

Visual Exploration of Large Transportation Asset Data using Ontology-Based Heat Tree

Tuyen Le¹, Chau Le², H. David Jeong^{3*} and Chuck Jahren⁴

¹*Ph.D., Dept. of Civil, Construction and Environmental Engineering,
Iowa State University*

²*Ph.D. Student, Dept. of Civil, Construction and Environmental Engineering,
Iowa State University*

³*Associate Professor, Dept. of Civil, Construction and Environmental
Engineering, Iowa State University*

⁴*Professor, Dept. of Civil, Construction and Environmental Engineering,
Iowa State University*

tle@iastate.edu, chle@iastate.edu, djeong@iastate.edu, cjahren@iastate.edu

Abstract

The national priority to enhance the service life of aging transportation assets has triggered the collection of a huge amount of asset data across the United States. Data visualization is widely recognized as a pressing need to assist highway professionals in quickly capturing overall trends and understanding the meaning of big data sets with millions of records and hundreds of data attributes. A key requirement for big data visualization is the selection of what data to be visualized among a large number of attributes. In addition, since most of the asset data are categorical attributes that are associated with complex hierarchical classifications, the capacity of comparing statistics throughout various classification levels on the same graph is needed. Despite the wide availability of methods, data visualization is still a major challenge to the practitioners of highway agencies due to the lack of a domain-specific automated tool that can successfully address those visualization requirements. This study proposes an ontology based visualization technique for automatically visualizing hierarchical transportation data. A domain ontology of highway asset data is developed to support the selection of related data attributes to be visualized. The proposed method also includes a novel technique for creating a heat tree that indicates the distribution of asset information over another hierarchy of categories. The system is expected to effectively assist professionals in developing the visual representation of data, which will in turn, translate into better and faster visualization, time savings, and enhanced decision making.

Keywords: *Data visualization, transportation asset data, highways, heat trees, hierarchical data, ontology, semantic models*

1. Introduction

Data-driven asset management has become a critical practice amid the increasing focus on improving the service life of transportation assets. This is one of the main priorities for many highway research institutes and state departments of transportation (DOTs) (MTC 2017). Numerous transportation asset management (TAM) systems such as pavement management system (PMS), bridge management system (BMS) have been widely used by Federal, State DOTs and Local agencies. Those systems include various decision-making

Received (July 28, 2017), Review Result (September 21, 2017), Accepted (October 12, 2017)

* Corresponding Author

frameworks that utilize asset inventory and condition data for predicting the performance, determining maintenance activities and resource allocation plans, and analyzing design methods.

In attempts to support TAM, a huge amount of data has been captured and made publicly available. Various national programs have been initiated that aim to collect data on a yearly basis. Examples of those efforts include the Highway Performance Monitoring Systems (HPMS), the Long-term Pavement Performance (LTPP) program, and the National Bridge Inventory (NBI). These programs require State DOTs to provide yearly submissions of public transportation asset data to the federal agencies. As a result, a huge amount of data regarding asset inventory, condition, traffic, and materials have been obtained for more than two decades. For example, a one-year HPMS dataset contains as much as millions of roadway sections across the nation with dozens of data attributes. The availability of up-to-date data theoretically enables immediate responses to the deterioration of assets.

Despite the availability of asset data, they cannot be translated into valuable information without effective data analytics methods and tools. Statistical analysis has an undeniable role for data mining. However, analyzing big data requires considerable skills/expertise and domain knowledge to effectively understand the meaning of data (SAS 2012, Khan and Khan 2011). It is evident that data visualization can effectively enable enhanced decision-making, better data analysis, and improved collaboration (SAS 2012). Exploratory data analytics which uses a visual method to help data analysts explore the trend and overall summary of the data is commonly employed. Visualized data allows professionals to quickly see patterns and evaluate the correlation between data attributes. However, deriving meaningful information from large data sets using traditional visualization techniques is challenging due to the increasing degree of the complexity and the volume of data (Ali *et al.*, 2016).

Transportation asset data are typically complex with hundreds of fields. Most of the data are categorical attributes that relate to a particular hierarchical classification tree. Those attributes are classified in a hierarchical manner for different levels of interests and audiences. However, the current methods employed by highway agencies mainly use tabular formats, bar/column charts, colored maps, stacked area graphs, and pictograms (Higgins *et al.*, 2017). Presenting data in such formats is difficult to convey the hierarchical relationship between different levels. Especially, using those conventional methods, practitioners are required to have considerable visualization expertise and they must have a deep understanding of the inter-relation among data to properly select attributes to be visualized. Choosing the right number of data attributes is a critical task as an appropriate reduction of dimensions would help to highlight the key pattern and can reduce processing time and the density of the image without losing interesting patterns (Ali 2016). Since identifying the semantic relationship between data objects is the first step of dimension selection (Khan and Khan 2011), automated determination of related variables is required to help highway practitioners overcome the obstacle when working on large and complex data sets.

