Virtual Observatories of the Future
ASP Conference Series, Vol. 225, 2001
R.J. Brunner, S.G. Djorgovski, and A.S. Szalay, eds.

Computer Science Issues in the National Virtual Observatory

Michael T. Goodrich
Deptartment of Computer Science, Johns Hopkins University, Baltimore, MD 21218

Abstract. This paper summarizes a discussion on general technological issues related to a national virtual observatory in a special session at the 2000 Workshop on Virtual Observatories of the Future, June 15, 2000. The discussion leaders were Michael Goodrich, of Johns Hopkins University, Jim Gray, of Microsoft, Michael Kurtz, of the Harvard-Smithsonian Center for Astrophysics, and Giuseppe Longo, of the Osservatorio Astronomico di Capodimonte.

1. Five Key CS Technologies

The discussion was opened by a presentation of five key technologies that are related to the needs of a National Virtual Observatory (NVO).

1.1. External-Memory Computations

When computations and data sets are too large to fit into the internal memory of a computer, they must necessarily involve external-memory computation. Moreover, because of the huge orders of magnitudes differences in the time to access storage in internal memory versus external memory, large-scale algorithms and data structures must be designed from the start with external memory in mind from the start. Thus, algorithm engineering technology developed for external-memory computations should be ported to the NVO. Relevant references include Agarwal *et al.* (1999), Arge (1995), Arge *et al.* (1998), Arge and Vitter (1996), Chiang *et al.* 1995), Crauser *et al.* (1998), Goodrich *et al.* (1993), Nodine *et al.* (1996), and Vengroff and Vitter (1996).

1.2. Geometric Indexing

NVO data is inherantly geometric, in that much of it can be modelled as points in a multi-dimensional space. Decades of research in computer science has shown that such data is best indexed geometrically using structures such as the hierarchical triangular mesh (HTM), k-d tree, or BAR-tree. Relevant references include Bentley (1990), Ding and Weiss (1993), Duncan *et al.* (1998), Duncan *et al.* (1999), Eastman (1981), Overmars and van Leeuwen (1982), Samet (1990a, 1990b, 1990c, 1992, 1995), and Sproull (1988).

1.3. Meta-Computing

An interesting concept for performing computations over the Internet involves using a loosely-configured coalition of computers to perform that computation. Such coalitions are known as meta-computers or computational grids, and they offer the potential for the efficient exploitation of significant parallelism for performing large computations. Relevant references include Hernández *et al.* (2000), Foster *et al.* (1998), Haupt *et al.* (1999), and Patten *et al.* (1999).

1.4. Information Security

Computations that are to be performed over the Internet and data resources that are to be housed on Internet computers must unfortunately content with computer security issues. Data must be kept so as to achieve certain security properties, which including propriety rights, integrity, availability, and (in some cases) confidentiality. Thus, modern cryptographic techniques may need to be incorporated into the architecture for an NVO. Relevant references include Foster *et al.* (1998), Hughes and Hughes (1998), Knudsen (1998), and Schneier (1996).

1.5. Information Visualization

Information that extracted from an NVO must be visualized. Several techniques exist for visualizing multi-dimensional and even relational data. These should be exploited. Relevant references include Alberts *et al.* (1997), Di Battistta *et al.* (1994), Di Battistta *et al.* (1999), Dodson, (1996), and Johnson (1993).

2. Discussion

In addition to these topics, a discussion leader raised the point that the community will need to decide if the NVO should be used to *perform* science or to *facilitate* science. Many in the audience commented that it should do both.

A discussion leader asked how computational resources will be performed, namely, whether or not sophisticated computations should be performed at client computers or at server computers. An audience member commented that he felt that simple computations could be performed without fees for all users, but that he felt that NVO time for complex computations should be requested in the same manner that instrument time is requested for physical observatories. Likewise, a small panel should allocated time for complex queries on the NVO in the same manner as panels allocate time at physical observatories.

Acknowledgments. The research of this author is supported by the NSF under grants CCR-9732300 and PHY-9980044.

References

Agarwal, P.K., Arge, L., Brodal, G., & Vitter, J.S. 1999, *I/o-efficient dynamic point location in monotone subdivisions*, In *Proc. 10th ACM-SIAM Sympos. Discrete Algorithms*, pp. 11–20

Alberts, D., Gutwenger, C., Mutzel, P., & Näher, S. 1997, *AGD-Library: A library of algorithms for graph drawing*, In *Proc. Workshop on Algorithm Engineering*, pp. 112–123

Arge, L. 1995, *The Buffer Tree: A new technique for optimal I/O-algorithms*, In *Proc. 4th Workshop Algorithms Data Structures*, number 955 in Lecture Notes Computer Science, pp. 334–345

Arge, L., Procopiuc, O., Ramaswamy, S., Suel, T., & Vitter, J.S. 1998, *Theory and practice of I/O-efficient algorithms for multidimensional batched searching problems*, In *Proc. 9th ACM-SIAM Sympos. Discrete Algorithms*, pp. 685–694

Arge, L., & Vitter, J.S. 1996, *Optimal interval management in external memory*, In *Proc. 37th Annu. IEEE Sympos. Found. Comput. Sci.*, pp. 560–569

Bentley, J.L. 1990, *K-d trees for semidynamic point sets*, In *Proc. 6th Annu. ACM Sympos. Comput. Geom.*, pp. 187–197

