and pertinent or not, he will assign a score to describe the content's quality. In other words, validators regulate and control the quality of information. So that content can be validated, it takes two validators elected by an expert to evaluate it. Finally, the expert's role is the monitoring of the validation process besides the election of validators. Fig.1 below represents the different steps of the static validation process:

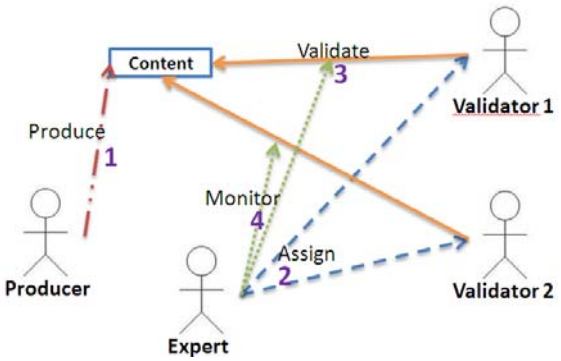

Fig. 1 The Process Of Static Validation.

The blog is a very powerful tool to publish ideas and to share knowledge on the web [8], on the other side; a published content suffers from a major disability: absence of relevancy and lack of information quality control. The static validation allows the classification of blog's contents (tickets) which makes web search easier and eliminates redundant information and poor quality content [9].

The major disadvantage of static validation is that the validated content will keep a static score during its lifetime. The evaluation performed to the content does not take into account the opinion of the rest of the Internet community.

2.2 Dynamic Validation

The dynamic validation works by monitoring the quality of the content during its lifetime on the web. In the case of blogs, the validation committee gives its opinion towards the quality of the information, this value (score) presents the first evaluation of the content performed by the validators. Later the users attribute values in favor of the ticket. This continuous assessment provides a dynamic validation during the whole content life time.

While the navigation on the web, the obesity and the overload information has composed a big problem for the users, in the dynamic validation a decision mechanism is proposed to compare the content quality to a threshold. For example, the validation committee decided that a ticket has average quality content. When the content is published on the web, it begins to accumulate comments and bad scores. Thanks to the dynamic validation, its quality takes into account the opinion of the community of users and when it becomes lower to a certain threshold it will be archived or eliminated automatically. Fig.2 illustrates the concept of dynamic validation of content during its lifetime on the web.

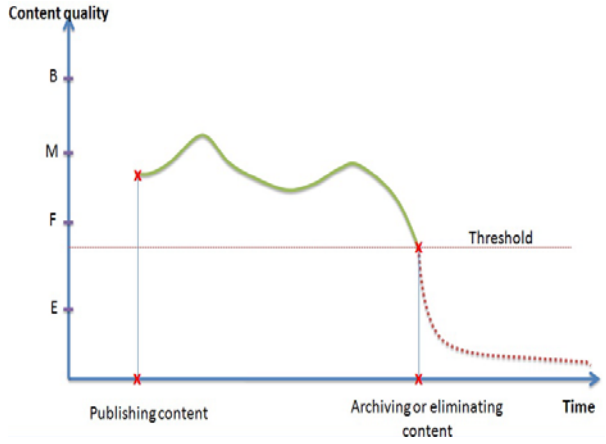

Fig. 2 Dynamic validation during content's lifetime

Threshold is defined by the validation committee, and it varies depending on the theme and the subject. For example, where the blogs which have a scientific aspect will have a high threshold, the social ones will have a low one. The excess of the threshold will trigger the archiving or the eliminating of the content depending on the nature of the blog and its policy regarding the quality's degradation.

3. Content Management Framework

Our approach intends to develop a Framework that handles the content and not the container. We have used in our approach the layers model concept because it allows the functional division of services in a way that make them complementary to each other. Also, we based our approach on the concept of information packet. We assume that instead of exchanging and storing raw content, we can add a header and a tail to the content in order to consider it as a logical packet of information in analogy with the IP packet [10]. According to our approach, every layer offers some features distinct from each other. In our case, this framework model well separates the management of content and container.

Fig.3 below shows the relationship between content and container in both actual web and our approach. Content will be encapsulated with the new header and tail instead of a raw content.

