Chapter 6

Collaboration in Virtual Space in Digital Humanities

Laszlo Hunyadi[1]

Introduction

The introduction of computing methodologies into the humanities has brought about significant changes both in the scope of humanities research and in the way that research is to be carried out. These new methodologies have offered the chance to ask questions that have never been asked before and to give new and more satisfying answers to questions previously asked. Importantly, interdisciplinarity as a scientific approach to the study of complex issues has gained more and more importance, and technological advances, together with the need for interdisciplinary studies, have made it increasingly essential to do research in collaboration. The development of humanities computing has been noticeably enhanced by the various emerging forms of virtual collaboration offered by the use of the internet. To name a few pioneering initiatives without trying to give a comprehensive overview, HUMBUL, which started in 1985, offered catalogued and reviewed websites relevant to the humanities at a time when today's search engines did not even exist. Nearly ten years on, the book *The Whole Internet* (Krol, 1992) could still be printed on a few hundred pages. The discussion group *Humanist* started in 1987 and is probably one of the oldest existing international online seminars still in existence and flourishing today. And of course the humanistic disciplines themselves have been enhanced by the emergence of easier and faster ways for colleagues to collaborate across the boundaries of subjects, institutions and countries.

The emergence of international organizations for humanities computing attests to the rapid growth in communities in the field: the Association for Literary and Linguistic Computing (ALLC), founded in 1973, and the Association for Computing in the Humanities (ACH), founded in 1978, initially served the communities mainly in Europe and North America, respectively, and the new publications— *Literary and Linguistic Computing* and *Computing and the Humanities*—which they established opened new forums for scholarly work in the field throughout the world. Community building and support has been continuous ever since: the

1 This article was partly supported by the joint Hungary–EU Grant No. TÁMOP 4.2.2-08/1-2008-0009.

emergence of the Alliance of Digital Humanities Organizations (ADHO) in 2007 as the umbrella organization of ALLC and ACH is a natural consequence of the growth and globalization of computing in the humanities. Other organizations are encouraged to join ADHO, and the Society for Digital Humanities/Société pour l'étude des médias interactifs (SDH-SEMI, founded in 1986 as the Consortium for Computers in the Humanities/Consortium pour ordinateurs en sciences humaines) has recently done so. ADHO hopes to promote the emergence of new, online publications in the field, such as *Digital Humanities Quarterly* (*DHQ*), which began publication in 2007.

In general, the emergence and development of interdisciplinary research in the humanities has been accompanied and/or supported by the emergence and development of various forms of collaboration in the past decades. The two seem to go hand in hand, one facilitating the other and offering new perspectives both for academic studies and research activities. What seems to be shared by many of these efforts is preference for virtual, online forms of collaboration. In what follows, we will give examples of virtual collaboration as it appears in today's practice, ranging from academic education to academic research to R&D activities. This attempts to show how digital humanities can contribute to advances in other research fields and also, vice versa, how it can benefit from traditionally "remote" disciplines.

The activities and thoughts associated with and inspired by the idea of collaborative work described in these pages are strongly influenced by the intellectual spirit I had the chance to encounter through the many years of my acquaintance with Professor Harold Short. One main message that has remained for me as a guiding light ever since is that we have to turn to our fellow colleagues with a keen interest in their work, offer and incorporate new ideas, and seek further and ever more productive forms of interdisciplinary research. The international community of digital humanities has always been a "safe haven" for me where I could experience a warmly welcoming and stimulating, but critical, intellectual environment for the development and testing of innovative ideas on the crossroads of exciting, mutually complementary as well as competing fields. It has been motivating for me to see a new discipline emerge and to understand that this new discipline will leave enough room for our original fields (linguistics in my case) while creating a "virtual cloud" of collaboration within and across individual disciplines. I can only hope to be able to contribute to handing over this spirit to the next generation.