Several data platforms have been introduced to support data visualization for the transportation sector, for example, the pavement visualization module of the InfoPave system (InfoPave 2017) and the interactive GIS-based asset visualization platform (Darter *et al.*, 2008). However, there is no tool for visualizing hierarchical transportation asset information. Additionally, the current methods focus on improving the interactivity between the user and the visualized data. Few studies addressed the very first step of visualization process. The existing methods fail to enable automated selection of attributes to be included in the graphical plot. There is a need for a better tool that can assist professionals in choosing relevant data and enable visualization of hierarchical asset data.

This study develops a novel automated method for visualizing highway asset data. The core components of the proposed method include the development of a domain ontology and a new algorithm for generating asset heat trees. The ontology, which presents the

domain knowledge of highway assets in a formal manner, is to help automatically determine associated data to be visualized. The study also includes the development of various taxonomies of categorical asset data. Asset data are visualized using a heat tree technique that can show both quantities and performance of assets against a given classification hierarchy. The study provides users including non-technical ones with an automated method that can enable simple visualization and a better understanding of the asset performance.

2. Background

2.1. Current Practices of Data Visualization among Highway Agencies

The existing tools and methods of visualization for transportation agencies have been well documented in the literature. The most recent and noticeable review is the one performed by Higgins *et al.*, (2017). This study surveyed the current practices and developed a detailed guidance on data visualization. Below are key findings regarding tools, methods, audiences, and challenges derived from the study.

- ArcGIS, Microsoft Office, Adobe InDesign, and JavaScript are among the most commonly used visualization tools.
- The most popular methods of visualization are bar/column charts, colored maps, stacked area graphs, and pictograms.
- Visualized products are primarily for non-technical audiences, for instance lay people, executives, and legislators, municipal planners, officials and technical advisory.
- Data visualization is mainly responsible by in-house skilled IT personnel, but consultants are also considered to perform some of the visualization tasks.
- In-house staff has strong data management skills but there is a lack of institutional resources, software/data experience, and high-quality data.

Table 1 summarizes various advanced chart types for visualizing transportation data identified by Higgins *et al.*, (2017). These methods can be used for visualizing the relationship among data attributes, and help to see the trend and overall summary of the data. However, none of them is able to illustrate the distribution of statistics over a hierarchy of categories. A network diagram is one that has a high potential for presenting the hierarchical classification. However, due to the lack of automated tools, developing such a network diagram for hierarchical transportation asset data requires much training and experience. Thus, tools that support automated generation of the visual presentation of asset data are needed.

Table 1. Common Visualization Methods [Adapted from Higgins et al. (2017)]

Plot type	Usage	General tools
 Geographic map	Geographic maps are used to show the correlation between non-geographic data and geographic areas and locations.	Esri ArcGIS, QGIS, Google Fusion Table
 Flow chart	This type is to show the flow of a quantity between places, stages, etc.	Microsoft PowerPoint, Visio, ArcGIS, Tableau, Microsoft Power BI, SankeyMatic

 <p>2-D scatter plot</p>	<p>2-D scatter plots are commonly used for showing the relationship between a given pair of two variables.</p>	<p>Microsoft Excel, Google Sheets, Tableau, Qlik, Microsoft Power BI</p>
 <p>Parallel coordinate</p>	<p>Distinct variables are represented as parallel vertical axes. Lines connecting different axes shows the relationships between data attributes.</p>	<p>R</p>
 <p>Heat map</p>	<p>Colors are used to represent the value of the cells in a two-dimensional matrix.</p>	<p>Microsoft Excel, Google Sheets, Tableau, Qlik, Microsoft Power BI</p>
 <p>Network</p>	<p>In this plot type, data are presented as a network of nodes and links. The size and color of nodes and links can represent the value of a data attribute.</p>	<p>Microsoft Excel (with NodeXL add-on), Tableau, Qlik (with D3.js extension), Google Fusion Tables, Microsoft Power BI</p>

2.2. Related Studies on Transportation Asset Visualization and Knowledge Gap

In attempts to assist professionals in visualizing transportation data, several techniques with implementation tools have been developed for diverse types of data. These tools provide intuitive and user-friendly interfaces that allow the end user to effectively interact with large and complex data. However, most of the previous studies focused on traffic data for enhancing the effectiveness and efficiency of the visualization methods (Zeng *et al.*, 2013, Ferreira *et al.*, 2013, Chen *et al.*, 2015, Chaolong *et al.*, 2016). Very few studies have been carried out for asset data. Below are the discussions on the most noticeable work for transportation assets.

Darter *et al.*, (2008) introduced a semantic model of mobile and stationary information of assets for visualizing transportation asset management using Geographic Information System (GIS). A software prototype was also developed that uses a Resource Description Framework (RDF) as a neutral data standard. The platform supports visualizing data for various mobile and stationary transportation assets including culverts, end treatments, and real-time traffic sensors. Another major effort was initiated by the Federal Highway Administration (FHWA) that aims to develop a web-based inventory of pavement data namely InfoPave (InforPave 2017). This database is designed specially to ensure easy access to the LTPP data. InfoPave provides various intuitive ways to interact with the data. However, this is mainly for selecting and extracting a subset of data. It does not allow for the analysis of the pavement condition and the relationship among data attributes.

