Bichitra: Online Tagore Variorum

Abstract

The digital edition of Rabindranath Tagore, *Bichitra: Online Tagore Variorum* aims to encompass all of Rabindranath's literary works in Bengali and English. This edition is an extremely interesting case study for the technical and logistical complexity of the task it sets out to achieve, which involved the development of customised technology and a visualisation strategy to engage the reader with textual variation at both a coarse-grained and a fine-grained level. The edition challenges the role of the literary archive to go beyond being a mere collection of texts and images, and to become instead a set of computational tools that allows the user not only to read but also to interact, explore and visualise the texts.

The Review

Bichitra is an ambitious project that seeks to create an online variorum edition of the works of Rabindranath Tagore (1861-1941), the Nobel Laureate Bengali poet and leading figure of the Indian Renaissance. The project claims to be the 'biggest integrated knowledge site devoted to any author, in any language to date'[1]. Indeed, it comprises most versions of nearly all of Tagore's works in Bengali and English.

A project of this size obviously needed a significant investment of time, energy and financial support to digitise and transcribe so many sources. Funded by the Government of India, as part of the celebration of the 150th anniversary of Tagore in 2011, Bichitra is led by Sukanta Chaudhuri, Professor Emeritus at Jadavpur University, Calcutta, who has collaborated with the School of Cultural Texts and Records (SCTR)[2]. Many other

---

[1] From Bichitra youtube presentation. <https://www.youtube.com/watch?v=GRNT9pf-sWA>

[2] The School of Cultural Texts website: https://web.archive.org/web/20140804101851/http://www.jaduniv.edu.in/ htdocs/view_department.php?deptid=135

important partners took part in the initiative[3], among whom Chaudhuri do not forget to mention the young collaborators who helped transcribing and encoding the texts, as to highlight the importance of the Bichitra experience for the formation of the Indian DH.

The uniqueness of Bichitra is a fact. In the light of the state of the art in the Indian DH, Bichitra is the first major project, launched on the international scale in 2013. The importance of explaining this initiative within its peculiar context of production is particularly evident, when assessing the quality of the research done, whose results are a nearly total customization. Bichitra is another 'digital island', whose technology was developed from scratch, in the attempt to address very specific contextual needs. First of all, the editor and his team had to face considerable problems to scan and transcribe approximately 137,520 pages, due to the many issues that Bengali handwriting poses for both existing OCR techniques and the textual encoding of the Bengali script. Secondly, the need to use a Bengali keyboard for entering the texts or the impossibility of finding or training enough operators in TEI, pushed Chaudhuri to adopt a custom encoding strategy. According to our opinion, these problems have been handled fairly well within the project by using a set of Unicode compliant Bengali fonts[4] and customised software tools to collate the sources. Among those, Bichitra's main technical achievement is the collation tool *Prabhed*, which takes center stage in the edition. This is a a complex system that detects variation on three different levels (section, segment, word), and visualise the results in form of interactive colour-coded bars, that users can manipulate. On the whole, this is the largest digital collection of Tagore's literary sources, which provides the users with thousands of images and transcriptions and a custom-built data mining system.

The strong Orange-Saffron homepage might at first seem an unconventional template option, if compared to the prevalence of white/gray/pastel tones, typical of other digital scholarly editions or publishing websites. Here Saffron stands for the II chakra, the colour of the liberation of earthly passions, made even brighter by the combination of yellow coloured icons. On the top, a red navigation bar provides multiple accesses to the content, together with a link to specific guidelines for each section. Due to the complexity of the edition, the guidelines are a recommended starting point for the scholar who is approaching to the website for the first time. They provide a detailed description of how the system works, explaining the criteria of inclusion/exclusion, the transcription and other useful how-to-use tips. For those who are eager to see the materials, on the left side of the homepage there is a window from which users have direct access to resources. The multiple access design of Bichitra allows the user to explore the content from different

---

[3] see the CREDITS.