Chiang, Y.-J., Goodrich, M.T., Grove, E.F., Tamassia, R., Vengroff, D.E. & Vitter, J.S. 1995, *External-memory graph algorithms*, In *Proc. 6th ACM-SIAM Sympos. Discrete Algorithms*, pp. 139–149

Crauser, A., Ferragina, P., Mehlhorn, K., Meyer, U., & Ramos, E. 1998, *Randomized external-memory algorithms for some geometric problems*, In *Proc. 14th Annu. ACM Sympos. Comput. Geom.*, pp. 269–268

Di Battista, G., Eades, P., Tamassia, R., & Tollis, I.G. 1994, *Algorithms for drawing graphs: an annotated bibliography*, *Comput. Geom. Theory Appl.*, 4, pp.235–282

Di Battista, G., Eades, P., Tamassia, R., & Tollis, I.G. 1999, *Graph Drawing*, Prentice Hall, Upper Saddle River, NJ

Ding, Y., & Weiss, M.A. 1993, *The k-d heap: An efficient multi-dimensional priority queue*, In *Proc. 3rd Workshop Algorithms Data Struct.*, volume 709 of *Lecture Notes Comput. Sci.*, pp. 302–313

Dodson, D. 1996, *COMAIDE: Information visualization using cooperative 3D diagram layout*, In F. J. Brandenburg, editor, *Graph Drawing (Proc. GD '95)*, volume 1027 of *Lecture Notes Comput. Sci.*, pp. 190–201

Duncan, C.A., Goodrich, M.T., & Kobourov, S.G. 1998, *Balanced aspect ratio trees and their use for drawing very large graphs* In *Graph Drawing*, Lecture Notes in Computer Science, pp. 111–124.

Duncan, C.A., Goodrich, M.T., & Kobourov, S.G. 1999, *Balanced aspect ratio trees: Combining the benefits of k-d trees and octrees*, In *10th ACM-SIAM Symp. on Discrete Algorithms*, pp. 300–309

Hernández, A.T.E, Cardinale, Y., Figueira, C. 2000, *SUMA: A scientific metacomputer*, In E. H. D'Hollander, J. R. Joubert, F. J. Peters, and H. Sips, editors, *Parallel Computing: Fundamentals & Applications, Proceedings of the International Conference ParCo'99*, pp. 566–573

Eastman, C.M. 1981, *Optimal bucket size for nearest neighbor searching in k-d trees*, *Inform. Process. Lett.*, 12, pp. 165–167

Foster, I., Kesselman, C., Tsudik, G., & Tuecke, S. 1998, *A security architecture for computational grids*, In *Proceedings of the 5th ACM Conference on Computer and Communications Security (CCS-98)*, pp. 83–92

Goodrich, M.T., Tsay, J.-J, Vengroff, D.E., & Vitter, J.S. 1993, *External-memory computational geometry*, In *Proc. 34th Annu. IEEE Sympos. Found. Comput. Sci.*, pp. 714–723

Haupt, T., Akarsu, E., & Fox, G. 1999, *WebFlow: A framework for Web based metacomputing*, *Lecture Notes in Computer Science*, 1593, p. 291

Hughes, M., & Hughes, C. 1998, *Applied Java Cryptography*, Manning Publications, Greenwich, CT

Johnson, B. 1993, *Treeviz: Treemap visualization of hierarchically structured information*, In *Proc. ACM Conf. on Human Factors in Computing Systems*, pp. 369–370

Knudsen, J.B. 1998, *Java Cryptography*, O'Reilly & Associates, Inc., Newton, MA

Nodine, M.H., Goodrich, M.T., & Vitter, J.S. 1996, *Blocking for external graph searching*, *Algorithmica*, 16(2), pp.181–214

Overmars, M.H., & van Leeuwen, J. 1982, *Dynamic multi-dimensional data structures based on quad- and k-d trees*, *Acta Inform.*, 17, pp.267–285

Patten, C.J., Hawick, K.A., & Hercus, J.F. 1999, *Towards a scalable metacomputing storage service*, *Lecture Notes in Computer Science*, 1593, p. 350

Samet, H. 1990a, *Applications of Spatial Data Structures: Computer Graphics, Image Processing, and GIS*, Addison-Wesley, Reading, MA

Samet, H. 1990b, *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, Reading, MA

Samet, H. 1990c, *Hierarchical data structures for spatial reasoning*, In L. F. Pau, editor, *Mapping and Spatial Modelling for Navigation*, pp. 41–58

Samet, H. 1992, *Hierarchical data structures for three-dimensional data*, In R. Vinken, editor, *From Geoscientific Map Series to Geo-Information Systems*, volume 122 of *Geologisches Jahrbuch A*, pp. 45–58

Samet, H. 1995, *Spatial data structures*, In W. Kim, editor, *Modern Database Systems, The Object Model, Interoperability and Beyond*, pp. 361–385

Schneier, B., 1996 *Applied Cryptography (Second Edition)*, John Wiley & Sons

Sproull, R.F 1998, *Refinements to nearest-neighbor searching in k-d trees*, Technical Report SSAPP 184, Southerland, Sproull and Associates

Vengroff, D.E., & Vitter, J.S. 1996, *Efficient 3-d range searching in external memory*, In *Proc. 28th Annu. ACM Sympos. Theory Comput.*