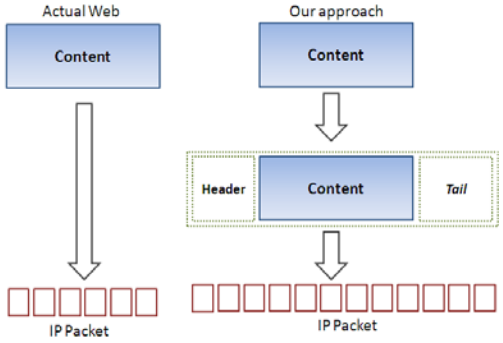

Fig. 3 The relationship between content and container

3.1 Information Packet

The information packet contains sections used to manage and deal with the content on Web 2.0. The sections are divided between a header and a tail depending on the functionality of each section [9]. The information packet has the following structure:

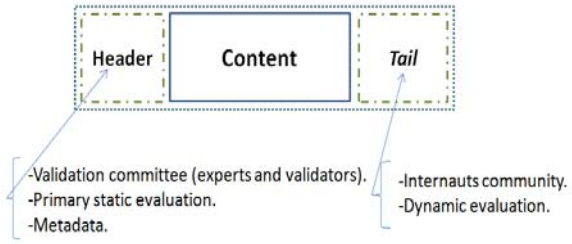

Fig. 4 Structure of information packet

The header contains essential information to identify content on the web. This information will be added during the static validation process. The header sections guarantee the following functionalities:

- Content uniqueness: this section contains the digital signature of the content. The content, through information packet, will be identified on the net. Thanks to this functionality it will be possible to eliminate overload and redundant content on the Web.
- Security and copyright: crypting content before publishing on the net in the case of pay services. The access rights to contents can be easily managed through this section. For example, violent content can be banned to underage internet users.
- Content integrity: to verify content integrity on the reception after being transported by internet protocol. This option is very useful because content can be altered and this measure will prevent data of being tampered.
- Validity of time: this section contains content expiry date which handles the management of the validated content during its lifetime; this functionality is inspired from TTL field IP protocol [11].
- User's identity: in order to participate in this web, users should be identified: producers, co-producers, validators and experts. Even with the identification of users, they will be able to surf anonymously.
- Content type: this section improves and refines search criteria in order to get the best pertinent search results based on content type (Blog, Podcast, Videocast, Wiki…)
- Content quality (static validation): contains a rating score assigned by validators. This information will be combined with dynamic validation score to obtain a final score.
- Extensibility: to support the fast development of web 2.0 applications, the information packet header can integrate customized functionality depending on the application specification.

The information packet tail includes dynamic validation data. This part will be calculated continuously to have an evaluation to the content quality. The header data will be set statically throughout all content lifetime on the web, on the other side; the tail is going to be updated in every interaction with the published content.

3.2 Implementation into Layers Architecture

The content management framework's implementation will be based on the layer model. Fig.5 represents this architecture:

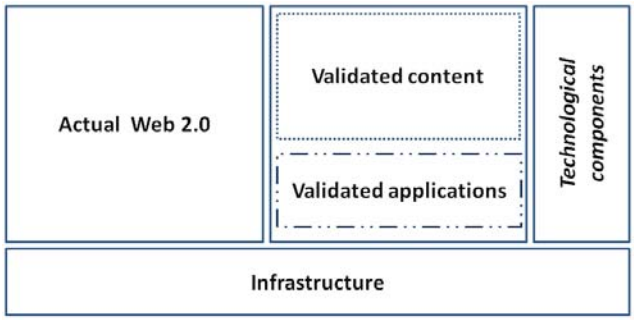

Fig. 5 The Framework's implementation into the layers stack

- Infrastructure layer: This layer represents all underlying infrastructure of Web services. This infrastructure includes: communications networks (wired, optical fiber, satellite …), hardware infrastructure (routers, servers, machines …) and communications protocols (TCP/IP, IPv6 [12]…).
- Actual Web 2.0 layer: includes all applications and contents published on actual Web 2.0.
- Validated applications layer: This layer represents the web applications that support the validated content management framework.
- Validated content layer: represents the validated content produced by the validated application layer and processed by the validation committee. The validated applications layer manages the information published within this layer and guarantees a validated pertinent content that can be easily found.
- Technological components layer: includes all the technologies that incorporate web 2.0 [13]: Ajax, Widgets, Mashups, APIs…

