European Commission | Horizon 2020 European Union funding for Research & Innovation

opencare

Deliverable 5.2: Toolbox for developing networkbased software for collective intelligence

Project Acronym	OPENCARE
Title	Open Participatory Engagement in Collective Awareness for REdesign of Care services
Project Number	688670
Work package	WP5 – Data processing for aggregating collective intelligence processes
Lead Beneficiary	UBx — University of Bordeaux
Editor(s)	Guy Melançon *University of Bordeaux*
Reviewer(s)	Guy Melançon *University of Bordeaux* Bruno Pinaud *University of Bordeaux* *Alberto Cottica* *Edgeryders*
Dissemination Level	Public
Contractual Delivery Date	31/12/2016
Actual Delivery Date	27/12/2016
Version	1.0
Status	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

Summary

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

Deliverable 5.2: Toolbox for developing networkbased software for collective intelligence

This document may be seen as a toolbox to help design the SNNA dashboard (as well as helping the design and development of the semi-automated aid to Open Ethnographer).

The toolbox comprises a review of key literature (strongly inspired and borrowing from (Renoust 2013)[*]), and points at algorithms and visualization/interaction techniques judged to be most useful.

SSNA and Visual Analytics

Visual Analytics is "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas & Cook 2006). Searching for information is actually even more than just finding it.

When searching, we find, we learn, and naturally explore the neighborhood of those findings (Marchionini, 2006). Users process data in an incremental and iterative manner (Springmeyer et al., 1992). They iteratively:

- Interact and explore with representations,
- Further investigate (or confirm insights)
- Organize data, filter, customize representations
- Express ideas (keep record of data transformations, observations, etc.) insights).

Interacting with visual representations facilitates the analysis of complex problems in order to detect known features (the expected) and discover new insights (the unexpected) in data (Wong and Thomas, 2004). Network visualizations improve visual discovery, with their ability to make users wonder about the relationships they observe, leading to "why" questions (Tamassia, 2013). Other user-centric, such as Dörk et al. (2011)'s *flaneur*, puts forward exploration as guiding principle in information seeking. Visual exploration is also a perfect way to deal with heterogeneous and noisy data as it is intuitive and requires no understanding of the complex mathematical and algorithmic procedures (Keim, 2002).

Sense making

Making sense of visualization is a complicated process that involves multiple factors, understanding of conventions in the visualization design, manipulations of

[*] Benjamin Renoust, affiliated with the National Informatics Institute (NII) in Tokyo has been listed as part of close collaborators to UBx in DoA (part B).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

the information, and general abstraction of the acquired knowledge. The different gaps in perception, and especially rationale gap underlined by (Amar & Stasko 2004), are critical issues any visualization must address. This rationale gap necessarily emerges between the actual relationship (at the data level) and the perceived relationship between elements in the visualization. Coping this gap consists in "being able to explain confidence in the relationship, as well as its usefulness". This analytic gap moreover may suffer from any distortion graphical representations inevitably carry, often inherited from data transformation or geometrical projections.

