

University of Groningen

Query driven visualization of large scale multi-dimensional astronomical catalogs

Buddelmeijer, Hugo

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Buddelmeijer, H. (2011). *Query driven visualization of large scale multi-dimensional astronomical catalogs*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



rijksuniversiteit
groningen

Query Driven Visualization of Large Scale Multi-Dimensional Astronomical Catalogs

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. E. Sterken,
in het openbaar te verdedigen op
maandag 4 juli 2011
om 13.15 uur

door

Hugo Buddelmeijer

geboren op 4 mei 1980
te Delft

Promotores: Prof. dr. E. A. Valentijn
Prof. dr. S. C. Trager
Prof. dr. J. B. T. M. Roerdink

Beoordelingscommissie: Prof. dr. J. M. van der Hulst
Prof. dr. R. F. Peletier
Prof. dr. M. Biehl

ISBN: 978-90-367-4975-6
ISBN: 978-90-367-4974-9 (electronic version)

“In the beginning the universe was created. This has made a lot of people very angry and been widely regarded as a bad move.” – Douglas Adams

This research was financially supported by the Netherlands Organisation for Scientific Research (NWO) under project number 643.200.501, as part of the STARE program.



Cover: Messier galaxy 101

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction: Large Astronomical Data Sets | 11 |
| 1.1 | Data Explosion in Astronomy | 12 |
| 1.1.1 | Growth of Data Volume and Analysis Capacity | 12 |
| 1.1.2 | Analysis of Large Scale Multi-Dimensional Data | 12 |
| 1.1.3 | Collaboration and Sharing | 13 |
| 1.1.4 | Interactive Exploration and Visualization | 13 |
| 1.2 | Galaxies as a Multi-Dimensional Problem | 13 |
| 1.2.1 | Galaxy Relations | 14 |
| 1.2.2 | Galaxy Evolution | 15 |
| 1.2.3 | Nature or Artifact: the Influence of Algorithms | 15 |
| 1.2.4 | Effects of Sample Selection | 16 |
| 1.3 | Techniques for Handling Large Scale Data | 16 |
| 1.3.1 | Astronomical Data Products | 17 |
| 1.3.2 | Bookkeeping and Data Lineage | 19 |
| 1.3.3 | Processing | 21 |
| 1.3.4 | Analysis and Visualization | 22 |
| 1.3.5 | Interoperability | 24 |
| 1.4 | The Astro-WISE Environment | 26 |
| 1.4.1 | Astro-WISE nodes | 27 |
| 1.4.2 | Astro-WISE key features | 28 |
| 1.5 | Summary and Prospect | 30 |
| 1.5.1 | Data Pulling in the Analysis Domain | 30 |
| 1.5.2 | Query Driven Visualization with Data Lineage | 31 |
| 1.5.3 | Comparison of Density Estimation Methods for Astronomical Datasets | 32 |
| 1.5.4 | Galaxy Evolution | 32 |
| 1.5.5 | Thesis Layout | 32 |

| | | |
|----------|--|-----------|
| 2 | Pulling Astronomical Catalogs | 33 |
| 2.1 | Introduction | 34 |
| 2.1.1 | A Functional Approach to Catalogs as Process Targets | 34 |
| 2.1.2 | Astro-WISE and other information systems | 35 |
| 2.2 | Introducing Source Collections | 36 |
| 2.2.1 | Source Collection Description and Representations | 36 |
| 2.2.2 | Source Collection Example | 38 |
| 2.2.3 | Key Features in Example | 39 |
| 2.3 | Source Collection Key Features | 41 |
| 2.3.1 | Store Data Lineage instead of Tables | 41 |
| 2.3.2 | Elementary Operators | 41 |
| 2.3.3 | Data Pulling | 42 |
| 2.3.4 | Partial Processing with Dependency Reorganization | 42 |
| 2.3.5 | Generic Operator for Attribute Calculation | 43 |
| 2.3.6 | Implicit Determination of Logical Relations | 43 |
| 2.3.7 | Scalability | 44 |
| 2.3.8 | Interactive Visualization | 44 |
| 2.4 | Source Collections Details | 45 |
| 2.4.1 | Catalog Operations: Source Collection Classes | 45 |
| 2.4.2 | Source Collection Persistent Properties | 45 |
| 2.4.3 | Sources | 48 |
| 2.4.4 | Attributes | 49 |
| 2.4.5 | Source Collection Dependency Graphs | 50 |
| 2.4.6 | Pulling Catalogs | 52 |
| 2.4.7 | Optimization of Dependency Graphs | 55 |
| 2.4.8 | Processing and Storing | 58 |
| 2.5 | Operators: Source Collection Classes | 62 |
| 2.5.1 | Operator: Detect Sources | 62 |
| 2.5.2 | Operator: Measure Photometry | 63 |
| 2.5.3 | Operator: Filter Sources | 63 |
| 2.5.4 | Operator: Select Sources | 64 |
| 2.5.5 | Operator: Concatenate Sources | 64 |
| 2.5.6 | Operator: Relabel Sources | 65 |
| 2.5.7 | Operator: Select Attributes | 66 |
| 2.5.8 | Operator: Rename Attributes | 66 |
| 2.5.9 | Operator: Concatenate Attributes | 67 |
| 2.5.10 | Operator: Attribute Calculator | 68 |
| 2.5.11 | Operator: External | 70 |
| 2.5.12 | Operator: Pass | 70 |
| 2.6 | Comparing the Sources in Source Collections | 72 |
| 2.6.1 | Algorithm Overview | 72 |
| 2.6.2 | Representations of Relations between Sets | 74 |
| 2.6.3 | Relations | 77 |
| 2.6.4 | Relation Manipulation Algorithm | 79 |
| 2.7 | Discussion and Extensions | 86 |