4. Case Study: Case of Blogs

4.1 Blogs and the Collaborative Web

Blogs are considered as one of the most important web 2.0 applications. In the last few years, the Blogosphere increased exponentially .According to *Blogpulse*, there are more than 148 million blogs in the world in late 2010, publishing 1 million posts per day [14].the growth of the Blog's popularity among internet surfers due to the fact that it doesn't need any technical background to create, publishes or uses a blog. *Technorati* claims in its annual report of state of the blogosphere of 2009 that 59% of bloggers use a free third party hosting service to create their blogs [15].

The blogging rises as an emergent way of publishing contents and information over the web. The number of people who rely on the web to seek information has increased dramatically over the last few years, which makes blogs a competitive source of information regarding the traditional media [16]. Blogs also are used for educational proposes such as in e-learning [17] or combined with other web 2.0 tools to build a collaborative e-learning approach [18].

4.2 Validated Blogs

Our goal is to design and develop a blogs that uses the concept of combined static-dynamic validation approach to resolve the blog's content relevancy and pertinence issue. We presented the concept Vbloges [9] and the design of the proposed blog's system is described through several UML diagrams in order to complete our recent work.

The use case diagram contains a several use cases that were identified from the needs analysis. Fig.6 represents the use case diagram of the blog's system. The primary actors that suppose to interact with the blog are: the users, the producers, the validators and the experts. An administrator is required to perform administration and supervising tasks .The blog system offers the possibility to login into the blog system using a username and password or a digital fingerprint. The blog can use an internal authentication or an external one so that the users can participate easily in the blogosphere with the same profile. In the case of biometric authentication, a device is needed in order to scan and identify the user's fingerprint.

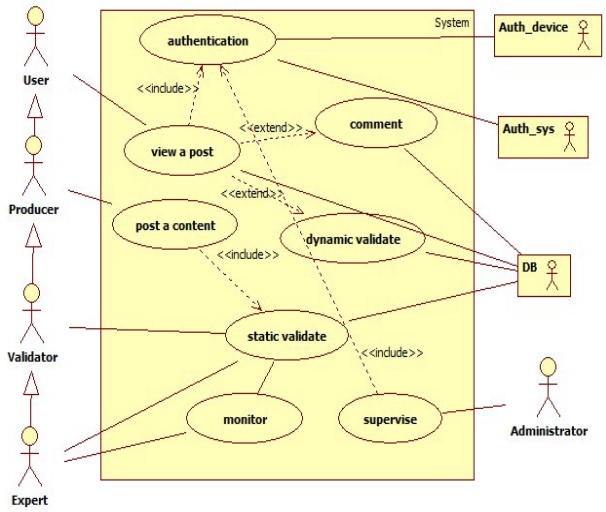

Fig. 6 Use case diagram

The structural aspect of our system is presented through a class diagram in Fig.7. Every blog is identified in the blogosphere with its own and unique ID. Thus we used *blogID* attribute in *Blog_system* class.