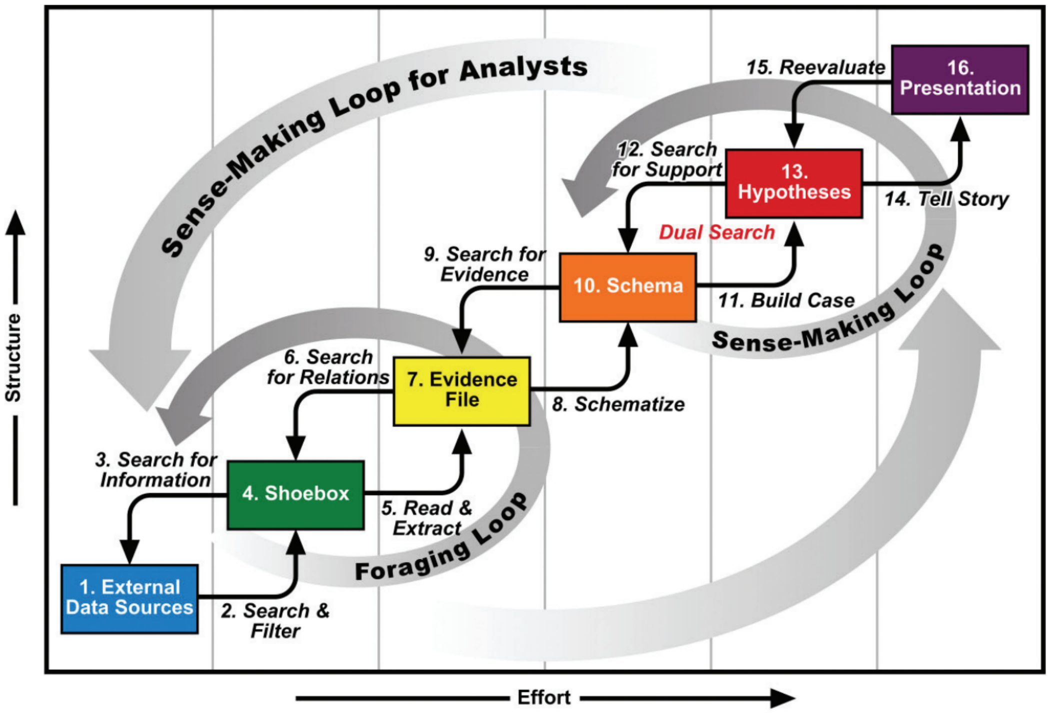

Figure 1. The sense-making loop for analytics by (Thomas & Cook 2005), adapted from (Pirolli & Card 2005).

This sense-making process is recursive and iterative. Data is typically examined at different levels of granularity. The process can also be bottom-up – from data to conclusions – or top-bottom – confirming hypotheses or theories with data. Users typically combine these approaches depending on findings and opportunities.

(Card et al., 1999) claim information visualization enhances cognitive abilities by:

- "increasing the memory and processing resources available to the users": encoding of the information and making it comparable, e.g. the length of a set placed on a node size
- "reducing the search for information": showing order where none obviously appeared, e.g. proximity between elements
- "using visual representations to enhance the detection of patterns": e.g. a force directed network diagram displaying tight communities
- "enabling perceptual inference operations": e.g. correlations of two variables when plotted

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

- "using perceptual attention mechanisms for monitoring": e.g. animation, colour coding, such as blinking red indicators for potentially threatening or dangerous situations
- "encoding information in a manipulable medium": e.g. allowing interaction through many means such as selections or scales

The previous cited literature thus underlines how visual analytics can support information seeking and analytical reasoning.

A more pragmatic perspective is that of (Munzner 2009) who proposed a model for the design of visualization systems. We refer the reader to deliverable 5.1 for a detailed description of Munzner's work that proves quite valuable for us.

Task taxonomies

We conclude this literature review by listing taxonomies of tasks supporting analytical reasoning (again borrowing from (Renoust 2013)), whether it's in the field of human-computer interaction, cartography, information retrieval, visualization, or visual analytics. (Meyer *et al*. 2012) highlights the gap in the literature between low and high-level tasks classifications, which strongly motivated the excellent work of (Brehmer & Munzner 2013), which describes a typology for visualization tasks at any level of abstraction.

Amar et al. (2005) provides a set of domain-free low-level analytic tasks from an extended set of user behavior observations. These tasks are centered on numerical operation and data selection upon which users achieve reasoning.

- Retrieve Value, find attributes of a data subset
- Filter, find a data subset satisfying some conditions on attributes
- Compute Derived Value, compute a numeric representation of a data subset
- Find Extremum, find a data subset with extreme attribute value in the observed data (sub)set
- Sort, rank a subset of data given an attribute
- Determine Range, find the span of an attribute value given a data (sub)set
- Characterize Distribution, characterize the distribution of an attribute given a data (sub)set
- Find Anomalies, given criteria and relationships, find the anomalies in a data (sub)set
- Cluster, group subsets of data of similar attributes
- Correlate, determine relationships between attributes among a data (sub)set

Amar and Stasko (2004) refer to *analytic gaps* as obstacles between visualization and higher-level tasks which more or less depict the gap mentioned in (Meyer et al., 2012) — on two levels *Rational gap* and *WorldView gap*, and proposes visualization tasks to avoid falling into these gaps.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

> The Rational gap refers to the difference between a perceived relationship through the visual system, and the ability of such system to explain reasons of this apparent relationship.

