Revue Ouverte
d'Intelligence
Artificielle

Introduction

Digital Humanities is often traced back to Roberto Busa's Index Thomisticus project, which, in 1949, sought to automatically create an index of Thomas Aquinas's Summa Theologica assisted by computers. Its origins can also be found in the works of Augustus de Morgan, who, in 1851, submitted a quantitative study of word frequency in order to characterize the style of various authors. Applying computer science to the humanities continued into the 1960s with Alvar Ellegård's attempt at automatically determining the authorship of the Letters of Junius, and with Frederick Mosteller and David L. Wallace's research aimed at identifying the authors of The Federalist Papers. The creation, in 1963, of the Centre for Literary and Linguistic Computing in Cambridge and of a text analysis research group at the University of Tübingen, as well as the emergence of the journal Computer and the Humanities in 1966, testify to the early interest shown in using calculations and computers in the domain of the humanities. This trend was further reinforced in the 1970s with the regular publication of the Association for Literary and Linguistic Computing's newsletter, and the founding of the ICCH (International Conference on Computing in the Humanities). Subsequently, formed in 1986, came Ansaxnet, the first discussion list for the humanities, and the TEI (Text Encoding Initiative) project, which set out guidelines for the coding and exchange of texts. Additionally, as far as Francophone initiatives are concerned, the JADT, or Journées d'analyses de données textuelles, has been held on a recurring basis every two years since 1992. It wasn't until the turn of the millennium, however, that the field of Humanities and Computing was transformed into Digital Humanities. This new term signalled a fundamental change in the role that computer science was to play in the future, as it would no longer merely be employed as a tool by the traditional scholarly disciplines but would also assist in forging new operators of interpretation. This thus created an epistemological paradigm shift in the humanities fields, which started going digital. In practical terms, this means that: in its preliminary era, that of Humanities and Computing, the computer helped scholars construct indexes and concordancers, determine authorship by using lexical statistics, and create electronic publications; whereas nowadays, with the switch to Digital Humanities, we are able to perform text mining, supervised and unsupervised machine learning, data visualization, semantic analyses such as named-entity recognition, graph theory, etc., in order to better understand and interpret texts.

For convenience sake, since it isn't possible to assess the developmental status of Digital Humanities in France in just a few lines, we would like to refer the reader to the 2014 report by Pierre Mounier and Marin Dacos[1]. It is clear from this, that the growth of consumer computing, which began in the 1970s, gradually permeated

---

[1] Pierre Mounier et Marin Dacos. "Humanités numériques état des lieux et positionnement de la recherche française dans le contexte international". Paris: Institut français, 2014.

all sectors of scientific research, including social sciences and humanities. Its impact naturally varied within the various "disciplines", but none were left on the sidelines. In the field of linguistics, which was readily incorporating mathematical tools, automated approaches gave rise, for example, to the highly active "Natural Language Processing" sector. The "new geography" of the early 1970s, which placed higher demands on quantitative modelling, enabled the development of the "Geographic Information System". Suffice it to mention, without making any value judgments, that very close and dynamic ties still exist between advanced quantitative modelling and economic and social history, and mathematical economics. The advent of "Big Data"[2] in the early 2000s generated another shift, also leaving its mark on contemporary digital humanities, through its commitment to using the digital format to process large volumes of data in an automated or semi-automated manner. The ease with which this data can be produced, circulated and processed has prompted us to entertain new ideas, such as constructing an interconnected history on a global scale. The sociology of networks, in turn, is developing very rapidly. Concurrently, this "Big Data" phenomenon also led to a resurgence of Artificial Intelligence[3], and one that amplified the ongoing transformation of the human and social sciences by incorporating, for example, advanced techniques for the study of textual corpora. More recently still, the use of topological data analysis techniques in the design of machine learning algorithms has created new avenues for its integration into the human and social sciences by allowing more complex data structuring.

By combining digital and mathematical techniques and extending this to the humanities, Claude Levi-Strauss's speech at UNESCO in 1958, which advocated for the "mathématiques de l'homme", set an ever-relevant course for the future: "Les besoins propres aux sciences sociales, les caractères originaux de leur objet imposent aux mathématiciens un effort spécial d'adaptation et d'invention. La collaboration ne saurait être à sens unique. D'un côté, les mathématiques contribueront au progrès des sciences sociales, mais, de l'autre, les exigences propres à ces dernières ouvriront aux mathématiques des perspectives supplémentaires"[4]. Digital technology now allows social science and humanities researchers to collect data on new media with new means of observation and to work with corpora of unprecedented sizes. Among the many challenges posed by these developments, two of them stand out as being paradigmatic with respect to the need for cross-pollination between these disciplines.

The first concerns the wide range of data available for analysing any given phenomenon: how could we reconcile the many and varied scales of observation that we now

---

[2]Les Big Data à découvert, Mokrane Bouzehoub et Rémy Mosseri (dir.), CNRS éditions, Paris, 2017, 364 p.

