Yahoo!answers and learning communities

YAHOO!ANSWERS AND LEARNING COMMUNITIES

Michelle PIERI
PhD
Università degli Studi di Milano-Bicocca
Piazza dell'Ateneo Nuovo, 1
20126 Milano
michelle.pieri@unimib.it

Matteo DOMINONI
PhD, Assistant Professor of Computer Science
Università degli Studi di Milano-Bicocca
Piazza dell'Ateneo Nuovo, 1
20126 Milano

Stefano PINARDI
Researcher
Università degli Studi di Milano-Bicocca
Piazza dell'Ateneo Nuovo, 1
20126 Milano

Abstract : In the last few years we have had an incredible increase of the number of users on the Internet and of the average time spent online, along with the proliferation and improvement of the information communication technologies. These changes have enabled the emergence of extremely popular web sites that support user communities centered on shared content. Typical examples of this kind of community portals include multimedia sharing sites (i.e. Flickr and YouTube) social networking sites (i.e. Facebook and MySpace), social bookmarking sites, (del.icio.us or StumbleUpon) and collaboratively built knowledge bases (i.e. Wikipedia). Yahoo!Answers, which was launched at the end of 2005, represents a new type of community portal. It is "the place to ask questions and get real answers from real people," that is, a service that allows users to post questions and/or answer questions asked by other members of the community. In addition, Yahoo! Answers facilitates the preservation and retrieval of answered questions, aimed at building an online knowledge base. In this contribution we will describe an on going interdisciplinary project aimed at integrating Yahoo!Answers in the on line learning community of Università degli Studi di Milano-Bicocca. The aim is both to have a community built knowledge database from the students activities and a way to measure a user's reputation within the community and to identify the most important individuals. The idea is to improve the e-learning with the opportunities offered by the Web 2.0, to transform the e-learning into e-learning 2.0. In this contribution we will focalise on the theoretical background of the project and on its phases.

Key worlds : Yahoo!Answers, Web 2.0, on line learning community, e-learning.

Yahoo!Answers and Learning Communities

Introduction

The Information and communication technologies (ICT), as numerous projects have shown over the last few years at the national and international level for people of all ages, have become a fundamental part of the learning experience. Currently possibilities for applications of the Web 2.0 in the learning environment are beginning to be explored.

The term Web 2.0 was coined in 2004 by Tim O'Reilly and Dale Dougherty from O'Reilly Media, an American publisher specialized in publications regarding the new technologies and networks. They also refer to the so-called second generation Internet services like, to cite a few of the most famous ones, the blogs, google (not only as a search engine, but also as in instrument for document sharing, googledoc, and to communicate, gmail), skype facebook, flickr, youtube and wiki. The Web 2.0 is easy to use, like social networking websites, wiki, communication instruments and folksonomy which emphasize collaboration and sharing among users. In fact, the Web 2.0, not only allows for content sharing, collaboration and communication among users, but also lets users, even those with little experience, to produce contents. The use of Web 2.0 in learning environments allows all the actors involved, such as teachers and students, to participate actively, giving them the possibility to generate and propose contents, to stimulate discussions and in general, to create real learning communities.

With the advent of the ICTs in education, over the years different changes of extreme relevance have taken place; the teacher was the key figure to the student, starting from a rigid hierarchical teacher-student model to one now in which the contribution of all participants, teacher and students is valued. This goes from a systematic order which is linear and sequential or a hypermediatic disorder, from the transmission of knowledge according to a behaviorist or cognitive model or the production of knowledge according to a constructivist model. On one hand, the individual has a more active role; on the other hand, the possibilities and the needs to choose and to personalize learning paths and experiences increases.

Every individual with his personal characteristics tends more and more to construct a personal learning environment which interacts with an ever growing number of technologies and social networks. As has happened in the passage from the "traditional" face to face learning to distance learning, today the passage from "traditional" distance learning to that of "2.0" is beginning to take place. The introduction of the last generation of ICTs in education, is tied, on one hand, to the characteristics of students today, the Net Gens; while on the other hand, the ever growing need for lifelong learning, which is available "everytime and everywhere" for people who have already finished their formal education but who need to continuously update and increase their knowledge, abilities and competences. ICTs in lifelong learning, since their arrival in education, have played a fundamental role; just think for example, of how e-learning and mobile learning meet the spatial and temporal needs of working people.

