The Map Generator Tool*

Valeria Fionda[1], Giuseppe Pirrò[2], Claudio Gutierrez[3],

[1] Department of Mathematics, University of Calabria, Italy
[2] WeST, University of Koblenz-Landau, Germany
[3] DCC, Universidad de Chile, Chile

Abstract. We present the MaGe system, which helps users and developers to build maps of the Web graph. Maps *abstract* and represent in a *concise* and machine-readable way *regions* of information on the Web.

1 Introduction

The Web is a large and interconnected information space (usually modeled as a graph) commonly accessed and explored via navigation enabled by browsers. To cope with the size of this huge (cyber)space, Web users need to track, record and specify conceptual regions on the Web (e.g., a set of Web pages; friends and their interests; a network of citations), for their own use, for exchanging, for further processing. Users often navigate large fragments of the Web, to discover and isolate very few resources of interest and struggle to keep *connectivity information* among them. The idea of a map of a Web region is essentially that of *representing* in a *concise way* information in the region in terms of *connectivity* among a set of *distinguished* resources (nodes).

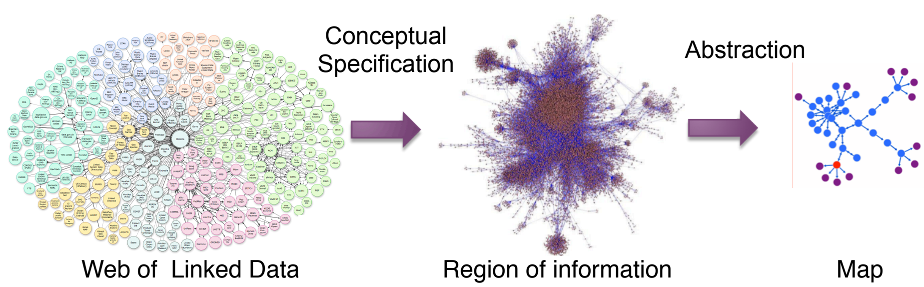

Fig. 1. Building maps of the Web.

With the advent of the Web of Data [4], maps to describe and navigate information on the Web in a machine-processable way become more feasible. The key new technical support is: *(i)* the availability of a standard infrastructure, based on the Resource Description Framework (RDF), for the publishing/interlinking of structured data on the Web; *(ii)* a community of developers; *(iii)* languages to specify regions of information on the Web [1]. Fig. 1 sketches the high level process of building maps of the Web via the Map Generator (MaGe) system.

* V. Fionda was supported by the European Commission, the European Social Fund and the Calabria region.

2 The Map framework

The idea of a map on the Web is to represent in a concise and comprehensive way connectivity information between pairs of distinguished nodes. Given a conceptual region G of information on the Web, there can be several maps of G with different level of detail (i.e., nodes and edges to be included).

Formally, let $\Gamma = (V_\Gamma, E_\Gamma)$ be a Web region, where V_Γ and E_Γ are the set of nodes and edges respectively. Then:

- $u \rightarrow v$ denotes an edge $(u, v) \in E$.
- $u \twoheadrightarrow v$ denotes a path from u to v in Γ.
- Let $N \subseteq V$. Then, $u \twoheadrightarrow_N v$ if and only if there is a path from u to v in Γ not passing through intermediate nodes in N.

Let $V_M \subseteq V$ be the set of distinguished nodes of the Web region $\Gamma = (V_\Gamma, E_\Gamma)$, i.e., those that we would like to represent.

Definition 1 (Map) *A* map $M = (V_M, E_M)$ *of* $\Gamma = (V_\Gamma, E_\Gamma)$ *is a graph such that* $V_M \subseteq V_\Gamma$ *and each edge* $(x, y) \in E_M$ *implies* $x \twoheadrightarrow y$ *in* Γ*.*

A basic (and highly used) example of map of the Web are bookmarks. In this case, V_M is the set of nodes highlighted or marked, and $E_M = \emptyset$, that is, there is no connectivity recorded among them. An important idea is that of a *good* map, i.e., a map which represents connectivity among the distinguished nodes and avoids redundant edges [3].