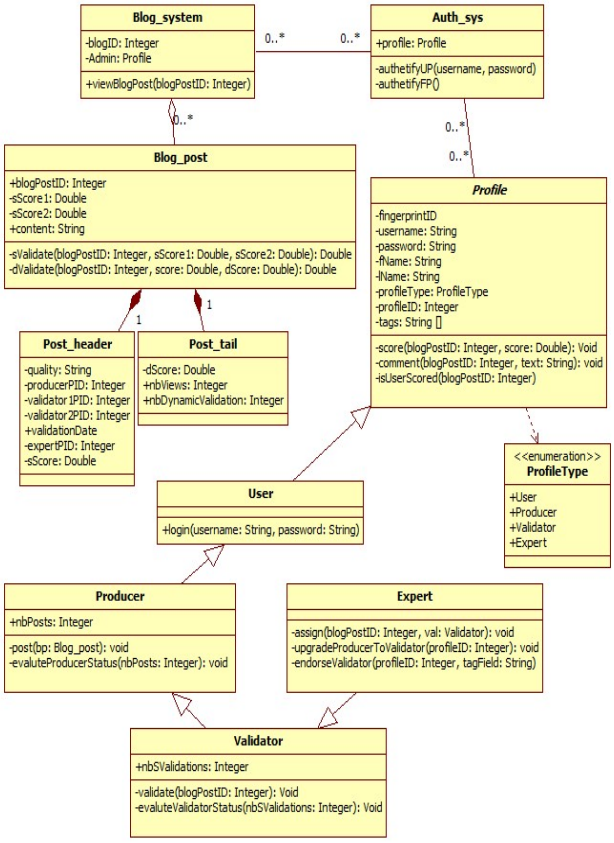

Fig. 7 Class diagram

According to our approach, the structure of the blog post is a composition of a header and a tail which we modeled using composition relationships between *blog_post* and the two classes: *Post_header* and *Post_tail*. The *Post_header* contains the information about the static validation process such as the following attributes:

- *quality*: post's quality
- *sScore*: the static score which is the combination of the two scores given by validators.
- *producerPID*, *validator1PID*, *validator2PID* and *expertPID*: contains the profileIDs of the users who participates in the production and the validation of the blog post.

The *Post_tail* contains all the details about dynamic validation including:

- *nbViews*: the times that this post has been viewed.
- *nbDynamicValidation*: the times that this post has been dynamically validated . The two attributes (*nbViews* and *nbDynamicValidation*) are useful to have an idea about the rate of user's participation on the validation process.
- *dScore*: every time a user gives his opinion on the blog post, this value is processed in order to obtain the new value. Each user has an influential factor that will be combined with his score, because for example we should not give the same weight of a score giving by an ordinary user and an expert. The following table represents the score's weight that we call the influential factors of users according to our approach:

Table 4: The users profile and their influential factors

Profile	User/reader	Validator	Expert
Influential Factor InFa	25	50	100

We note that the producer of the blog post cannot participate in the validation process: neither the static nor the dynamic one.
The dynamic validation score S_d of a post:

$$S_d = \frac{\sum_{k=1}^{n} S_{u(k)}.InFa_k}{\sum_{k=1}^{n} InFa_k} \qquad (1)$$

Where $S_{u(k)}$ is the score giving by the user *k* and $InFa_k$ his influential factor.

The abstract class *Profile* contains the most common attributes needed by other classes of users such as: *profileID*, *username*, *password*… The *tags* attribute will be useful for experts to help them assigning validators interested with the blog post's subject in order to validate it.

The blogosphere is the community made up of blogs. In this context, the blog system can use its inner authentication system or use an external one that belongs to another system. This part is modeled by using a multiplicity between *Blog_system* and *Auth_sys* in the class diagram.

Fig.8 shows a simplified sequence diagram of the authentication process. We take into account both possible authentication methods: using a biometric fingerprint, in this case an authentication device will be needed. The second method is the classical login using a username and a password.

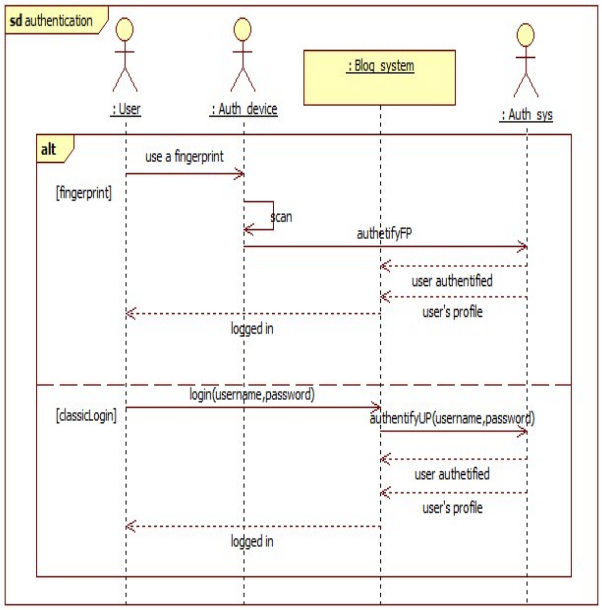

Fig. 8 Authentication sequence diagram

The users can easily post in the system after logged in. This operation is explained in the sequence diagram of posting a content Fig.9. The user in this case is the author or the producer of this blog post