The Net Gens, also called native speakers or digital natives in the literature, (the opposite of the digital immigrants), the Millennial Generation (Howe and Strauss, 2000; 2003) or also Generation Me (Twenge, 2006) were born between 1980 and 2000. They never knew the world without the ICTs or the net. As Diana and Jim Oblinger underline (2005) "as long as they've been alive, the world has been a connected place, and more than any preceding generation they have seized on the potential of networked media."

The ICTs and the net play a fundamental role in the life of the Net Gens. The Net Gens use the ICTs to communicate, play, learn, create, socialize, look for and exchange information (Eubanks, 2006). The ICTs are the culture in which young people today build and exchange meaning (Ardizzone and Rivoltella, 2008). Being born and raised in an era of "media saturation" and unlimited access to digital technologies has lead the Net Gens to have a way of thinking, communicating and learning which is completely

different from previous generations (Oblinger and Oblinger, 2005; Prensky, 2006; Tapscott, 1998). As Mark Prensky (2001) underlines "...as a result of this ubiquitous environment and the sheer volume of their interaction with it, today's students think and process information fundamentally differently from their predecessor...they thinking patterns have changed" and “today's students are no longer the people our educational system was designed to teach”. In fact, the Net Gens relationship to information is totally different with respect to previous generations: “they develop hypertext minds, they leap around” (Oblinger and Oblinger, 2005).

Among members of this generation, a process of linear thought is much less common with respect to a “bricolage” thought process, or the ability to integrate parts of information from different sources (Oblinger and Oblinger, 2005). Teachers today no longer only need to know how to use technology and to choose the best ICT available for their teaching, but they must also consider the characteristics of the generation the students come from. “Our Digital Immigrant instructors, who speak an outdated language (that of the pre-digital age), are struggling to teach a population that speaks an entirely new language” (Prensky, 2001). As Piero Bertolini (2000) underlined, perhaps the moment has come to move from teaching technology (teach how to use technology) or teaching with technology (use technology to teach) to teaching inside the technology (consider the technology as a natural environment in which the learning process takes place).

This is the context for the idea of our project.

A social glue for sparse university communities

The evolution of distance learning is inextricably linked to the development of Web technologies: how the Web is evolving into something called Web 2.0 in the same way e-learning is evolving in E-learning 2.0. (Downes, 2005; Ludovisi, 2008; Ferri, Scenini and Mizzella, 2010).

Until now, e-learning, itself, consisted mainly in the delivery and use of online courses, just think for example of the famous MIT OpenCourseWare project. The dominant technology, linked to this way of doing e-learning, is that of learning management systems (LMS). Among the most diffuse LMSs open systems there is Moodle (Chavan and Pavri, 2004).

The contents of the courses are offered in the LMS and organized in a standard way with courses divided into modules and lessons and exercises given in the form of questionnaires, tests and discussions and with the support of thematic forums and blogs.

The starting point of our project is the examination of some of the critical points of this organization. In addition to the cost of design of each course, which in principle is amortized by reusing the contents of the course itself, these activities have significant management costs. To be truly effective e-learning of this type must have a support structure, with teachers and tutors to stimulate various activities, answering questions and keeping the discussion alive and moderating it. The analysis of these critical issues is related to activities involving a large number of students, such as those at a first year university.

We believe that to make high quality, effective distance education, it is essential to avoid the isolation of the students involved in the e-learning activities. The classic instruments present in the LMS, such as discussion forums, present limitations in interaction management:for example, if students of different years or different courses want to look for answers on specific topics, indexing techniques (methods and practices that allow to find the information) of the interventions may not be fully effective or appropriate for the purpose (Gyongyi, Koutrika, Pedersen and Garcia-Molina, 2008).

We believe that an important part of the learning activityis driven by nonlinear processes, with mechanisms of trial and error (Downes, 2005). These processes should be implemented more informally as far as possible to encourage spontaneous participation of the individual in the community (Lubensky, 2006). Wikis, for example, allow a significant participation but in a formalized way too: they seem more suited for a community of experts or at least for a final stage of the learning process, where the formalities have been well learned.

Forums provide a more informal approach from this point of view, but because of their dialogic

structure, for the user it can be difficult to reconstruct the entire question-answer process that originated the discussion, which may make the Forum ineffective for knowledge sharing. Also, forums are usually organized thematically on very restricted and regulated community, and the neophyte may face serious difficulties in obtaining the requested information.

An instrument that has captured our attention for its diverse approach in knowledge sharing is the one proposed by Yahoo! Answers: an informal mechanism that allows peer help to get answers on the most diverse subjects (Adamic, Zhang, Bakshy and Ackerman, 2008). What is striking is the overwhelming growth of this system and its use from a great variety of users: when you run a search on different themes it is not rare to find answers generated by Yahoo! Answers in the Googletop rank list among the pages of proposals.