Collaboration in Virtual Centers for "Real" Education

There is a perception in a number of universities in Europe that humanities faculties are currently facing troubled times, only partly shared by some other academic disciplines. The decrease of state funding has noticeably affected the humanities, while the number of students enrolled in many traditional majors

is stagnating or even gradually falling. The cut in state funds has affected the humanities especially badly because—in contrast to the sciences—external funds are much harder to generate. Whereas in the sciences demands for technological advances obviously generate additional funds, the traditional approach of the humanities, not being technology- or even "consumer"-oriented, results in the humanities disciplines becoming less competitive in the race for funds. In addition, the decrease of enrolments is ominous: fewer students choose traditional humanities disciplines, mainly due to market pressure—they wish to have a degree which opens their way to a decent career and prestige. Regardless of sound argumentation based on well-tested traditional values, entering the competition both for funds and for students with old perceptions seems not too promising. However, studying present-day and prospective future demands may offer a solution: the introduction of computing in humanities education and research may have the effect that, similarly to the introduction of computing in several other disciplines, the use of computing methodologies might enhance traditional studies in the humanities and result both in new answers and new discoveries. This is the route some leading universities in Europe and America have chosen. They introduced computing in various aspects of the humanities, thus enhancing and further widening the scope of humanities computing for education, research and application development. One of the institutional forerunners in humanities computing was King's College London, where a new department was set up to pursue the goal of institutional development of humanities computing. The emergence of digital humanities (DH) as a new discipline, as we know it now, was the result of continuous development in many areas of humanities computing, both in education and research. The idea was gradually developed that DH as an academic offering could find its place among disciplines in the humanities and even offer a contribution to disciplines beyond the humanities as well.

When we at the University of Debrecen were first contemplating introducing humanities computing into the Faculty of Humanities, we hoped that by doing so we could give fresh momentum to humanities education and humanities research. We expected that, as a result, we could build a community that would bring about a substantial renewal of the practice of traditional humanities and create a new, more comprehensive "space" for humanities education and research. We discovered, however, that to start a new program with no history in the country at all is no easy matter. It is first necessary to establish DH as a recognized academic discipline and then get the accreditation to begin a program in a particular institution. Both requirements need an institutional background that we, at the outset, equally lacked. Normally, the traditional way of running an academic program is to associate it with a single (rarely more) academic discipline, but we soon realized that DH cannot be associated with a single "traditional" humanities discipline. Therefore our endeavor was twofold: establish a single organizational unit for DH and then a single discipline associated with it.

Initially, we were aware that there were individual academics teaching a subject that in one way or another included some applied computational methodology. There was, however, no collaboration or any other professional relationship among the individual teachers. Thus, we decided as the first step to establish a community that would later on undertake the education as well.

In order to build the institutional background necessary for any hope of success in national accreditation, we needed some traditional form of organization that would submit the proposal for accreditation, either a department or a research center. Since DH was a new discipline not belonging to the profile of any existing department or research center, we decided to establish a new, virtual Center for Digital Humanities. It needed to be virtual because it would rely on existing teaching positions; these positions would remain in their respective departments and offer an extra service to run courses in humanities computing as well.

The Center for Digital Humanities set out its program: it planned to elaborate a proposal for an MA in Digital Humanities, secure the teaching staff, and offer additional activities. The proposal was essentially based on two sources: similar DH programs elsewhere for reference, consultations, and possible co-operation; and the availability and interest of the potential teaching staff. The two specializations offered by the program would reflect two classes of broader interest: *cultural heritage preservation* and *language technology*. The disciplines that are thus represented are quite numerous, enabling students to enter the program with a BA in, among other subjects, modern and classical languages and literatures, history and ethnology. The disciplines taught in the MA program range from classical humanities and information science to music, even to related issues in architecture and the sciences.

We cannot leave this story without documenting its outcome. In the first round, our submission for recognizing DH as an independent discipline was unsuccessful. For the second submission, we had to convince the respective bodies that: (a) what we offer is not information science proper, but a humanities-oriented teaching and research scenario with a strong reliance on approaches, methodologies, and techniques of IT; and (b) what information science, especially some exams, such as ECDL, offers does not—contrary to some superficial assumptions—cover those fields that are the characteristics of DH: the integration of humanities and some other disciplines with information science from a particular, unique perspective. Based on strong arguments supported by spectacular achievements in the wider field of digital humanities, DH was finally recognized as an independent academic discipline at the masters level. Shortly after that, the University of Debrecen received national accreditation to also run the program. Following a successful series of lectures in 2010 as an introductory course, with the enrollment of more than two dozen students with a wide variety of subjects as their primary humanities background, we look forward with enthusiasm to launching the first full-fledged program in 2011. Beyond its own advancement, DH will also serve as an example in Hungary of how interdisciplinary co-operation across traditional academic boundaries can contribute to the modernization of higher education.