To bridge this, Amar and Stasko propose the following tasks:

- Expose uncertainty in measures and outcome, for example by showing population, average and standard deviation of a statistical summary.
- Concretize relationships by explicitly displaying them, for example linking and brushing among derived data highlighting the source data.
- Formulate cause and effect possibly allowing live editing to investigate effects from causes.

> TheWorldview gap is the gap between what is displayed from what needs to be displayed (including the limitations of a display).

The authors propose then these tasks:

- Determination of domain parameters
- Multivariate explanations
- Confirm hypotheses

(Buja *et al*. 1996) describe some multivariate data tasks for interactive visualization. Although they might be part of the visual analytics culture now and feel quite basic, these higher-level tasks still remain a good reference:

- Finding structural patterns: focusing individual views in order to figure out some patter or overview (choosing data subsets/ attributes/visual parameters)
- Dynamic queries: linking multiple views in order to highlight relationships (search for an element in all views, brush among axes and highlights in the other views)
- Making comparisons: arranging many views in a comparable way (with similar layouts, or a scatter plot matrix for example).

(Gotz & Zhou 2009) identified a whole catalog of 21 different low-level tasks from quantitative observations of users' behavior of Visual Analytics systems. Three categories have been identified.

- Exploration actions separates visual exploration tasks and data exploration tasks
- Insights actions manipulate the visual insights or knowledge insights taken from exploration, and
- Meta actions operates at the application level on user's activity).

More recently (Heer and Shneiderman, 2012) provides three high-level categories of tasks and 12 different tasks:

- Data and view specification (visualize, filter, sort, and derive)
- View manipulation (select, navigate, coordinate, and organize), and
- Analysis process and provenance (record, annotate, share, and guide).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

The categories correspond to high-level tasks, and are critical to enable iterative visual analysis. These tasks are relevant for visualization creation, interactive querying, multiview coordination, history, and collaboration.

Although the tasks listed are described at different levels, they all are relevant for visual analytics.

It is with these taxonomies in mind, and the guiding principles presented above that we framed our work on designing the opencare SSNA dashboard.

Network Visualization

Graphical representations of networks are based on graph models and focus on interactions between entities. Two types of representations are widely used for general graphs (as opposed to specific graphs such as trees): matrix views and node-link diagrams. Each have advantages and deficiencies, mostly depending on structural properties of the networks that are visualized (Ghoniem *et al*. 2005) (Holten *et al*. 2011).

Matrix views rely on the representation of a graph's adjacency matrix as a graphical object. Instead of showing number in rows and columns, they can encode relationships with colours or intensity. As mentioned in deliverable 5.1, we haven't considered matrix views mainly due to user preference. We may come back to these representations later if needed.

The node-link diagram has been supporting social network analysis in the literature as early as (Moreno 1934). It intuitively depicts relationships with nodes represented as glyphs and edges as lines or curves. These types of representations came as an obvious choice since our users were already accustomed to them. This type of representation benefits from the results in graph drawing research, focusing on efficient layout algorithms and aesthetic representations of networks (Tamassia 2013).

Network layouts

Many layout algorithms are available (Herman *et al*. 2000) (Tamassia, 2013) of which force-directed layouts are the most popular for general graphs. A force-directed layout (or also a spring-embedding, energy-based layout) algorithm aims to draw a graph only from its structural information. The principle is simple, is embeds forces in the nodes and links of graph and iteratively moves each node in the layout until it reaches a stable state.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