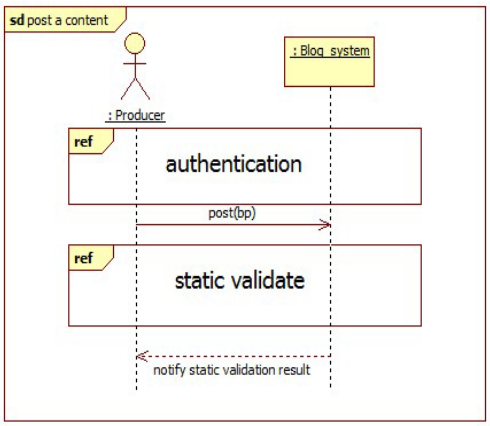

Fig. 9 Post a content sequence diagram

After posting the content, the blog post remains unvalidated and unpublished; therefore it needs to be statically validated by the validation board which is composed by an expert and two validators. The static validation sequence diagram in Fig.10 represents the interaction between actors and the system in the static validation process.

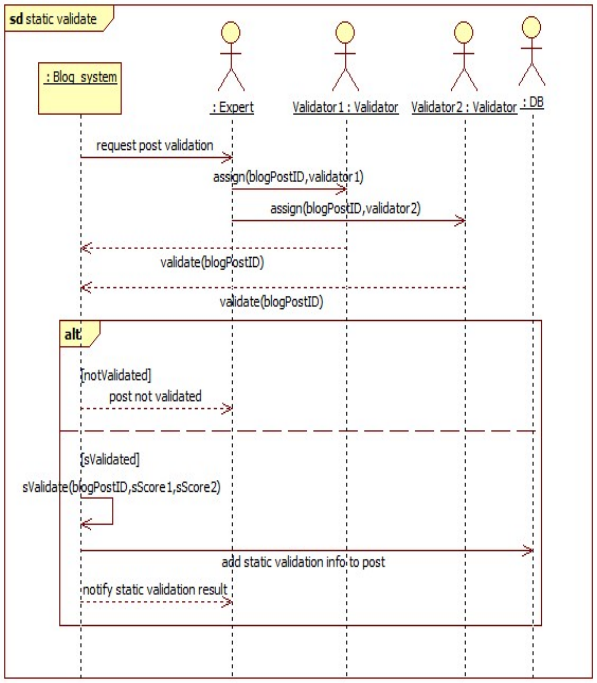

Fig. 10 Static validation sequence diagram

Once the blog receives a post, it requests form an expert a static validation. Two validators will be assigned by this expert based on their *tags* (validators in the same field of the post's subject). The decision of the two validators will determine if this content will be published (content validated) or not (remains unpublished and unvalidated). In this case of positive decision, each validator gives a score and a brief description about the content's quality. As a result, the blog system will process the scores in order to obtain the score of static validation via *sValidate* method which needs as arguments *blogPostID*, *sScore1* (validator 1's score) and *sScore2* (validator 2's score). This process offers an initial reviewing for the content's quality.

After being statically validated, the blog post will be available (validated and published) for other users to read it. The post viewing sequence diagram in Fig.11 includes all the possible use cases related to viewing a post.

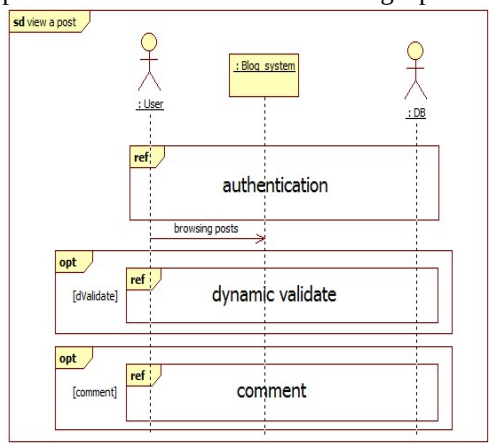

Fig. 11 View a post sequence diagram

While reading or viewing the post, users can comment or participate in the validation process. The sequence diagram of the dynamic validation is represented in Fig.12.