[4] Developed by OmicronLab based in Dhaka, Bangladesh.

points of view. One can filter the sources by genre, title, or just by browsing the alphabetical list and clicking on the title. Once a manuscript has been selected, for example, its image is displayed together with its transcription. The texts are interlinked at the document level and can be also explored/manipulated in different ways, simply by switching from the page-image view to the filtered transcription — a final text for the end-user, or directly launching Prabhed. While the search engine is basic, the collation software is a complex system that generates a layered series of visualisations and is definitely the most original feature of the project . Essentially, Prabhed allows the user to interact with the multiple versions, and to form an impression of the patterns of variation within them at an abstract graphical level, while gradually leading up (or perhaps more in keeping with the metaphor of 'distant-reading', digging down) to the actual texts themselves.

Due to the lack of an editorial statement, I decided to interview Chaudhuri, taking the opportunity to ask a few questions about the philosophy behind his representation of Tagore. In keeping with the project's title — Bichitra means 'the various' in Bengali — Chaudhuri specifies that 'the textual/editorial philosophy behind Bichitra is inclined towards a 'fluid and radial representation' of the literary phenomenon, which is the very nature of Tagore's textual tradition, as 'he wrote and rewrote so extensively that no single work is preserved in less than two or three different versions'. According to his words, 'the conglomerate of versions of a particular work available on Bichitra may be said to be the work in its multiplicity: no single text, nor any new text extracted from one or more texts can replace the full textual range of Tagore's work, which constitutes its true and full identity'[5].

Chaudhuri's words help to better define the rationale behind the edition, which on the whole has been called a successful experiment in the Digital Humanities. There are certainly things that could be improved, especially regarding the usability of Bichitra. In any case, one should keep in mind that Bichitra is an extremely ambitious project and the biggest problems with which the editors have struggled are generally unsolved. For example, many humanities scholars coming to the archive with traditional expectations may feel that the design impedes access to the texts, but as far as I know, despite the essential role that comparing different versions of the text plays in fluid editions, there are no satisfying solutions to display more than three versions on the screen in a really usable way — that is offering the high control that specialized users want. That is probably the main reason why Prabhed seems still fairly difficult to use when high-level differences are computed (the figure shows how the page looks when comparing segments of a poem). As can be seen, it is quite hard to keep track of so many colours, especially as it gets much more complicated when comparing segments of a longer work, for instance

---

[5] From my interview.

from the genre fiction. Nonetheless, the interface provides a way to compare texts side by side – here is another example from the play – that doesn't really help in the case of more than three versions, since no more than three pop-up windows can fit on the screen.

Secondly, some may feel that the use of TEI encoding would have made the materials more accessible. To those who may question the validity of his approach Chaudhuri explains, that this was not only dictated by contextual needs but philosophical reflections linked to important epistemological and cultural themes. Apart from his skepticism about nested markup, when he wonders 'if textual computing can ever free itself completely from mark-up …not only for transcription but for all kinds of search and retrieval? (I am talking chiefly of encoding for content, not encoding for layout.) Can we register attributes by indexing rather than encoding? variants by parallel transcription and selective filtering?'[6] — from what Chaudhuri keeps saying about encoding, he goes further in envisaging the massive digitisation of texts in non latin characters, and the question then becomes to what extent existing standards can be adapted to the representation of heritages originating in a cultural environment very different from that in which encoding standards were originally designed.

One can only hope that as the archive moves forward, it will gain the momentum and financial support to further explore the computational potential of the Tagore corpus. With this in mind, Prof. Chaudhuri is trying to acquire funds for long-term maintenance and updating, and the team is already working on a newer version of Prabhed and, eventually, on improving its social presence[7]. To conclude, the potential of Bichitra lies in the significant realignment of what one might term the 'rules of engagement' between reader and text. They indicate an underlying conception of the textual archive as one that resonates between the dominant practices of textual encoding and editing, focusing on nuance, detail, and the emerging technique of 'distant reading', and promises to use emerging computational techniques to enable 'reading' at scale – i.e. to see the 'macro' and 'micro' approaches as interdependent (Jockers, Flanders 2013). Bichitra's approach to this dichotomy is a novel one. In fact, within this conception of the very large textual object (Hope, Witmore 2004), textual analysis is seen as the place for a fruitful transformation and interplay, which pays as much attention to scale as to the internal (between the corpus) and external relations of the materials. At the same time, it combines this editorial rigor with a broader vision of Tagore's literary corpus as a fertile ground for further computational intervention and analysis.