Virtual Space: The Academic Benefit of Professional Collaboration

Today's technology allows us to disseminate knowledge to a much wider audience than is possible in a traditional classroom setting. To reach as wide an audience as possible is especially important in the case of a new discipline such as DH.

This is what led us to make our participation at the annual European Researchers' Night, an international event, through videoconferencing. In the last three consecutive years our hub in Debrecen has connected participants from Debrecen, Budapest, Oulu and Dublin via cyberspace. As in an ordinary event, we had "plenary talks" and "round table discussions" alike. The audience from all sites could enjoy the event and be part of it by asking questions and reflecting upon any issues arising. A videostream was also broadcast to enable "unregistered" viewers to participate. The impact of these two events on the audience was quite measurable: being mostly students, they received the talks and discussions with much enthusiasm. Even some term papers originated from them as a reflection on the scientific merit of these events, suggesting that DH has found a fruitful ground to further grow and develop. The impact on students of these events is also measurable by the fact that last year's was organized and run by PhD students, suggesting that DH will have its future among the younger generation.

In addition to these European Researchers' Night events, the same technology has also been tested and utilized in further educational activities within the framework of our Culture and Technology Seminar Series. The series were jointly designed by the universities of Debrecen, Leipzig, Oulu, and King's College. Later Glasgow, Bologna, and Dublin also joined in our effort to bring the major ideas of DH closer to undergraduates in the respective universities. Students of various disciplines across the continent were connected weekly by the central idea as well as the technology. Professors from the respective universities (E. Burr, D. Buzzetti, L. Hunyadi, W. McCarty, M. Moss, L.L. Opas-Hänninen, and others) gave inspiring talks on their research, all focusing on a particular aspect of digital humanities, ranging from linguistics to literature, history, computer science and engineering. Later, as interest grew, calls for participation were also extended to PhD students. We had exciting and highly motivating presentations by graduate students from Bologna, Debrecen, and Dublin, three universities located far away from each other but connected in the virtual space by a common interest in DH. Already the fact that regular scientific meetings in such a virtual space, under the umbrella of digital humanities, was made possible shows that DH as a discipline has found its place in higher education.

The idea that culture and technology could organically meet, sharing the approach of DH, was further developed in Leipzig where the first Summer Schools in Culture and Technology were held in the summers of 2009 and 2010. There are further plans as well: based on the success of seminars run by doctoral students, we intend to offer a virtual conference for young scientists in DH, to be organized simultaneously at various university sites in Europe and broadcast to the participating institutions and possibly beyond. We hope that such a conference will

become yet another means of scientific motivation for those young people who are engaged in any aspect of DH and are interested in developing collaboration without borders. This is how the inter- and multidisciplinary concept of DH equally inspires the "virtual" and the "real."

The Benefit of Professional Collaboration in Virtual Space for Research and Development

Digital humanities is, by definition, inter- and multidisciplinary. Being part of the DH community, one has a strong sense of the benefits of crossing traditional academic boundaries in order to collaborate with respective partners from other disciplines in order to "arrive home," that is, to give answers to questions often proving too complex for one's single primary discipline. We can see the benefits of such collaboration in at least three areas: (a) at the institutional, organizational level; (b) in activities of dissemination; and (c) in running individual projects. Below, I will suggest some concrete examples of these benefits.