Definition 2 (Good map) *A* map $M = (V_M, E_M)$ *of* $\Gamma = (V_\Gamma, E_\Gamma)$ *is good if and only if:*
1. $\forall x, y \in V_M \; x \twoheadrightarrow_{V_M} y$ *in* Γ *implies* $x \rightarrow y$ *in* M
2. $\forall x, y \in V_M \; x \rightarrow y$ *in* M *implies* $x \twoheadrightarrow_{V_M} y$ *in* Γ*.*

Good maps have the nice properties *(i)* uniqueness and *(ii)* low complexity of computation. Indeed, given a region $\Gamma = (V_\Gamma, E_\Gamma)$ and a set of distinguished nodes $V_M \subseteq V_\Gamma$ there exists a unique good map $M = (V_M, E_M)$ that is computable in $\mathcal{O}(|V_M| \times (|V_\Gamma \setminus V_M| + |E_\Gamma|))$ by an adaptation of the BFS algorithm.

3 MaGe: Building Maps of the Web

Maps are built on top of regions of information on the Web. To automate the process of generating regions, MaGe uses the NautiLOD [1,2] language. Given an expression *ex*, NautiLOD enables to extract a Web region $\Gamma = (V_\Gamma, E_\Gamma)$ such that V_Γ and E_Γ are the set of nodes and edges *visited* while evaluating *ex*. Once the region has been obtained, MaGe computes good maps as sketched in Section 2 considering the set of distinguished nodes $V_M = \{s\} \cup T$, where s is the node in the region that corresponds to the seed URI where the evaluation of *ex* starts and T are the nodes satisfying *ex*.

MaGe has been implemented in Java and is available for download[4]. It includes two main modules: the *selection* and the *abstraction* modules. The first

[4] The MaGe website: `http://mapsforweb.wordpress.com`

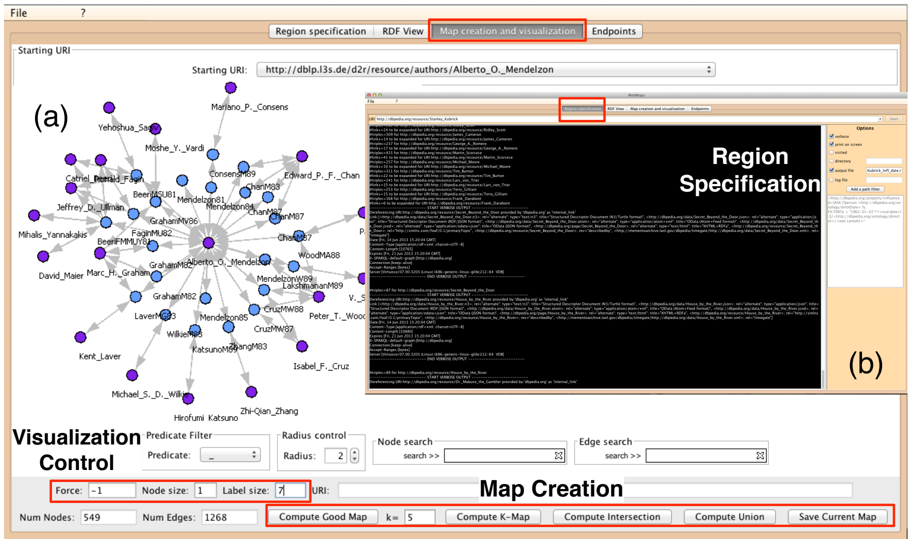

Fig. 2. The GUI of the MaGe tool.

one is responsible for the implementation of the NautiLOD language. In particular, given a seed URI and an expression, this module retrieves a Web region and a set of distinguished nodes. The second module, given the Web region and the set of distinguished nodes leverages the map framework to build maps. The decoupling between selection and abstraction enables to use the two functionalities also separately. MaGe is endowed with a GUI, which is shown in Fig. 2. It includes four main tabs. The first one (Fig. 2 (b)) is used to specify the region via a NautiLOD expression. The second and fourth display the region retrieved in RDF and the expression endpoints, respectively. The third tab (Fig. 2 (a)) deals with the creation of maps and their visualization. Both regions and maps can be saved in RDF allowing their storage, sharing, reuse and exchange. We now provide an example that we plan to show (along with others) in the demo. A video explaining how to use the tool is available at `http://youtu.be/BsvAiX3n968`.