Yahoo! Answer is a system to ask questions and obtain direct answer (Gyongyi, Koutrika, Pedersen and Garcia-Molina, 2008; Adamic, Zhang, Bakshy and Ackerman, 2008). In Yahoo! Answer there are"reward"mechanisms to stimulate the generation of quick responses with a rank on the usefulness of the answers. In general the audience to which the questions are addressed is made up of millions of users: the billionth question has recently been generated. This mechanism has the characteristics necessary to induce the generation of fast responses in a very general community, it seems particularly interesting for real-time support for a course. In principle, the community may go well beyond the course boundaries: in general you can reach an audience of many more orders of magnitude greater than that of normal courses online.

The cons of this approach is that there is no guarantee of the quality of the answers obtained. The involved community is absolutely uncontrolled. This creates a serious problem: the absolute lack of control over the quality of the questions themselves, which may be too broad and not sufficiently focused. Caustic criticism that is often given to this type of instrument is that it leads to creating communities of "the blind leading the blind." (Liu and Agichtein, 2008).

Our goal is to create a social glue for sparse university communities exploiting Yahoo! Answers as a social tool partially correcting the above described deficiencies of traditional LMSs and effectively sharing the knowledge. On the other hand, our approach is aimed at focalizing the Yahoo! Answers mechanism, filtering the generated knowledge with an evaluation about the quality of the questions and answers.

The project is initially aimed for the community of the Università degli Studi di Milano-Bicocca, that has approximately 30,000 students from different subject areas and interests. At first,we want to identify the already existing learning communities inside our campus that already have a presence on the network (sites, blogs, forums , etc.) to involve them – together with our researchers and a selected community of students – asking them to participate in the project and actively promote the community. This phase is necessary both to kick off the community, and to create a small group of experts who could observe and measure the effects of the spontaneous interactions inside the growing e-learning population.

In order to connect inside and outside the academic borders, the small, spontaneous, and very specific e-learning communities that are typical of a university campus, every user will have the opportunity to “export”his/her questions from the internal e-learning communities(for example from the Moodle forums) to the Yahoo! Answer community. This mechanism will allow the single users of a community with specific interests to reach a wider and more heterogeneous population of potentially millions of people, who could read and eventually answer the given questions. The idea is to reciprocally connect isolated communities and students, exploiting the mechanism of a very popular and diffused “questioning-answering” social instrument.

Technically speaking, it will be done by constructing a software architecture, that will connect a traditional e-learning instrument (Moodle), which is already available to our students, to a set of complementary services dedicated to “live communication” (media, communication, and web servers) and to the Yahoo! Answer social site using a MS Live authentication service (Yahoo, and MS Live authentication services are reciprocally federated).

The next goal will be to collect, select and assess the generated information, using both a mechanism of subjective evaluations of the students and statistical-machine learning techniques in order to classify the quality of the given answers, to select the "best" ones and identify the best peer helpers (Harper, Moy and Konstan, 2009; Wu, Wang and Cheng, 2008; Suryanto, Lim, Sun and Chiang, 2009). In particular we want to understand what the characteristics of a“best answer” and of a“best peer helper” are. We consider as candidate best peer helpers the students with expertise or willingness to participate, we want to exploit statistical mechanisms that will help to identify them (Gyongyi, Koutrika, Pedersen and Garcia-Molina, 2008; Adamic, Zhang, Bakshy and Ackerman, 2008; Koutrika, Bercovitz, Kaliszan, Liou and Garcia-Molina, 2009)

We believe that the implementation of this mechanism will reduce the problem of students’ isolation by encouraging their participation through a simple, intuitive interaction with a broader community and the implicit reward to quickly getting an answer; we want to store and classify the created knowledge, to analyze the spontaneous mechanisms of e-learning interactions and "last but not least”, to create a "live learning community” that could extend beyond the typical dimension of a single university course.

The phases of this project are the following.

Step 1. We have contacted the departments of the Università degli Studi di Milano-Bicocca where Moodle is used as an e-learning platform. The first step is to introduce an already developed software mechanism (a plugin) in the Moodle platform allows our students to transparently reach (to authenticate into) the Yahoo! Answers community: we want to give the possibility to students who voluntarily agree to participate to automatically submit the questions both to the forums thematic course of Moodle, and to the Yahoo! Answers universities sections. These sections are already present but scarcely used.

