Towards a computationally aware approach to humanistic data interfaces

Sollazzo, Anna
2805036S@student.gla.ac.uk
University of Glasgow, United Kingdom

Mapping human understandings of structure and processes to a machine interpretable form is an inevitably lossy process. This is especially true in humanities contexts where the specificity, ambiguity, and complexity valued in humanistic inquiry clash with the systematicity, determinism, and normalisation fundamental to computational logics. Data interfaces are, by definition, a site of this conceptual negotiation; however, as highlighted by Johanna Drucker, this process can be collapsed and invisibilized by mechanisms that emerged from engineering traditions and that correspondingly perform a reification of information consistent with an epistemological tradition that tends to privilege axioms over interpretations. Much of Drucker's work in elaborating more humanistic visualisation protocols has focused on the addition of visual channels—abstractions—as a means of expressing the positionality, ambiguity, and constructedness of the information presented (Druker 2011). Media theorist Anthony Masure argues, to the contrary, that the reliance of interfaces on compounded layers of abstraction (Cellard / Masure 2018) is precisely the reason that they are "rarely reflective, in their design, of epistemological traditions [between computing and the humanities]" (Masure 2018). Computational abstractions are mechanisms designed to lessen the dissonance between human conceptualisations and machine models; it is their dissection that stands to reveal the 'lost' interpretative process.

One facet of the Digital Humanities is the application of digital techniques to humanities problems. Fundamental to this process is grappling with the way in which these questions must be transformed in order to fit into computational paradigms, and how this can influence and reshape the research, in ways that can often pass unnoticed without explicit reflection. This presentation will explore the potential of a data interface that deliberately forefronts computational processing to interrogate this mapping. It will first describe three identified overarching characteristics of humanistic interface design—transparency, generativity, and interpretability—and provide examples of how they manifest in a concrete context using the data tools produced by long term DH initiative, the Comédie-Française Registers Project. It will then turn to the design and implementation of CLAIRON (Figure 1), a proof of concept interface in this alternative style applied to the same data. CLAIRON is inspired by natural language interfaces for databases (NLIDB), but unlike other tools in that class, it does not attempt to make the question to query translation appear in any way seamless or self evident.

Figure 1: Natural language input and query translation with highlight mapping for the input *Quelles sont les comédies en trois acres ou en cinq actes qui sont jouées moins de 15 fois et ont une recette moyenne de plus de 900 livres ?*

Using glassbox scaffolding (Hmelo / Guzdial 2005), the interface focuses on engaging with what David Berry terms the "underlying computationality" (Berry 2011) of objects of digital scholarship by exposing and encouraging the user to actively engage with the complexity and compromises of each step of the process—for example, entity resolution (Figure 2) and structure inference (Figure 3)—and to confront the limits of machine 'understanding'. The discussion will touch on the salient technical details of implementation; look at how CLAIRON approaches the core challenges of any NLIDB system, entity mapping and join path inference (Kim et al. 2020); explain the mechanisms used to walk the line between polymorphism and domain specificity; examine the ways in which it addresses the aforementioned characteristics of humanistic interfaces; and highlight sources of error, deliberate exclusions and limitations, and possible extensions.

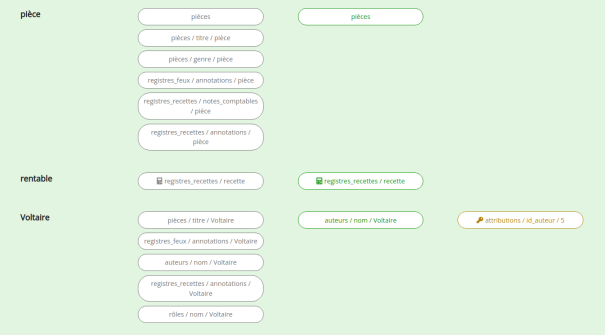

Figure 2: Entity match lists and selections for the input *Quelle est la pièce la plus rentable de Voltaire ?*.

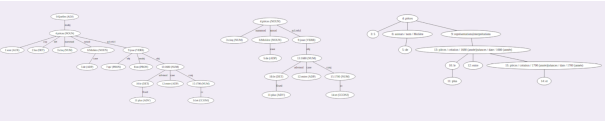

Figure 3: The first three parse and query trees used to structure the input *Quelles sont les cinq pièces de Molière qu'on joue le plus entre 1680 et 1700 ?*.

The final piece of the presentation will speak to how CLAIRON relates to DH pedagogy. On a concrete level, it will highlight how different elements of the interface lend themselves to its use as a learning support for databases. More abstractly, it will open discussion around the key competencies of humanities computing education. Using Micheal Mateas' notion of procedural liter-

acy (Mateas 2005), it will argue that the focus on computational procedures and interrogation of model mapping of the interface theorisation underpinning CLAIRON is, to an extent, absent from current practices, but essential to developing a sufficient understanding of computing in the abstract so as to be able to reshape and subvert its rules and norms in order to enact a practice of computing that fundamentally aligns with humanistic epistemologies and values.

Bibliography

Berry, David M. (2011): "The computational turn: Thinking about the digital humanities", in: *Culture machine* 12.

Cellard, Loup / **Masure, Anthony** (2018): "Le design de la transparence", in: *Multitudes* 4: 100-111.

Drucker, Johanna (2011): "Humanities approaches to interface theory", in: *Culture machine* 12.

Drucker, Johanna (2011): "Humanities approaches to graphical display", in: *Digital Humanities Quarterly* 5, 1: 1-21.

Hmelo, Cindy E. / **Guzdial, Mark** (1996): "Of black and glass boxes: scaffolding for doing and learning", in: *Proceedings of the 1996 international conference on Learning sciences*: 128-134.

Kim, Hyeonji / **So, Byeong-Hoon** / **Han, Wook-Shin** / **Lee, Hongrae** (2020): "Natural language to SQL: where are we today?", in: *Proceedings of the VLDB Endowment* 13, 10: 1737-1750.

Masure, Anthony (2018): "Vers des humanités numériques «critiques»" https://dlis.hypotheses.org/2088 [04.11.2022]

Mateas, Michael (2005): "Procedural literacy: educating the new media practitioner", in: *On the Horizon* 13, 2: 101-111.