(a) At the institutional, organizational level, we must give credit to a significant event with far-reaching consequences for the future of DH, originally initiated and gradually pursued by Harold Short and John Unsworth, backed by their respective organizations: the "merger" of ALLC and ACH under the umbrella of ADHO. Originally, ALLC and ACH used to function on a semi-geographical basis; by the establishment of ADHO (Alliance of Digital Humanities Organizations: <www.digitalhumanities.org/>) in 2005, the DH community has come of age by offering global collaboration and expandability, while at the same time preserving the "local" focuses of the constituting organizations. Going global means that, based on previous successful large-scale projects and both geographically and professionally wide scope actions, it has now become possible to institutionalize the achievements and concentrate our academic, human, and material resources to accomplish even further reaching goals. The first impact at the organizational level was the joining in 2007 of SDH-SEMI, the Canadian organization for DH, a move that simply accomplished a professionally supported natural process. Recently, this movement towards globalization is perceived more and more strongly with regard to Japan: the establishment of ADHO gave significant impetus to strengthen and further enhance the presence of DH in that country, with the possibility of establishing a regional chapter some time in the future. It is also expected that the increasingly global reach of DH through ADHO will result in further academic and organizational advances in other parts of the world as well.

(b) Activities of dissemination are very important in the development of an ever growing community, through making available new advances and sharing best practice in particular fields. Several seminars and other similar events have been initiated or run by the constituting organizations of ADHO in many parts of the world. I will only give one example here, that of a conference and seminar organized by the West Bengal University of Technology, Kolkata in 2007, where

experts from ALLC were invited to participate. This conference dealt with the issue of digitization of a vast number of Indic languages, languages that either did not have at that time any encoding system or, on the contrary, had more than one but none of them appropriate for standard machine representation and analysis. The participation of delegates from ALLC was judged to be very useful in that, step by step, they were guided through actual and accomplished projects of digitization as possible examples of the use of standard methodologies. The list of local participants included a wide range of professions, from linguistics to IT to engineering and even nuclear physics, showing that these professionals deeply understood the needs of their country in the preservation of objects of their cultural heritage. Regardless of their academic orientation, they were ready to contribute to this magnificent goal by starting from the very basics and moving steadily towards accomplishment. ALLC was proud to be a contributing part of this—in a sense—historic moment.

(c) The essential strength of inter- and multidisciplinary approaches is clearly shown in joint projects requiring the contribution of DH (by itself multidisciplinary) to the accomplishment of tasks involving further disciplines beyond DH. Below, we will give an account of some leading ideas behind a fairly complex project involving traditional humanities, information science, and engineering, with computational linguistics playing a significant role.

The project HuComTech ("The Theoretical Foundations of Human–Computer Interactions": <http://hucomtech.unideb.hu/hucomtech>) at the University of Debrecen is based on the general observation that man–machine communication systems usually lack that "appeal" of naturalness which makes such an interaction easy and straightforward for the human participant. The reason is not just the fact that computer graphics are still in the phase of development (on the contrary, many games or even movies can now be made to appear realistic) or speech recognition systems are still restricted to a special or specialized vocabulary, but that we do not yet know enough about human–human communication in general. The underlying idea is that, if we know how to formally represent the structure of a given human–human interaction, we can build computer systems that will "behave" in a similar fashion. But the major problem is that, being dependent on many individual and in cases never repeated factors (profile, ontology, background knowledge on the one hand, and given psychological, cultural, moral, and so on, settings on the other), human–human interaction seems to be hard to represent in a regular form. The aim of the project, then, is to learn the regularities of human–human interactions based on type scenarios and respective ontologies, separate these regularities from their context-dependent individual characteristics, and finally build and interpret a corresponding structure for the given interaction. This task requires the collaboration of many people from various disciplines, and the project thus involves computer scientists, engineers, computational linguists, specialists in communication, as well as psychologists. The task is inspiring and at the same time challenging for all. The novelty of research is mainly represented by its special interdisciplinary nature and the way in which it contributes to the

completion of the research program: it is its cross-disciplinary DH approach that unites the often seemingly incompatible counterparts. The major issue, as mentioned above, is that of the formal representation of human communication. Although several theories exist for communication, most of them miss the criteria of formal description necessary for implementation by the computer. Our task, then, is to search for and identify those building blocks of communication that, in an orderly fashion, contribute to what we can tell the computer as structure. It is the computational aspect of DH that leads us in the search for implementable structure and that does not let us go astray under the influence of certain functional features inherent in any communication, but that go beyond the scope of a technologically implementable human–computer interaction system.