The effects will be to broaden the audience to a community which is wider than the “usual” single-course single-university population and to enrich the university sections of Yahoo!Answer with new challenging question-and-answer threads.

Step 2. The second step is to trace and save the indexing of the question-and-answer threads and store them into servers managed by our University. This is technically possible without having direct access to the Yahoo databases because the actors who generate the questions are easily identifiable as members of our audience, according to project implementation schemes (see Fig.1).

In Yahoo! Answers social network anyone is qualified to give an answer, but only the person who asks the question can judge the quality of the given answers. We cannot and do not intend to directly modify this mechanism, which is responsible for the popularity of the system. However, this method creates a great quantity of generic, unverified and very low quality answers [see: questioning answer yahoo], we consider this from the point of view of the quality of information, as “noise” affecting the system.

We want to introduce a “filter” to reduce or eliminate this type of“noise”. We want to “internally” extend the evaluation mechanism about the quality of answers, giving not only the originators of the questions but every student of our community, the possibility to make an evaluation of the quality of the given questions and answers. At first, this “filter” will reduce the noise that affects these “unsupervised” social tools. This method will also be supported by a classification mechanism (using statistical machine-learning and text-mining techniques), which at first will index the question-and-answer threads and then will use the users’ evaluations to classify question-and-answer quality.The dimensions of the potential audience that will go over the traditional borders of the specific faculties courses will help us to assess these classifications methods.

Step 3. The last phase of the project is to make a real evaluation of the effects of the instrumentation. This will be done with a systematic evaluation of the material generated, both in terms of quantity and quality, trying to

determine the actual propagation of the questions and the diffusion of good information ("best answer") in the users' community.

Conclusions

We want to create an architectural instrumentation that will integrate and extend the mechanisms highlighted in Answer! Yahoo within a traditional Learning Management System, like Moodle. We want to study the effects in terms of quantity of new information generated and to understand the quality of the questions and answers that subsequently has been created. On one hand, the intention is to adapt LMS applications to existing mechanisms of social interactions, introducing methods that are perceived as more intuitive and natural by students. On the other hand, we intend to control the quality of the information, introducing a filter mechanism that is guided by the evaluations of the community of students, in particular by the most qualified peer helpers, which is enhanced by supplementary mechanisms of statistical "intelligent" classification. We want to use and exploit this instrumentation and this mechanism to help and to improve institutional educational purposes, but we think that these methods could be easily extended far beyond the institutional boundaries.

Acknowledgements

We want to thank F.Buccoli Bechelor CS Student for his information about MS Live and the related architectural design.

Bibliography

Adamic, L. A., Zhang, J., Bakshy, E., Ackerman, M. S. (2008), Knowledge sharing and yahoo answers: everyone knows something. In *Proceeding of the 17th international Conference on World Wide Web* (Beijing, China, April 21 - 25, 2008). WWW '08. ACM, New York, NY, pp. 665-674.

Ardizzone, P., Rivoltella, P. C. (2008), *Media e tecnologia per la didattica*, Vita e Pensiero, Milano.

Bertolini, P. (2000), Percorsi formativi e nuovi territori educativi, in *Wwwchiesainrete. Nuove tecnologie e pastorale, interventi, relazioni e contributi del Seminario di Studio promosso dall'ufficio Nazionale per le Comunicazioni Sociali Servizio Informatico* (http://www2.glauco.it/chiesainr/assisi/Assisi.pdf).

Chavan, A., Pavri, S. (2004), *Open-source learning management with Moodle Linux Journal*, 128, 2.

Downes, S. (2005), *E-learning 2.0*. Elearn Magazine.

Eubanks, S. (2006), *Millennials – Themes in Current Literature*, prepared for Azusa Pacific University, April 24, 2006.

Ferri, P., Scenini, F., Mizzella, S. (2010), *I nuovi media e il web 2.0. Comunicazione, formazione ed economia nella società digitale*, Guerini Editore, Milano.

Gyongyi, Z., Koutrika, G., Pedersen, J., Garcia-Molina H. (2008), Questioning Yahoo! Answers. *Int'l Workshop on Question Answering on the Web, in conj. with WWW*.

Harper, F. M., Moy, D. Konstan, J. A. (2009), Facts or friends?: distinguishing informational and conversational questions in social Q&A sites. In *Proceedings of the 27th international Conference on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 759-768. DOI= http://doi.acm.org/10.1145/1518701.1518819.

Howe, N., Strauss, W. (2000), *Millennials rising: The next great generation*. Random House, New York.

Howe, N., Strauss, W. (2003), *Millennials Go to College: Strategies for a New Generation on Campus*, AACRAO and LifeCourse Associates, Washington, DC.