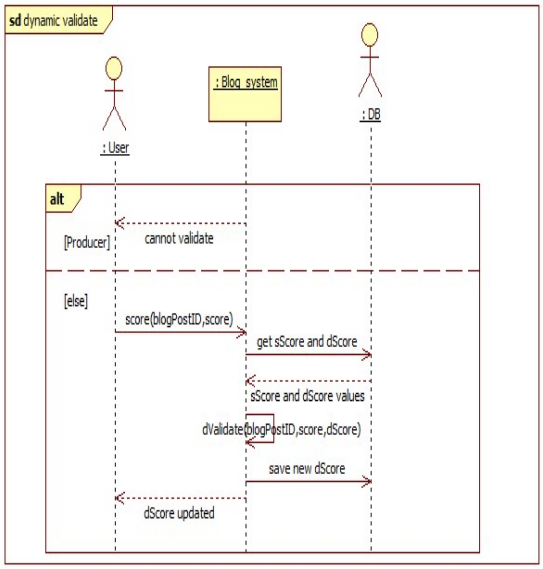

Fig. 12 Dynamic validation sequence diagram

The producer or the author of the post cannot participate in the dynamic validation in order to keep the objectivity of the content's quality. Once the user scores the post, the blog system evaluates the new dynamic score by *dValidate* method. The dynamic score contains the users's opinions about the post's quality.

The content produced in the *blog* goes through several states. Fig.13 represents the content lifecycle and its different states in the blog system:

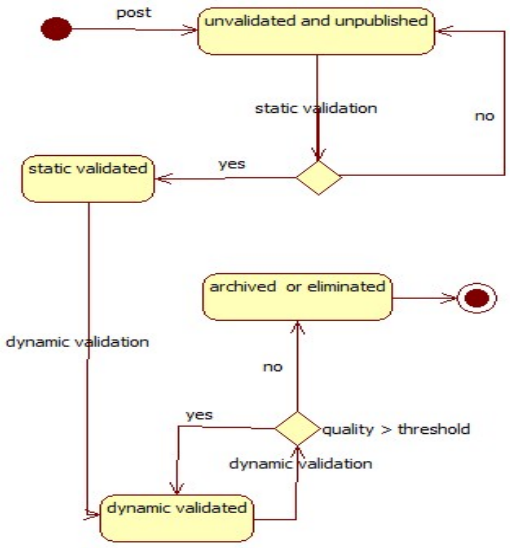

Fig. 13 Content lifecycle

Once the user produced a post, its remains unvalidated and unpublished unless booth validators approved and validated it. After that, the content considered is statically validated and published in the blog. By then, users have access to view the post and give their opinion on its quality by the dynamic validation process. Whenever a user validates and gives a score to the content, the system evaluate the new dynamic score and compares the new value and the threshold set by the validation board. If it's always greater than the threshold, the blog keep publishing the content. Otherwise the system will archive or eliminate the content depending on the policy of the validation board of this blog.

5. Conclusion

Unlike the web 1.0, the collaborative web or else the web 2.0 involves the users on the information production. As a result, the web has grown up in a huge way raising many problems such as: the information obesity, the lack of the content quality and the relevant research. The purpose of our approach focuses on handling these problems by proposing a framework that supports the content to provide an effective, organized and structured one. The information quality is going to be monitored during time using combination of the static and the dynamic validation. The structuring of the information in logical packets provides additional features to manage the content on the web.

As an application to our approach, we propose to develop and extend by integrating the concept of the validated content management framework to existing tools of web 2.0 such as Wikis and social networks in order to reach a validated web.

References

[1] T. O'Reilly, "what Is Web 2.0", 2005, Available: http://oreilly.com/web2/archive/what-is-web-20.html, accessed 2 March 2010.

[2] G. Vickery and S, Sacha Wunsch-Vincent, Participative Web And User-Created Content: Web 2.0 Wikis and Social Networking, Organization for Economic Publisher, October 2007.

[3] J.F Gervais, Web 2.0 : les internautes au pouvoir , Edition Dunod, 2007.

[4] B. Sbihi, "Web 2+: Vers une nouvelle version du web 2.0", Journal of Information and Communication Technologies, Vol. 35, No. 2, 2009, pp. 12-24. Available: http://isdm.univtln.fr/PDF/isdm35/Sbihi_isdm35.pdf.

[5] B. Boiko, , Content Management Bible, 2nd edition, Wiley, 2005.