[3]Kersting, K. & Meyer, U. Künstl Intell, From Big Data to Big Artificial Intelligence? KI - Künstliche Intelligenz 32 (1), 2018.

[4]Levi-Strauss C. (1954). Les mathématiques de l'homme. *Bulletin International des Sciences Sociales*, VI(4), pages 643-653. [The mathematics of humankind. The specific needs of the social sciences, the uniqueness of their objectives, will require a special effort on the part of mathematicians in terms of adaptation and invention. It wouldn't be a one-sided relationship. On the one hand, mathematics will contribute to the progress of the social sciences, and, on the other hand, the specific requirements of the latter will shape new mathematical perspectives.]

have access to? In other words, in practice, how do we combine all the information extracted from digital publications, digital footprints on social media, ethnographic interviews, monographs, etc.? A quick detour through the life sciences will put us back on track here. A little less than 20 years ago, coupled with computing techniques, the life sciences launched an ambitious research program on integrative biology aimed at incorporating the data collected on various levels, from genomic to metabolic, into the analysis of "the biological human". This unifying project could inspire the development, within the Digital Humanities, of an "integrative human science", which, through thorough interdisciplinary dialogue, would be able to offer up innovative methodologies capable of operationally combining, via suitable artifacts, the "micro" and "macro" and the "quanti" and "quali" in order to tackle the complexities of the studied phenomena head-on. A project of this nature could be seen as reconnecting the foundational aspects of artificial intelligence with its present advanced developments since the issue of the multi-scale processing of information is central to deep learning.

The second challenge concerns questioning the humanities about digital technology. Let's go back to UNESCO, but in much more recent times. At the Internet Governance Forum that was held in November 2018, a workshop was devoted to the "Software Heritage" project, which is the world's largest library of software source codes[5]. It has a short history since the oldest code to be preserved by the project is that of the Apollo 11 program. This present-day "Library of Alexandria", however, already contains nearly 7 billion source files. The questions raised by the analyses of the corpora referred to in this special issue are renewed by these data that inextricably links computer science, particularly artificial intelligence, and the humanities.

As mentioned, the co-construction, occurring at the crossroads of the information sciences and the human and social sciences, now requires more than just computer science and statistics: it calls for artificial intelligence. From this perspective, we found it pertinent to gather some examples of these interactions. The descriptions of these experiments should help to identify the paths that this interdisciplinary research could take, as well as the associated technological barriers and pitfalls to be avoided. How could the tools provided by Artificial Intelligence be integrated into the knowledge that we have already acquired in the domain of social sciences and humanities? Would Artificial Intelligence be able to avoid going entirely off track if faced with hypothesis-free digital data? In what ways do these interactions between Artificial Intelligence and the humanities have the potential to produce new forms of knowledge?

A virtuous circle of mutual exchange exists between Artificial Intelligence and Digital Humanities: Artificial Intelligence transforms the work of researchers in the traditional scholarly disciplines, commonly referred to as the humanities, into the human sciences and into the social sciences. In turn, Digital Humanities provides Artificial Intelligence with new challenges.

In this issue, we pay particular attention to the implications of the former, although the latter warrants at least a brief introduction.

---

[5]J.F. Abranatic, D. di Cosmo, S. Zacchinolli (2018). *Building the Universal Archive of Source Code*. Communication ACM.

Artificial Intelligence seeks to reproduce cognitive functions such as perception, understanding and decision-making through computer simulation[6]. For certain perception tasks, particularly visual perception ones, the machine often achieves success rates close to, and sometimes even higher than humans, but unlike them, this is at the expense of having to provide a very broad range of labelled examples for it to "learn" from. And even if one has access to a sizable training set, the machine can still be deceived when faced with a fairly noisy image, whereas the human will not slip up. Understanding should be seen as the translation of perceived information, such as recorded images or words, or even texts written in natural language within a formal language that allows automatic inferences and complex queries. However, for many reasons, machine results remain significantly lower than human ones. Finally, the decisions that need to be made when there is a lack of data or time require a certain amount of intuition that the machine does not possess. So, there is a considerable gap that needs bridging in terms of improving our machines and algorithms, and one that we cannot fill solely by improving our processing and storage capabilities. We will need a far better understanding of human faculties in order to accomplish this. Since both the scholarly disciplines and the social and human sciences are now contributing to this enhanced understanding, they are likely to be of assistance in bridging this divide.