---

[6] From the lecture — DH 2015.

[7] From my interview.

Bibliography

Anderson, J. D. "Mr. Rabindranath Tagore on Bengali Prosody". *The Journal of the Royal Asiatic Society of Great Britain and Ireland*, (Oct., 1914), pp. 1038-1053. http://www.jstor.org/stable/ 25189251 (last accessed 30/07/2014).

Ed. by Bhattacharya, A., Renganathan M. "The Politics and Reception of Rabindranath Tagore's Drama: The Bard on the Stage", Routlege, NYC, 2015.

Cerquiglini, B. "Eloge de la variante", Paris : Seouil, 1989.

Eggert, P. "What we edit and how we edit; or, why not to Ring-Fence the Text" on Ecdotica 10(2013).

Fiormonte D**.** "Towards a Cultural Critique of the Digital Humanities". Historical Social Research, Vol. 37 (2012) p. 59.

Fiormonte D., Martiradonna, V., Schmidt, D. "Digital Encoding as a Hermeneutic and Semiotic Act: The Case of Valerio Magrelli", **<**http://www.digitalhumanities.org/dhq/vol/4/1/000082/000082.html**>**

Fischer, F. "The Pluralistic approach. The first scholarly edition of William of Auxerre's treatise on liturgy", http://computerphilologie.digital-humanities.de/jg08/fischer.html#FN1

Hope J., Witmore, M. "The very Large Textual Object: A Prosthetic Reading of Shakespeare", 2004 <http://extra.shu.ac.uk/emls/09-3/hopewhit.htm>

Jockers M. J., Flanders, J. "A matter of scale" *Faculty Publications - Department of English*, March 18, 2013 <http://digitalcommons.unl.edu/englishfacpubs/106.> (last accessed 30/07/2014).

Mordenti R. "Prospettive e problemi per l'edizione critica digitale dello Zibaldone Laurenziano (Plut. XXIX, 8) di Giovanni Boccaccio" Accademia dei Lincei, 20 June 2012.http://www.lincei.it/ files/centro_linceo/Rela_Mordenti.pdf (last accessed 30/07/2014)

Nelson, T. H. "Embedded Markup Considered Harmful." October 2 (1997). http://www.xml.com/ pub/a/w3j/s3.nelson.html.

Olivelle P. 1998, "Unfaithful Transmissions Philological Criticism and Critical Editions of the Upanisads", *Journal of Indian Philosophy*, Vol. 26, No. 2 (April 1998), pp. 173-187.

Pierazzo, E. "The Encoding of Time in Manuscripts Transcription: Towards Genetic Digital Edi- tions," in Schmidt, S, Siemens, R., Kumar, A. and Unsworth, J. (eds), *Digital Humanities 2007 Conference Abstracts*, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign.

Pound, E., "On Rabindranath Tagore". http://fortnightlyreview.co.uk/2013/04/rabindranath-tagore/ (last accessed 30/07/2014).

Schillingsburg, P. "Literary documents, texts and works represented digitally", on Ecdotica Journal 10(2013).

Schmidt D. "The inadequacy of embedded markup for cultural heritage texts". Literary and linguistic Computing. 25, 3 (2010) p. 337.

Schmidt, D. "Towards an Interoperable Digital Scholarly Edition", *Journal of the Text Encoding Initiative* [Online], Issue 7 | November 2014, Online since 12 November 2014, connection on 28 January 2015. URL : http://jtei.revues.org/979 ; DOI : 10.4000/jtei.979

Schmidt, D. "The role of markup in the digital humanities". http://www.cceh.uni-koeln.de/files/Schmidt_final.pdf

Yeats, W. B., "Introduction to the Gitanjali". <http://www.sacred-texts.com/hin/tagore/gitnjali.htm> (last accessed 30/07/2014)