Exploring, Browsing and Interacting with Multi-Scale Structures of Knowledge

Quentin Lobbé, David Chavalarias and Alexandre Delanoë

Exploring, browsing and interacting with multi-scale structures of knowledge

Quentin Lobbé · David Chavalarias ·
Alexandre Delanoë

Abstract Since the dawn of humanity, writing has been one of the first mnemotechnology: a technique not only designed to fix a thought on a medium but also a dynamic tool for the elaboration of a collective memory. Written texts can thus be considered as vectors of knowledge as well as providers of socio-historical contexts. The accumulation of Mesopotamian clay tablets (4000 BC) or the elaboration of the Vivarium library (535 AD-555 AD) gave early evidence of a growing will to collect and provide access to isolated elements of knowledge. Later on, with the transition from manuscript to book, textual contents outgrew erudite communities and started to touch all layers of the population, up to the present day: we are now daily fueled by unlimited flows of articles, novels, messages, tweets, etc. The recent ICT revolution has given birth to an unprecedented world of digital traces and has impacted a wide number of knowledge-driven domains such as education or policy making.

Science, in particular, has been one of the first area to experiment this digital shift. Databases of scientific publications are scaling up and it is now possible to dive into the amazing richness of most of these catalogs. Qualitative sciences are also taking advantage of the ICT revolution by integrating large cultural data sets (digitized historical documents, social networks footprints, archived Web sites, etc.) within their own scopes of analysis. Digital-born fields of research have thus emerged at the crossroad between social and computational sciences. But whether we speak of digital humanities or cultural analytics, it seems that all these domains end up facing the same issue: how to navigate across growing constellations of words and texts?

Because there are structures inside knowledge, a given text can always be studied in relation to others or in light of a specific socio-cultural context. By way of textual traces, human beings are calling out to one another: citations, retweets, controversy, etc. We are the architects of a giant web of elements of knowledge whose very structures and shapes convey information of their own [1]. Like ants or bees, through the aggregation of individual con-

Complex Systems Institute of Paris Ile-de-France (ISCPIF - CNRS)

tributions, we collectively achieve complex tasks that are out of the reach of individuals. This phenomenon is called *collective intelligence* and relies on a core mechanism called *stigmergy*; that is, the indirect coordination between an agent and an action through the environment. From scientific archives to Web pages and online ratings, our digital societies are literally embedded in a stigmergic environment.

Nowadays, complex systems approaches enable us to reconstruct the collective shapes and ontogeny of large corpora of texts. We call *reconstruction methods* all techniques implemented to understand a complex object or natural phenomenon by means of both the observation of patterns and the analysis of processes. Such methods are part of the larger family of *phenomenological reconstructions*, designed to find reasonable approximations of the structure and dynamics of a given phenomenon. Reconstruction methods can be summarized by the generic workflow $O \rightarrow R \rightarrow V$, where O represents a complex object associated to a set of properties. Based on a collected data set, O is next reconstructed as a formal object R described in a high-dimensional space on the basis of a collected data set. The process ends with the dimensional reduction of R, so that it can be projected as a human-readable visualization V. In this presentation, we focus on the *phylomemy reconstruction* method which consists in reconstructing inheritance networks of elements of knowledge on top of timestamped corpora of textual documents [2]. But while the most recent research have investigated the multi-level and multi-scale properties of phylomemies in R [3], the question of their visualization in V remains an open challenge. How can we explore, browse and interact with wide elements of knowledge through a phylomemy? How can we translate their inner processes in a graphical way?

In this presentation, we'll first review the state of the art of *text analysis and knowledge visualization* methods. We will then focus on the field of *science dynamics* before drilling down through the phylomemy reconstruction method. In doing so, we will highlight the inner properties and components of the phylomemies we aim to visualize. After having investigated the notions of synchrony and diachrony, we will define two complementary axes of visualization: the seabed view and the kinship view, and discuss ways to interact with them through micro, mezzo and macro lenses of exploration. Using the free software *Gargantext*, we will then develop our approach thanks to a new open source vizualisation system called *Memiescape* and validate it by browsing through the historical landscapes of various domains of research (see Fig. 1). We will end up by introducing a generic methodology for browsing through multi-scale structures of knowledge.

Keywords Phylomemy reconstruction · Knowledge visualization · Science map · Co-word analysis
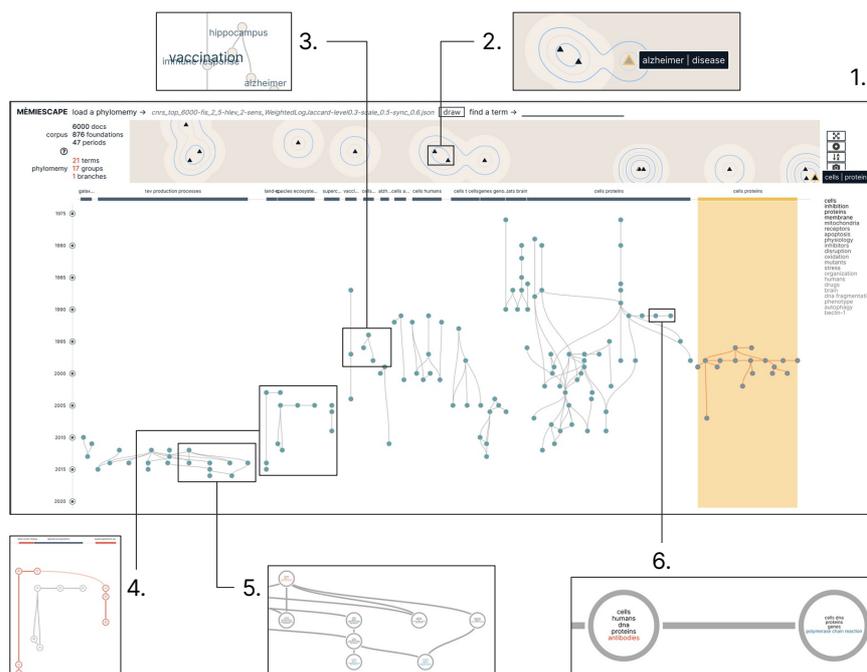
Fig. 1 Screenshots taken from *Memiescape* to visualize the phylomemy of a corpus gathering 6000 top cited papers extracted from the *Web of Science* (WoS). Written between the 1980s and the present day, at least one of each publication's authors is affiliated to the french CNRS.

References

1. Chavalarias David, Formes collectives, Le Genre humain, 1, 145–152 (2019)
2. Chavalarias David and Cointet Jean-Philippe, Phylomemetic patterns in science evolution—the rise and fall of scientific fields, PloS one, 8, 2 (2013)
3. Chavalarias David and Lobbé Quentin and Delanoë Alexandre, Draw me Science - multi-level and multi-scale reconstruction of knowledge dynamics with phylomemies (forthcoming)