| | | |
|----------|---|-----------|
| 2.7.1 | Comparison with MIDAS Table File System | 86 |
| 2.7.2 | Image and Association Operations | 87 |
| 2.7.3 | Improved Attribute Calculators | 89 |
| 2.7.4 | Data Lineage for Further Analysis | 89 |
| 2.7.5 | Increased Scope | 90 |
| 2.7.6 | Improved Algorithms | 91 |
| 2.8 | Summary | 92 |
| 3 | Catalog Data in Astro-WISE | 95 |
| 3.1 | Introduction | 96 |
| 3.1.1 | Processing in Astro-WISE | 96 |
| 3.1.2 | Existing Classes Overview | 97 |
| 3.1.3 | Source Collection Rationale | 97 |
| 3.1.4 | Source Collection Overview | 99 |
| 3.2 | Existing Classes | 100 |
| 3.2.1 | DBObject | 101 |
| 3.2.2 | DataObject | 101 |
| 3.2.3 | ProcessTarget | 101 |
| 3.2.4 | SourceList | 102 |
| 3.2.5 | AssociateList | 102 |
| 3.2.6 | CombinedList | 103 |
| 3.3 | Source Collection Class | 105 |
| 3.3.1 | Source Collection Definition | 105 |
| 3.3.2 | Usage of Source Collections | 107 |
| 3.3.3 | SQL Layer | 110 |
| 3.3.4 | Python Layer | 113 |
| 3.3.5 | Example Dependency Graph and SQL | 115 |
| 3.4 | Source Collection Classes | 117 |
| 3.4.1 | Operator: SourceList Wrapper | 118 |
| 3.4.2 | Operator: Select Attributes | 118 |
| 3.4.3 | Operator: Rename Attributes | 118 |
| 3.4.4 | Operator: Concatenate Attributes | 119 |
| 3.4.5 | Operator: Filter Sources | 119 |
| 3.4.6 | Operator: Select Sources | 119 |
| 3.4.7 | Operator: Concatenate Sources | 120 |
| 3.4.8 | Operator: Relabel Sources | 120 |
| 3.4.9 | Operator: Attribute Calculator | 120 |
| 3.4.10 | Operator: External | 121 |
| 3.4.11 | Operator: Pass | 121 |
| 3.5 | Source Collection Tree | 122 |
| 3.5.1 | Initializing a Source Collection Tree | 122 |
| 3.5.2 | Dependency Graph Modifications | 122 |
| 3.5.3 | Processing Source Collections | 124 |
| 3.5.4 | Creating New Source Collections by Pulling Data | 125 |
| 3.5.5 | Interaction and Visualization | 127 |

| | | |
|----------|--|------------|
| 3.6 | Attribute Calculators | 128 |
| 3.6.1 | Attribute Calculator Definition Class | 128 |
| 3.6.2 | Attribute Calculator Parameter Class | 129 |
| 3.6.3 | Code | 129 |
| 3.6.4 | Instantiating New Attribute Calculators | 130 |
| 3.7 | Set Relations | 130 |
| 3.7.1 | Set Relations Class | 131 |
| 3.7.2 | Set Relations Set Class | 132 |
| 3.7.3 | Source Collection Specifics | 132 |
| 3.8 | Extensions | 133 |
| 3.8.1 | Improved Operators | 133 |
| 3.8.2 | Integration with other Astro-WISE Functionality | 135 |
| 3.8.3 | Improved Storage | 135 |
| 3.8.4 | Scalability | 136 |
| 3.9 | 15 square degree SDSS pilot | 138 |
| 3.9.1 | Data Ingestion | 138 |
| 3.9.2 | Derived Attributes | 140 |
| 3.9.3 | Sample Selection | 141 |
| 3.9.4 | Galaxy Densities | 142 |
| 3.9.5 | Combining All Attributes | 142 |
| 3.9.6 | Processing | 142 |
| 4 | Comparison of Density Estimation Methods | 145 |
| 4.1 | Introduction | 146 |
| 4.2 | Density estimation methods | 147 |
| 4.2.1 | k-nearest neighbor method | 148 |
| 4.2.2 | Adaptive Epanechnikov kernel density estimation | 148 |
| 4.2.3 | Adaptive Gaussian kernel density estimation (DEDICA) | 151 |
| 4.2.4 | Delaunay Tessellation Field Estimator (DTFE) | 153 |
| 4.2.5 | Error measures | 153 |
| 4.3 | Datasets | 154 |
| 4.3.1 | Simulated datasets with known density fields | 154 |
| 4.3.2 | Astronomical datasets with unknown density fields | 156 |
| 4.4 | Results | 160 |
| 4.4.1 | Simulated Datasets | 161 |
| 4.4.2 | Application to SDSS datasets | 164 |
| 4.5 | Conclusions and Recommendations | 170 |
| 4.5.1 | Artificial and Simulated Datasets | 170 |
| 4.5.2 | SDSS Datasets | 171 |
| 4.5.3 | DEDICA | 171 |
| 4.5.4 | Computational Complexities | 172 |
| 4.5.5 | Recommendations | 173 |
| 4.A | Mock samples – selection effects | 174 |
| 4.B | DTFE Monte-Carlo sampling | 178 |

