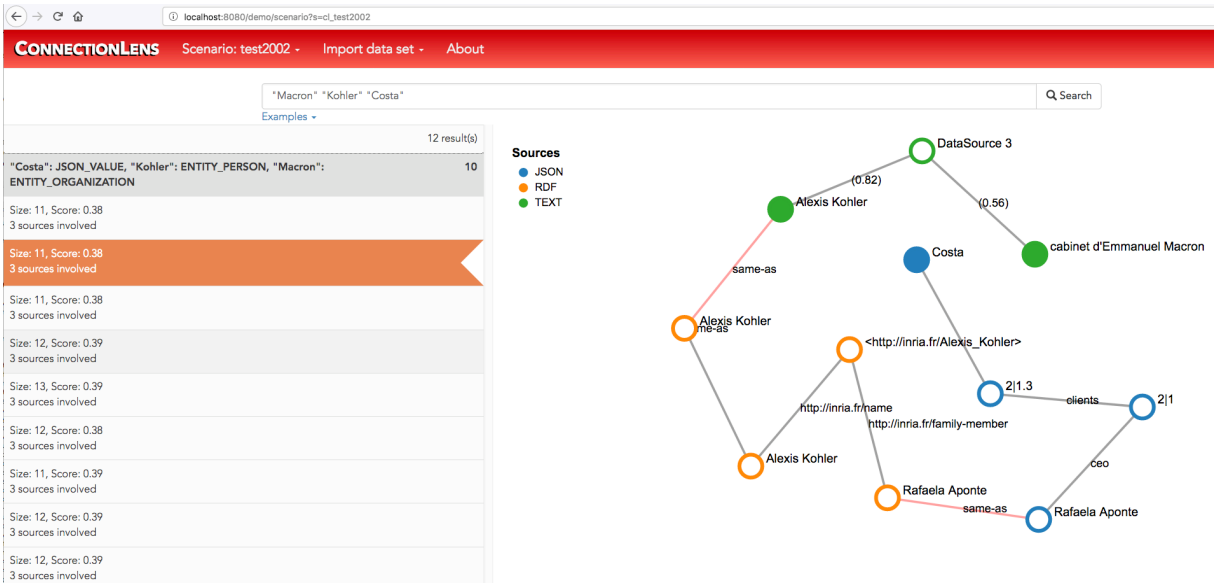
Interactive visualizations of heterogeneous Big Data for data journalism
(Ioana Manolescu and Emmanuel Pietriga)

Data journalism projects often need to find *connections* across heterogeneous data sources, such as text documents and structured databases (relational, RDF, JSON, etc.). Such connections may enable, for instance, to "*find all the private companies involving members of the Parliament or people closely associated to them*" by exploiting, on one hand: a set of text documents describing the involvement of people in contracts; and on the other hand, a structured database specifying the members of the parliament and possibly their relatives.
ConnectionLens [1] is designed to enable data journalists to find connections between user-specified search terms across heterogeneous data sources. ConnectionLens treats a set of heterogeneous, independently-authored data sources as a single virtual graph, in which nodes represent fine-grain data items (relational tuples, attributes, key-value pairs, RDF, JSON or XML nodes) and edges correspond either to structural connections in the data, or to similarity (sameAs) links. ConnectionLens is developed as part of an ongoing collaboration with Les Décodeurs, the fact-checking team of the Le Monde newspaper.



Screen shot of the current ConnectionLens interface: it displays search result but is not interactive and it does not allow exploring the data. The proposed topic is to devise novel visualization techniques for such heterogeneous data.

ConnectionLens is capable of finding and exploiting connections present across heterogeneous data sources without requiring users to specify any join predicate. One significant challenge, however, is **to enable users to understand and efficiently explore the multivariate network [2] formed by these connections** between sets of heterogeneous entities using interactive visual exploration techniques.
The student will be tasked with designing and prototyping solutions for incrementally populating a canvas with elements of interest (entities of various types, connections of different kinds) using direct manipulation, and choosing appropriate visual mappings for those different types of entities on-the-fly following the expressive design approach we have been researching recently [3], possibly investigating the use of motion as an

additional encoding variable [4] and emphasizing changes in the diagram using dynamic network visualization techniques.

Advisors: Ioana Manolescu (ioana.manolescu@inria.fr, Inria and LIX, Turing building, office 1029) & Emmanuel Pietriga (emmanuel.pietriga@inria.fr, Inria and U. Paris Saclay)

References

1. Camille Chanial, Rédouane Dziri, Helena Galhardas, Julien Leblay, Minh-Huong Le Nguyen, and Ioana Manolescu. 2018. Connectionlens: finding connections across heterogeneous data sources. Proc. VLDB Endow. 11, 12 (August 2018), 2030-2033. http://dx.doi.org/10.14778/3229863.3236252, https://hal.inria.fr/hal-01841009
2. Carolina Nobre, Miriah D. Meyer, Marc Streit, and Alexander Lex. The state of the art in visualizing multivariate networks, Computer Graphics Forum, vol. 38, no. 3, pp. 807, 832, 2019. https://vdl.sci.utah.edu/publications/2019_eurovis_mvn/
3. Hugo Romat, Caroline Appert, and Emmanuel Pietriga. 2019. Expressive Authoring of Node-Link Diagrams with Graphies. Under review at IEEE Transactions on Visualization and Computer Graphics. https://hugoromat.github.io/graphies/
4. Hugo Romat, Caroline Appert, Benjamin Bach, Nathalie Henry-Riche, and Emmanuel Pietriga. 2018. Animated Edge Textures in Node-Link Diagrams: a Design Space and Initial Evaluation. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA. DOI: https://doi.org/10.1145/3173574.3173761, https://hal.inria.fr/hal-01726358
5. Benjamin Bach, Emmanuel Pietriga, and Jean-Daniel Fekete. 2014. GraphDiaries: Animated Transitions and Temporal Navigation for Dynamic Networks. IEEE Transactions on Visualization and Computer Graphics 20, 5 (May 2014), 740-754. DOI: http://dx.doi.org/10.1109/TVCG.2013.254, https://hal.inria.fr/hal-00906597