[6] G. Pullman, Content Management: Bridging the gap between theory and practice, Baywood Publishing Company, 2008.

[7] B. Sbihi, and K.E. El Kadiri, "Web 2.2: Toward classified information on the Web", International Journal of Web Applications, Vol. 1, No. 2, June 2009, pp. 102- 109.

[8] B. Walsh, Clear Blogging: How people blogging are changing the world and how you can join them, Apress, 2009.

[9] B. Sbihi, K.E. El Kadiri, and N. Aknin, "The Vblogs: Towards a New Generation of Blogs", International Journal of Computer Science Issues, Vol. 7, issue 3, No. 2, May 2010, pp. 9-14.

[10] G.C. Vinton, and E.K. Robert, "A Protocol for Packet Network Intercommunication", IEEE Transactions on Communications, Vol. 22, No. 5, May 1974.

[11] G. Pujolle, O. Salvatori, and J. Nozick, Les réseaux, Eyrolles, 2008.

[12] J. Davies, Understanding IPv6, Second Edition, Microsoft Press, 2008.

[13] M. Norman, and C. Heilmann, Web Development Solutions: Ajax, APIs, Libraries, and Hosted Services Made Easy, Friends of ED Publisher, 2007.

[14] Blogpulse Stats, http://www.blogpulse.com/ [visited on 21/10/2010]

[15] Technorati, State of the Blogosphere 2009, http://technorati.com/blogging/feature/state-of-the-blogosphere-2009/ [visited on 21/0/2010]

[16] T. J. Johnson, B. K. Kaye, S. L. Bichard, W. J. Wong, "Every Blog Has Its Day: Politically-interested Internet Users Perceptions of Blog Credibility", Journal of Computer-Mediated Communication, Vol. 13, Issue 1, October 2008, pp. 100–122.

[17] Wei-Jane Lin, Hsiu-Ping Yueh, Yi-Ling Liu, Masayuki Murakami, Koh Kakusho, Michihiko Minoh, "Blog as a Tool to Develop e-Learning Experience in an International Distance Course", Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06), 2006, pp. 290-292.

[18] B. Sbihi, K. E. El Kadiri, "Towards a participatory E-learning 2.0 A new E-learning focused on learners and validation of the content", International Journal on Computer Science and Engineering, Vol. 2, No. 1, January 2010, pp. 1-7.

Anouar ABTOY received the Master degree in electronics and telecommunications in 2008 from Abdelmalek Essaadi University in Tetouan, Morocco. Currently, he is a PhD Student in Computer Science. He is also is a student member of the IEEE Computer Society. Ongoing research interests: Web 2.0, collaborative intelligence, online identity, evaluation and assessment of information and content.

Noura AKNIN received the the PhD degree in Electrical Engineering in 1998 from Abdelmalek Essaadi University in Tétouan, Morocco. She is a Professor of Telecommunications and Computer Engineering in Abdelmalek Essaadi University since 2000. She is the Co-founder of the IEEE Morocco Section since November 2004 and she is the Women in Engineering Coordinator. She has been a member of the Organizing and the Scientific Committees of several symposia and conferences dealing with RF, Mobile Networks, Social Web and information technologies.

Boubker SBIHI received the PhD degree in computer science. professor of computer science at the School of Information Science in morocco. He is also the head of Information management department. He has published many articles on E-learning and Web 2.0. He is part of many boards of international journals and international conferences.

Ahmed EL MOUSSAOUI received the PhD degree in electronics at the University of BRADFORD in 1990. In 2007 he received the international master in E-learning in the Curt Bosh institute in Switzerland. He has been a member of the Organizing and the Scientific Committees of several symposia and conferences dealing with RF, Mobile Networks and information technologies. He has participated in several projects with France and Spain. Currently, his is the vice president of Abdelmalek Essaadi University in Tétouan - Morocco.

Kamal Eddine EL KADIRI received the PhD degree in computer science and he is professor of computer science at Faculty of Sciences of Tetuan in Morocco. Currently, He is the director of the National School of Applied Sciences (ESNA) of Tétouan and also the director of LIROSA laboratory. He has published several articles on E-learning and Web 2.0. He is also part of many boards of international journals and international conferences.