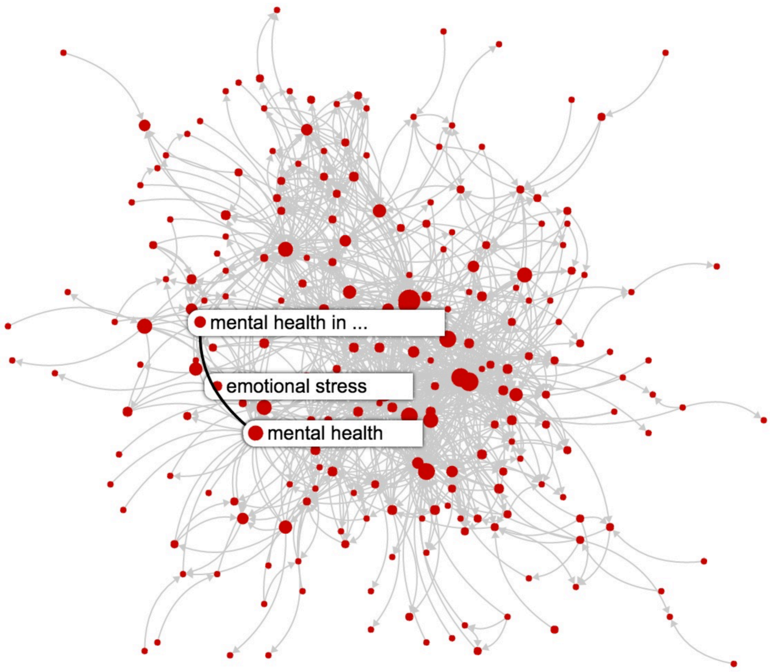

Figure 2. Typical force-directed layout of a graph.

Hierarchical layouts (Sugiyama & Tagawa 1981) deal with networks where nodes can be assigned to levels, starting from source nodes down to sink nodes. Tree layouts (Walker 1990) form an even more specific set of layouts applicable to trees (graphs with no cycles). This turns out to be relevant for opencare as discussion threads indeed show a tree structure.

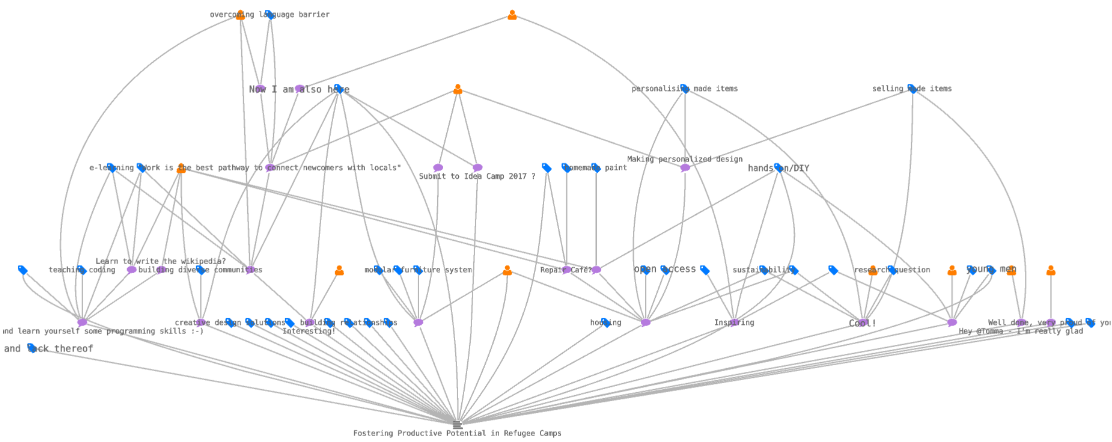

Figure 3. Typical hierarchical layout of a graph.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

Learning by examples

We adopted an approach where the previously stated theoretical results and principles, and guidelines were confronted with users in participatory sessions. To this end, users were invited early in the process to "play" with the data and provide feedback.

To this end, several tools were used and made available to users, mainly during live sessions (see deliverables D1.1 and D1.2).

Database views were built to ease the extraction of data (json format)

- Waiting for the dashboard: a script to quantify the OpenCare conversation
- Fetching OpenEthnographer codes and annotations in JSON format

The Tulip Graph Visualization Framework[†] has now become the standard tool used within the consortium (Masters of Networks sessions usually starts by having newcomers install the software and run a demo).

The scripting capabilities of Tulip allows to distribute data files and script to run to test ideas.

The Detangler application[‡] is also featured during live sessions to explore the use of a combined and synchronized view of both the social network and the tag co-occurrence network (see deliverable D5.1 for more details on what these networks precisely are).

Conclusion

The review of the literature thus provides guidelines and principles to fuel the design of the SSNA dashboard.

Additionally, the UBx team has a deep and experimented knowledge on graph drawing, and no the use of visual variables in an interactive environment.

The Detangler application (Renoust *et al*. 2015) serves here as a guide. We expect the dashboard to expand Detangler's capabilities by extending some of its feature or adding new ones.