Maps of Influence. An influence network is a graph where nodes are persons and edges represent influence relations. We leverage information taken from `dbpedia.org` and the property `dbpprop:influenced`.

Example 3 *Build a map of a region containing people that have influenced, or have been influenced by Stanley Kubrick (SK) up to distance 6. The distinguished nodes must be scientists.*

The region can be specified via the following NautiLOD expression. Here, the URI of SK in DBpedia (`dbpedia:Stanley_Kubrick`) is used as seed node:

```
dbpprop:influenced<1-6>[ASK {?p rdf:type dbpedia:Scientist.}]
```

In the expression, the notation `<1-6>` is a shorthand for the concatenation of (up to) six steps of the predicate `dbpprop:influenced`, while the `ASK` query in the test `[ ]` is used to filter the distinguished nodes (i.e., scientists).

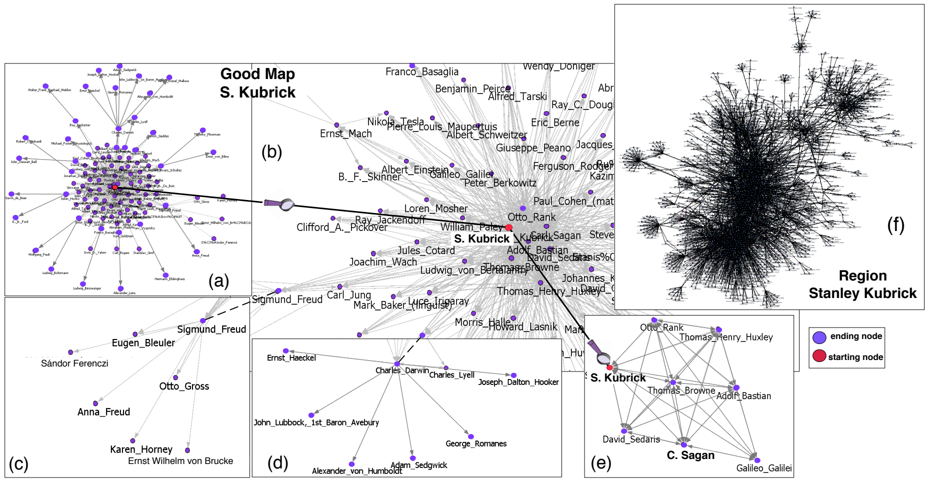

Fig. 3. Region (f) and good map (a) for SK with some zooms (b)-(e).

Fig. 3 (f) reports the region associated to the influence network of SK. The region contains 2981 nodes and 7893 edges. Indeed, it is very difficult to identify the distinguished nodes and more importantly connectivity among them and with the seed node. Fig. 3 (a)-(e) show the good map of this region (109 nodes; 2629 edges). The abstraction provided by the good map enables to identify the influence path, for instance, between SK and C. Segan (Fig. 3 (e)).

4 Conclusions

The availability of machine-processable information at a Web scale opens new perspectives toward the development of systems for the harnessing of knowledge on the Web. We contend that maps, key devices in helping human navigation in information spaces, are also meaningful on the Web space. They are useful navigation cues and powerful ways of conveying complex information via concise representations. Effectively, they play the role of *navigational charts*, that is, tools that provide users with abstractions of regions of information on the Web. We have implemented the MaGE system to generate maps. During the demo we will show maps in different domains including bibliographic networks.

References

1. V. Fionda, C. Gutierrez, and G. Pirrò. Extracting Relevant Subgraphs from Graph Navigation. In *ISWC (Posters & Demos)*, 2012.
2. V. Fionda, C. Gutierrez, and G. Pirrò. Semantic Navigation on the Web of Data: Specification of Routes, Web Fragments and Actions. In *WWW*, 2012.
3. V. Fionda, C. Gutierrez, and G. Pirrò. Knowledge Maps of Web Graphs. In *KR*, 2014.
4. T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 2011.