Scientific works concerning written communication, for example – from the analysis of authors' corpora to new forms of writing – furnish Artificial Intelligence with concepts, methods and tools that contribute to the formalization of various forms of speech architecture, and thus to the simulation of certain aspects of thinking. In a similar vein, the social component is proving to be instrumental in interpersonal exchanges and will undoubtedly present a challenge for Artificial Intelligence in the not-too-distant future. Thus, even though it is easier to test and study insect colonies, and even if these studies provide very fruitful results in terms of the phenomena of cooperation, they do not rule out the need for analysing the human collective. It is becoming increasingly necessary to implement artificial intelligence techniques that are adapted to human societies, rather than asking societies to adapt to the devices constructed for them. Understanding social organization and analysing societal expectations are issues that sociology, and the Social Sciences and Humanities in general, have long been invested in studying and for which they possess specific skills. Where would the development of the autonomous car be without usage analysis? More importantly, what are the relevant parameters that need considering in order to make the correct inferences rapidly and in real time, and that would allow for the speedier recognition of a scene of violence, and that would generate an almost instantaneous modification of a robot's behaviour when facing an unusual situation? Does part of the answer lie in our ability to incorporate subjective considerations – in all their social and historic depth, and that depend upon the context of the action – into our algorithms? Would the headlong rush towards data not simultaneously be what was responsible for the rebirth of Artificial Intelligence as well as its recent dead end?"

---

[6]Intelligence artificielle – état de l'art et perspectives pour la France – 2019 (p 40).

An Artificial Intelligence that serves the human and the society must necessarily assimilate knowledge about both the human being and about the society – this is the very core of the Human and Social Sciences profession. Conversely, these disciplines must analyse the impact of Artificial Intelligence on humankind and society, thereby providing them with the opportunity to decide on future developments. Naturally, since Artificial Intelligence techniques have permeated these scientific fields, maintaining objectivity in this regard might prove difficult.

As previously mentioned, the research outlined in the articles selected for this special issue does not focus as much on the contribution of Digital Humanities to Artificial Intelligence as it does on the use of Artificial Intelligence by Digital Humanities. Each article sheds light on the bonds that can be forged between these two fields. The strength of these bonds is reflected in both the wide array of research methods used by the assorted contributors and institutions, as well as in the diversity of their academic backgrounds.

The article written by Etienne Cuvelier, Sébastien de Valeriola and Céline Engelbeen illustrates the contribution of machine learning to research in medieval history. More specifically, the first issue that they tackle concerns identifying the various sources used by medieval encyclopedists. Also, by focusing on one of the three parts of perhaps the most extensive encyclopedia of the Middle Ages, the Speculum Maius, written in the 13th century by Vincent de Beauvais, the authors explored an automatic identification method based on drawing comparisons (using an adaptive metric) between the Speculum's explanatory notes and their pre-identified potential sources. Faced with the magnitude of this task, the need for automated processing is quite understandable: more than 13,000 notes and over 60 potential sources. In addition to the standard performance evaluation of the proposed approach in terms of error identification, the analysis of the results raises questions – relating to both the decisions made in the implemented algorithm and to missing sources, and perhaps even to errors of unknown origin that thus raise further questions (errors introduced voluntarily or involuntarily by the author, or differences in versions between those of the author and those referenced in the corpus studied?) – that are likely to enrich the historiographical scope of medieval encyclopedic knowledge.

The article by Maria Papadopoulos and Christophe Roche demonstrates the value of using ontoterminologies within the humanities to help define a core terminology and the structure of its associated concepts. They illustrate their argument with a study of ancient Greek clothing, by examining how experts in the field define and name the different garments or parts thereof. It is then possible to define a set of essential characteristics, or primitives, which the experts agree upon, and from which an ontoterminology can then be built. The authors present the TEDI platform as an example of software that can be used to achieve this goal.

Aurélien Benel's article outlines the history of early research in the 70s, particularly in France, focusing on the formalization and simulation of reasoning in archaeology. This clearly reflects current research on the semantic web, which, having originated during that period, was based on semantic artificial intelligence with the semantic

representation of knowledge by means of semantic networks, data frames, prototypes, etc. However, according the author – and what makes his work original, aside from the historical aspect' – his research was not merely aimed at creating a representation of meaning in order to simulate it, i.e. semantics, but also at determining the internal coherence of sign systems, i.e. semiotics. It is this tension between semantics and semiotics' – particularly the influence of the latter, which is directly linked to Digital Humanities' – that is highlighted in this article.

The article by Emmanuelle Bermès and Eleonora Moiraghi questions the broader use of Artificial Intelligence, and digital technology in general, for the various activities (collection, description, classification, storage and preservation, information and communication) of a heritage institution' – in this case, the Bibliothèque nationale de France [National Library of France, or "BnF"]. After having introduced the fields concerned by this study, the authors focus on three projects dealing with diverse corpora: (i) the future of online digital heritage via the internet archives of the Great War, (ii) the analysis of usage traces linked to the connection logs of the Gallica Digital Library, and (iii) the construction of semantic indexing through deep learning techniques in the BnF's iconographic document collections.

The article concludes with a presentation of the Corpus project, launched in 2016, which aims to "build a corpus supply service providing text and data searches for research". The authors' research outlines not only the various dimensions explored by Digital Humanities but also the interdisciplinary dialogue taking place in an operating environment supported by a prominent institution.

Jean-Gabriel Ganascia
LIP6, Sorbonne Université

Bertrand Jouve
LISST, CNRS, Université Toulouse Jean-Jaurès

Pascale Kuntz
LS2N, Université de Nantes