| | | |
|----------|--|------------|
| 5 | Query Driven Visualization by Abstraction of Data Pulling | 181 |
| 5.1 | Introduction | 182 |
| 5.1.1 | Data Pulling and Declarative Querying | 182 |
| 5.1.2 | Full Data Lineage and Exploration | 183 |
| 5.1.3 | Abstraction and Interoperation through SAMP | 183 |
| 5.1.4 | Astro-WISE | 183 |
| 5.2 | Interoperability through SAMP | 183 |
| 5.2.1 | Simple Application Messaging Protocol | 184 |
| 5.2.2 | Existing Catalog Related Messages | 184 |
| 5.2.3 | Data Pulling Messages | 184 |
| 5.2.4 | Object Messages | 185 |
| 5.3 | SAMP HUB and Clients | 186 |
| 5.3.1 | Existing SAMP applications | 186 |
| 5.3.2 | Astro-WISE and SAMP | 187 |
| 5.3.3 | Query Driven Visualization Prototype | 187 |
| 5.A | Appendix: SAMP Protocol and Messages | 188 |
| 5.A.1 | SAMP Protocol | 188 |
| 5.A.2 | SAMP Data Types | 188 |
| 5.A.3 | SAMP Messages | 192 |
| 5.A.4 | Query Driven Visualization | 193 |
| 5.B | Appendix: SAMP in the awe-prompt | 196 |
| 5.B.1 | Data in Astro-WISE | 196 |
| 5.B.2 | SAMP Classes | 196 |
| 5.B.3 | Communication with HUB | 197 |
| 5.B.4 | Sending Data | 197 |
| 5.B.5 | Sending Interaction Messages | 198 |
| 5.B.6 | Sending General Messages | 199 |
| 5.B.7 | Receiving Table Related Messages | 199 |
| 5.B.8 | Query Driven Visualization Data Structures | 200 |
| 5.B.9 | Receiving Query Driven Visualization Messages | 200 |
| 5.B.10 | Storing Information | 201 |
| 6 | Galaxies and their Environment | 203 |
| 6.1 | Introduction | 204 |
| 6.1.1 | Quantifying Environment | 204 |
| 6.1.2 | Query Driven Visualization and Data Pulling | 204 |
| 6.1.3 | Applications and Interoperation | 205 |
| 6.2 | Data Sets | 206 |
| 6.2.1 | Spectroscopic Completeness from KIAS VAGC | 206 |
| 6.2.2 | Morphologies from Galaxy Zoo | 206 |
| 6.2.3 | Initial Sample | 206 |
| 6.2.4 | Initial Datasets Creation | 207 |
| 6.3 | Density Defining Populations | 208 |
| 6.3.1 | Pulling Density Defining Populations | 209 |
| 6.3.2 | Created Source Collections | 212 |

| | | |
|----------|--|------------|
| 6.4 | Quantification of Environment | 213 |
| 6.4.1 | Pulling Densities | 216 |
| 6.5 | Analysis Samples | 217 |
| 6.5.1 | Pulling Attributes | 217 |
| 6.5.2 | Creating Figures | 217 |
| 6.6 | Results | 220 |
| 6.7 | Kernel Density Estimation | 223 |
| 6.8 | Comparison with KIAS Morphology | 227 |
| 6.9 | Conclusions | 230 |
| 6.9.1 | Galaxy Evolution | 230 |
| 6.9.2 | Evaluation of Techniques | 231 |
| 7 | Summary and Outlook | 233 |
| 7.1 | Query Driven Visualization | 234 |
| 7.1.1 | Automatic Data Discovery and Creation | 234 |
| 7.1.2 | Data Lineage | 236 |
| 7.1.3 | Flexibility | 237 |
| 7.1.4 | Astro-WISE | 237 |
| 7.2 | Galaxies and their Environment | 237 |
| 7.3 | Outlook | 238 |
| 7.3.1 | Data Lineage Standards | 238 |
| 7.3.2 | Ultra-fine Partitioning of Data Lineage | 239 |
| 7.3.3 | Incremental Visualization | 239 |
| 7.3.4 | Current and Future Astronomical Projects | 240 |
| 7.3.5 | Galaxy Evolution | 241 |
| 7.4 | Conclusions | 242 |
| | Nederlandse Samenvatting | 243 |
| | Dankwoord | 251 |