Eurographics/ IEEE-VGTC Symposium on Visualization 2010
G. Melançon, T. Munzner, and D. Weiskopf
(Guest Editors)

Volume 29 (2010), Number 3

Integrating Node-Link-Diagrams and Information Landscapes: A Path-Finding Approach

W. Kienreich & V. Sabol & C. Seifert

Know-Center Graz, Austria

Abstract

Knowledge repositories usually contain semantic structures as well as massive amounts of content. Information Landscapes have commonly been used to visualize large content repositories, and Node-Link Diagrams have commonly been used to visualize semantic structures. We present preliminary research on a path finding approach to the integration of Information Landscapes and Node-Link Diagrams. Our approach is based on a link-specific height map which reflects the semantic properties of the content in relation to the link. The paths generated by our approach tend to reduce the amount of contradictory visual evidence and to respect terrain features.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation I.3.8 [Computer Graphics]: Applications—

1. Introduction

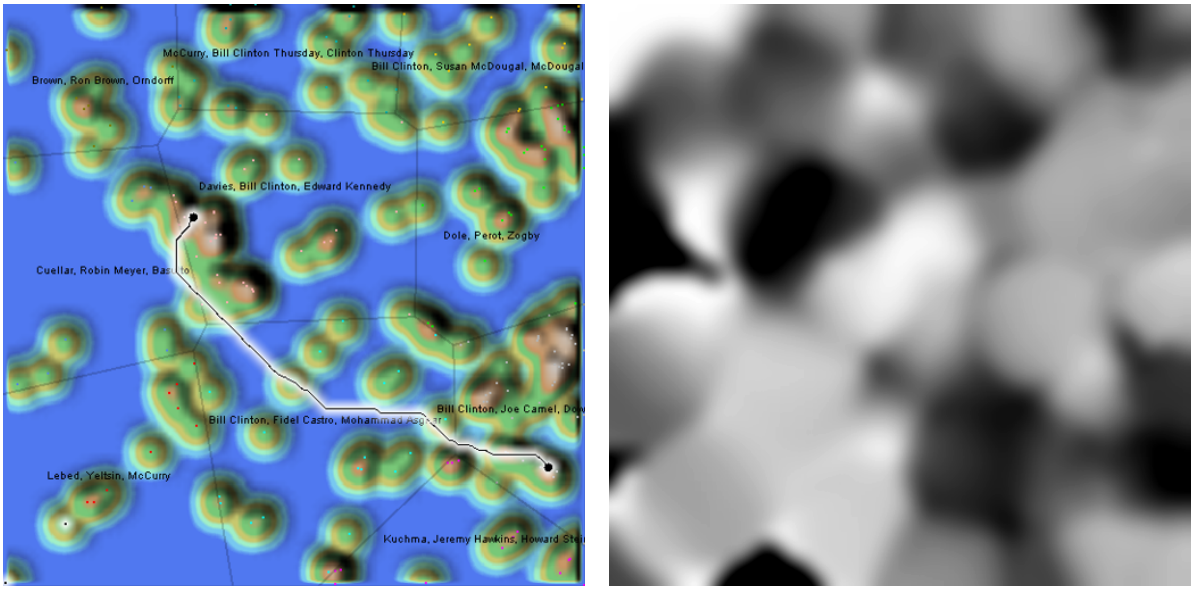
The nature of large content repositories has recently undergone significant evolution. Advances in information extraction and semantic enrichment have facilitated the automatic identification of semantic entities and relations from content. Online phenomena like collaborative tagging and social networks have harnessed the wisdom of the crowds to manually identify semantic structures. Today's knowledge repositories consequentially contain semantic structures as well as massive amounts of content. This development poses a challenge to traditional repository visualization techniques.