Koutrika, G., Bercovitz, B., Kaliszan, F., Liou, H., Garcia-Molina, H. (2009), CourseRank: A Closed-Community Social System Through the Magnifying Glass3rd Int'l AAAI Conference on Weblogs and Social Media, May 17 - 20, San Jose, California.

Liu, Y., Agichtein, E. (2008), On the evolution of the Yahoo! Answers QA community. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in*

Information Retrieval (SIGIR). pp.737–738.

Ludovisi, F. (2008), E-learning e web 2.0: una dimensioneb sociale dell'apprendimento virtuale, URL: http://www.isfol.it/DocEditor/test/File/Focus_Isfol_Conoscenza_n.1.pdf.

Oblinger, D. G., Oblinger, J. L (eds.) (2005), *Educating the Net Generation. EDUCAUSE*, Washington, D.C.. From: http://www.educause.edu/books/educatingthenetgen/5989 (accessed March, 2007).

Prensky, M. (2001), *Digital natives, digital immigrants. On the Horizon*. NCB University Press, 9 (5), October 2001. From http://www.marcprensky.com/writing/default.asp (accessed April, 2007).

Prensky, M. (2006), *Don't bother me Mom-I'm learning*. Paragon House Publishers, Minneapolis.

Suryanto, M. A., Lim, E. P., Sun, A., Chiang, R. H. (2009), Quality-aware collaborative question answering: methods and evaluation. In *Proceedings of the Second ACM international Conference on Web Search and Data Mining* (Barcelona, Spain, February 09 - 12, 2009). R. Baeza-Yates, P. Boldi, B. Ribeiro-Neto, and B. B. Cambazoglu, Eds. WSDM '09. ACM, New York, NY, 142-151. DOI= http://doi.acm.org/10.1145/1498759.1498820.

Tapscott, D. (1998), Growing up digital: The rise of the net generation. McGraw Hill, New York.

Twenge, J. M. (2006), *Generation me: Why today's young Americans are more confident, assertive, entitled – and more miserable than ever before*. Free Press Simon and Schuster, New York.

Wu, H., Wang, Y., Cheng, X. (2008), Incremental probabilistic latent semantic analysis for automatic question recommendation. In *Proceedings of the 2008 ACM Conference on Recommender Systems* (Lausanne, Switzerland, October 23 - 25, 2008). RecSys '08. ACM, New York, NY, 99-106. DOI= http://doi.acm.org/10.1145/1454008.1454026.

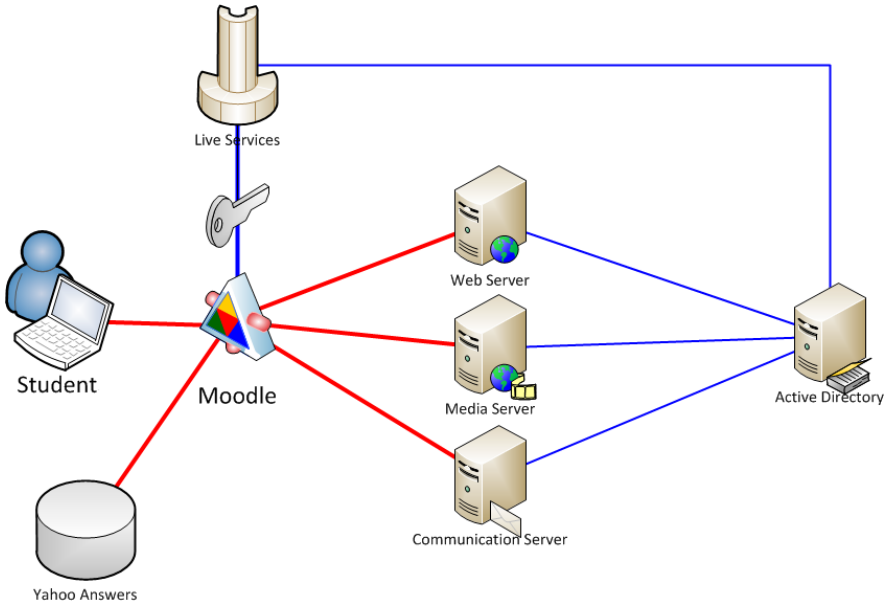

Fig.1 The Architecture. At the center of the communication structure a traditional instrument of elearning (Moodle).A connection to the MS Live service will allow the access to a set of complementary communication instruments and to Yahoo Answer. In red the communication connections, in blue the authentication paths.