Accordingly, the focus of the project is to propose a model inspired at the crossroads of the humanities, engineering, and information science, and to validate this model in a practical, working implementation. Supported by approaches aiming at unveiling the formal basic structure of communication and also at the model-based technological implementations of an arbitrary communicative event (cf. Polanyi, 1988, 2004; Jurafsky, 2004; Thione et al., 2004a, 2004b), but extending the scope to multimodality and bidirectionality, our model itself is generative (cf. Hunyadi ,2011): it inherits its approach from theoretical linguistics (the latter originally inspired by the needs of computational linguistics for language technology). The approach is modular, a significant property equally important and manageable for theoretical linguistics and technology. Namely, the complexity of the issue of understanding the basic underlying structure of communication and implementing it within technology can be reduced and, consequently, relatively conveniently handled by subdividing the flow of communication into a number of self-contained but structurally interrelated modules for the description and generation of appropriate and specific internal relations. In our model, these modules are assumed to be responsible for (a) the generation of the formal skeleton of all possible communication events, (b) the functional extension of the formal basis of a given event, and (c) the pragmatic extension of the functionally already extended, rich formal basis. This latter module contains the interface between the verbal (and, from a technological point of view, informal) description of an event and its formal technological implementation. All this is done in a unified manner: each module consists of a finite set of primitives only specific to the given module and a finite set of rules of derivation to derive the set of all possible structures within that given module. Modules are connected by applying the same unifying principles and rules to them within a structural hierarchy. The big challenge is how to map non-formal, practically verbal descriptions of events of communication onto formal, parametric data, the latter being the only possible input relevant to technology. This is where our approach is thought to offer a strong theoretical contribution to the problem. Applying to our model the theory of fuzzy systems to capture linguistic descriptions of certain cognitive concepts (cf. Zadeh, 1965; Bunt and Black, 2000), we map linguistic variables onto technological parameters

in order to handle the two, conceptually apparently very distant, fields of human communication and technology in a single unified system.

In addition, a further benefit of applying a (humanities-based) generative model to the technology of human–machine interaction is that this model is then bidirectional: it equally models analysis and synthesis, two aspects of communication that virtually happen at the same time but that are usually implemented as two independent and (at least partly) incompatible models. Instead of building separate systems for analysis and synthesis using principally different models, our generative model allows for capturing the relevant aspects of human–human communication as they exist in unity in our own behavior. Even though the task is definitely highly complex, we believe that an attempt to technologically implement at least a restricted subset of communicative relations can serve as a means of validation for the model and its application.

The complexity of the issues described above suggests that this project can only be carried out in strong co-operation by a number of disciplines. There are many challenges: linguists, engineers, and IT people have radically different conceptual and working methodologies and cultures. We believe that DH that is inter- and multidisciplinary by itself can effectively contribute to the successful establishment of synergy between these groups that are professionally distant but closely related by the definition of the final goal.

Our virtual Center for Digital Humanities, which has already proved to be successful in making a proposal for the academic program MA in Digital Humanities and which has led further academic activities in virtual space, plays a leading role in this technological project, with the hope that the complex approach of DH will bring this project to successful completion.

Multidisciplinary Collaboration in a Virtual Lab

With significant advances in computing and the use of the internet, it has been made possible to share one and the same physical infrastructure among a number of researchers and research communities, both for individual projects that are connected by way of the use of the same infrastructure and for joint complex tasks. I will present here our current effort to build such a shared physical and virtual research infrastructure, where traditional disciplines of the humanities reach out to and become an integral part of engineering application development.

As a logical continuation of the HuComTech project mentioned in the previous section, we at the University of Debrecen are in the process of creating a cognitive robotic lab that is: (a) real (physical) in the sense that it is based on a real technological implementation of a primary lab, with all the necessary components like testing space, robots, external sensors, and a complex computing system; and (b) virtual in the sense that we can network into this system a number of "stripped down," secondary labs which lack the robots but which each has a testing space and is equipped with the necessary external sensors and the appropriate computing

system. These two kinds of labs are integrated into a single research space using the VIRCA (Virtual Collaboration Arena; cf. <www.virca.hu/>) framework. There are at least two far-reaching benefits of this research setting: (a) each of the secondary, robotless labs can carry out experiments with the robot of the distant primary lab; (b) our labs (either the primary or the secondary) can also be integrated into a larger network of similar labs running the same protocol. As a consequence, it is not only the case that we can use the infrastructure of other distant labs, but we also have the opportunity to invite remote projects to join us and carry out joint research and experiments with and within our project.