Information Landscapes have commonly been applied to visualize large content repositories [SKM*09]. An Information Landscape represents the topical similarity between documents through the spatial proximity of their positions on a two-dimensional map. Node-Link diagrams have commonly been applied to visualize semantic graph structures [KHL*07], representing semantic entities as nodes and semantic relations as links connecting the nodes. The integration of Information Landscapes and Node-Link-Diagrams would facilitate the comprehensive visualization of repositories containing both content and semantic structures. Because the Node-Link-Diagram and the Information Landscape can be constructed to share the same coordinate system, the former visualizations could simply be drawn atop of the latter. However, the straight lines or regular curves produced by a Node-Link Diagram break the geographic map metaphor imposed by the Information Landscape because they ignore terrain features. Furthermore, a link connecting two nodes may pass through terrain features which are not semantically related to the link or the nodes. This behavior produces contradictory visual evidence which hinders comprehension of the visualization.

2. Our Approach

We propose a path finding approach for integrating Information Landscapes and Node-Link diagrams. Our approach is based on a vector-space representation of documents and on a graph representation of the semantic structure. We assume that semantic entities and relations can be assigned vectors in the vector space model. In our experiments, we have employed named entity extraction techniques to identify semantic entities in text documents and co-occurrence analysis to identify relations between semantic entities.

A projection algorithm based on clustering and force-directed placement is used to compute two-dimensional positions for the documents and for the semantic entities. A height-map reflecting document density is used to construct the Information Landscape (compare figure 1(a). The Node-Link Diagram is initialized by placing nodes at the computed entity positions. Links between node pairs are then com-



(a) Path found in the Information Landscape (b) Height map used by path finding algorithm

Figure 1: *Combination of Information Landscape and Node-Link diagram. 1(a) The path avoids semantically unrelated topical clusters (e.g., top-left path segment). 1(b) The corresponding height-map used for the path finding algorithm (white corresponds to high altitude values). Both images share the same coordinate system.*

puted using a path finding algorithm. In our experiments, we have employed the A*-Algorithm [HNR68].

The core concept of our approach is to find the path for each link in a separate height-map which reflects the semantic properties of the documents in relation to the link (compare figure 1(b)). The altitude at a position in the height-map corresponds to the cosine similarity between the term vector associated with the link and the term vectors of documents near this position, weighted by their distances to this position. Regions with high altitudes (white in figure 1(b)) are avoided by the path finding algorithm.

In the Information Landscape the resulting path tends to avoid semantically unrelated topical clusters (i.e. hills), but will move across hills representing similar concepts, as shown in figure 1(a). Thus, the amount of contradictory visual evidence is reduced. Also, the paths do not break the landscape metaphor because they tend to respect terrain features in general.

3. Conclusion

We have presented preliminary research on a path finding approach to the integration of Information Landscapes and Node-Link diagrams. Our experiments indicate that our approach produces paths which reduce the amount of contradictory visual evidence and integrate well with the geographic map metaphor imposed by the Information Landscape. In the future we plan to evaluate the quality of the paths, for example using local stress measures, and test the usefulness of our combined approach with test users. We also will continue to refine our approach, for example by interpolating the link term vector along the path.

Acknowledgements

The Know-Center is funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - under the auspices of the Austrian Ministry of Transport, Innovation and Technology, the Austrian Ministry of Economics and Labor and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency (FFG).

References

[HNR68] HART P., NILSSON N., RAPHAEL B.: A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on 4*, 2 (july 1968), 100 –107.

[KHL*07] KATIFORI A., HALATSIS C., LEPOURAS G., VASSILAKIS C., GIANNOPOULOU E.: Ontology visualization methods - A survey. *ACM Comput. Surv. 39*, 4 (2007), 10.

[SKM*09] SABOL V., KIENREICH W., MUHR M., KLIEBER W., GRANITZER M.: Visual knowledge discovery in dynamic enterprise text repositories. In *IV '09: Proceedings of the 2009 13th International Conference Information Visualisation* (Washington, DC, USA, 2009), IEEE Computer Society, pp. 361–368.