References

Amar, R. and Stasko, J. (2004). A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations. In Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on, pp. 143–150.

[†] See http://tulip.labri.fr

[‡] See http://164.132.58.138:31497/ or http://detangler.labri.fr:31497

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

Amar, R., Eagan, J., and Stasko, J. (2005). Low-level components of analytic activity in information visualization. In IEEE Symposium on Information Visualization, 2005, pp. 111–117.

Brehmer, M. and Munzner, T. (2013). A multi-level taxonomy of abstract visualization tasks. In Proceedings of the 2013 InfoVis Conference, Transactions in Visualization and Computer Graphics. IEEE.

Buja, A. ; Cook, D. ; and Swayne, D. F. (1996). Interactive highdimensional data visualization. In Journal of Computational and Graphical Statistics, vol. 5, no. 1, pp. 78–99.

Card, S. K., Mackinlay, J. D., and Schneiderman, B. (1999). Readings in information visualization: using vision to think. Morgan Kaufmann.

Dörk, M., Carpendale, S., and Williamson, C. (2011). The information flaneur: a fresh look at information seeking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1215–1224.

Ghoniem, M., J.-D. Fekete, et al. (2005). "On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis." Information Visualization (Palgrave) **4**(2): 114-135.

Gotz, D. and Zhou, M. X. (2009). Characterizing users' visual analytic activity for insight provenance. In Information Visualization, vol. 8, no. 1, pp. 42–55.

Heer, J. and Shneiderman, B. . 2012. Interactive dynamics for visual analysis. In Queue, vol. 10, no. 2, p. 30.

Herman, I., M. S. Marshall, G. Melançon. (2000). "Graph Visualisation and Navigation in Information Visualisation: A Survey." IEEE Transactions on Visualization and Computer Graphics **6**(1): 24-43.

Holten, Danny, et al. (2011) "An extended evaluation of the readability of tapered, animated, and textured directed-edge representations in node-link graphs." 2011 IEEE Pacific Visualization Symposium, pp. 195-202.

Keim, D. A. (2002). Information visualization and visual data mining. In Visualization and Computer Graphics, IEEE Transactions on, vol. 8, no. 1, pp. 1–8.

Marchionini, G. (2006). Exploratory search: from finding to understanding. In Communications of the ACM, vol. 49, no. 4, pp. 41–46.

Meyer, M., M. Sedlmair, T. Munzner. (2012). The four-level nested model revisited: blocks and guidelines. Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization. Seattle, Washington, ACM: 1-6.

Moreno, J. L.. (1934). Who shall survive? A new approach to the problem of human interrelations., Nervous and Mental Disease Monograph Series, vol. 58, Nervous and Mental Disease Publishing Co.

Munzner, T. (2009). "A Nested Process Model for Visualization Design and Validation." IEEE Transactions on Visualization and Computer Graphics 15: 921-928.

Pirolli, P. and Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In Proceedings of International Conference on Intelligence Analysis, vol. 5, pp. 2–4.

Renoust, B. (2013). Analysis and Visualisation of Edge Entanglement in Multiplex Networks, Doctoral dissertation, University of Bordeaux.

Renoust, B., G. Melançon, et al. (2015). "Detangler: Visual Analytics for Multiplex Networks." Computer Graphics Forum **34**(3): 321-330.

Springmeyer, R. R., Blattner, M. M., and Max, N. L. (1992). A characterization of the scientific data analysis process. In Proceedings of the 3rd conference on Visualization'92, pp. 235–242. IEEE Computer Society Press.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.

Sugiyama, K., S. Tagawa, et al. (1981). "Methods for Visual Understanding of Hierarchical Systems Structures." IEEE Transactions on Systems, Man and Cybernetics **11**(2): 109-125.

Tamassia, R. (2013). Handbook of graph drawing and visualization. CRC Press.

Thomas, J. J. and K. A. Cook, Eds. (2006). Illuminating the Path: The Research and Development Agenda for Visual Analytics IEEE Computer Society.

Walker, J. Q. (1990). "A Node-positioning Algorithm for General Trees." Software — Practice and Experience **20**(7): 685–705.

Wong, P. C. and Thomas, J. (2004). Visual Analytics. In IEEE Computer Graphics and Applications, vol. 24, no. 5, pp. 20–21.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688670.