What is the role and interest of humanities and digital humanities in building the above network of real and virtual labs, if the setting seems to be so technical and technology oriented? We learn something significant from engineering and the application of computing methodologies: it can be said that the humanities is strong in building theories such that their validation is usually done by testing their logical consistency. On the other hand, engineering and computing are also necessarily supported by theories, but the nature of their disciplines requires validation by testing them against their physical implementation. Second, whereas findings in, among others, physics or chemistry are verified in real-world tests and experiments, humanities will also have a significant contribution to engineering and especially robotics only if any appropriate theory aimed at engineering development is exposed to testing through and within technology. It is especially true of a theory of communication aimed at cognitive robotics, where any theory has to be tested against the "behavior" of the given robot. In this way a humanities theory can be verified, refined, or rejected as a result of exposure and testing against its technological, robotic implementation. The network of real and virtual cognitive robotic behavior with humans will be able to remotely manipulate the behavior of robots and observe their response, and, ultimately, test the validity of a given theory of human behavior.

We are aware that the use of a network of real and virtual labs as described above reaches beyond the scope of humanities and even digital humanities, but we also believe that the establishment of such a research environment is very much in line with the general purpose of DH: we need to include in humanities research the power and possibilities of computing in order: (a) to offer a more complete answer to some traditional issues in humanities; and (b) to reach out to multidisciplinary research where we believe our knowledge and methodologies can contribute to the understanding and implementation of more complex issues, that are traditionally non-specific to humanities. Time will show if this direction that we are taking now by building an infrastructure of real and virtual labs will indeed yield the results we are envisaging at present.

Summary

The purpose of this chapter has been to show through some concrete examples that humanities research has come a long way since the introduction of computational

methodologies, to develop into a new discipline called digital humanities. The new technologies adopted from computing have made it possible to introduce novel scientific perspectives by facilitating the emergence of further forms of collaboration among scientists. Inter- and multidisciplinary research has become a "must" to solve certain emerging new issues, giving rise to virtual research collaboration as well. Research, academic education, and technological advances go hand in hand while transforming the whole long-established landscape of humanities, leading it into the digital era which promises so much. Pioneering institutions and organizations, as well as newly emerging ones, lead this march, with the indispensable participation and co-operation of a rapidly growing number of single individuals.

References

Bunt, Harry and William Black, 'The ABC of Computational Pragmatics', in Harry Bunt and William Black (eds), *Abduction, Belief and Context in Dialogue: Studies in Computational Pragmatics* (Amsterdam: John Benjamins, 2000): 1–46.

Hunyadi, László, 'Multimodal Human–Computer Interaction Technologies: Theoretical Modeling and Application in Speech Processing', *Argumentum* (2011 [in press]).

Jurafsky, D., 'Pragmatics and Computational Linguistics', in Laurence R. Horn and Gregory Wards (eds), *Handbook of Pragmatics* (Oxford: Blackwell, 2004).

Krol, Ed, *The Whole Internet: User's Guide and Catalog* (Sebastopol, CA: O'Reilly & Associates, 1992).

Polanyi, L., 'A Formal Model of Discourse Structure', *Journal of Pragmatics* 12 (1988): 601–39.

Polanyi, L., 'A Rule-based Approach to Discourse Parsing', in *Proceedings of SIGDIAL '04* (2004), Boston, MA.

Thione, G.L., M. van den Berg, Ch. Culy and L. Polanyi, 'Hybrid Text Summarization: Combining External Relevance Measures with Structural Analysis', *Proceedings of the ACL Workshop Text Summarization Branches Out* (2004a), Barcelona.

Thione, G.L., M. van den Berg , Ch. Culy and L. Polanyi, 'LiveTree: An Integrated Workbench for Discourse Processing', *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, 25–26 July 2004, Barcelona, Spain (2004b): 110–17.

Zadeh, I.A., 'Fuzzy Sets', in *Information and Control* 8.3 (June 1965): 338–53.