Traffic and assets, however, share a certain number of common characteristics. For example, a traffic incident and a transportation asset both have geographic and non-geographic attributes. Systems particularly designed for traffic incidents can be directly implemented for asset data. With this regard, a web-based tool called Incident Cluster Analysis (ICA) (Pack *et al.*, 2009; Wongsuphasawat *et al.*, 2009) is one of the best examples. ICA integrates various types of plots including ranking, histograms, 2-D plots, and parallel coordinates into a map. This integration offers the end user a convenient way to explore all relevant information for a particular event on the map. This is a mutual interaction in which a selection of objects in the plot yields a highlight of corresponding assets on the map; and vice versa, detailed incident information is shown when specific instances on the map are selected.

In summary, visualization tools currently employed by highway agencies are still lagging behind the data volume. The traditional methods do not support automated generation of visualization plots. Especially they fail to allow for the comparison of a statistic given

different hierarchical categories. For this reason, the construction of multiple plots is needed. New techniques are needed to support multi-level visualization using a unique plot to better highlight the distribution of statistics throughout a classification tree.

3. Heat Tree Visualization System

3.1. Methodology Overview

The goal of this study is to develop a system that can support automated generation of visual graphs of complex and large transportation asset data sets. One key requirement of the system is the capability of comparing data between the categories of a taxonomic tree. For instance, visualizing the roadway condition for different types of the surface taxonomy is a good example. Another primary criterion is the level of automation in identifying relevant data attributes to be included in the plot.

In order to fulfill the above requirements, the proposed system is designed to include the following three components: (1) a domain ontology of transportation asset data, (2) a set of taxonomic classification hierarchies for categorical attributes, and (3) a heat tree method for visualizing an attribute in a hierarchy of categories. The relationships among these three components are depicted in Figure 1. Specifically, the ontology is a formal representation of the knowledge of the roadway domain, it provides guidance on the association among data groups and allows for proper selection of what data attributes to be included in the graph of heat trees. The second component is a resource of classifications for those categorical attributes constituting the ontology. These hierarchies are key input information for the generation of heat trees. The final component, a heat tree technique, is to show the distribution of information of interest over another hierarchical data.

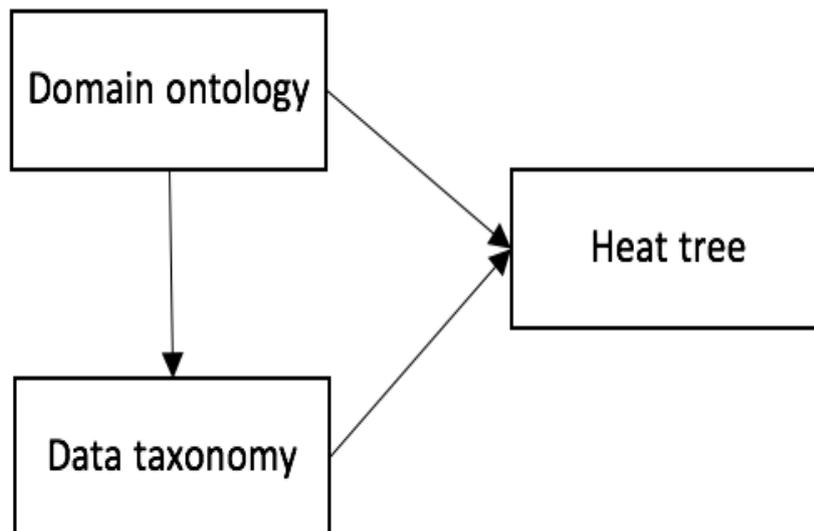


Figure 1. System Architecture

3.2. Ontology of Pavement Asset Data

Deciding what data to be included in a visualized figure is typically the first step of the data visualization process. It requires the end user to understand the meaning and the semantic relation among data attributes. This is troublesome when the data contains hundreds of attributes. In order to automate that process, a semantic model that presents the domain knowledge in a machine-readable manner with an explicit representation of relationships is needed. This helps to discard out non-related dimensions and consequently

reduce running time. To fulfill that need, this study employs the ontology approach to support the automated determination of what variables to be included in the visualized plot.

The generation of heat trees, as presented later in this paper, is based on the logic specified in an ontology of pavement asset information. The ontology, as shown in Figure 2, is a network of nodes and links. Nodes refer to data classes, and links represent semantic relations. The ontology developed in this study is built on the HPMS data model. The ontology includes six main concepts including *design feature*, *weather*, *condition*, *traffic*, *maintenance*, and *administration*. These concepts are defined by a number of attributes (see Table 2). There are two types of relationships among the data categories including *effects* and *has*. These relations indicate how different data groups relate to each other. The plot of heat tree for a certain data variable should include only those entities that have direct connection edges with it. For instance, the graph of *design feature* would include only the *administration*, *climate*, or *traffic* information. Other unrelated pairs such as ‘*maintenance - design feature*’ are dismissed as they have no direct relationship.

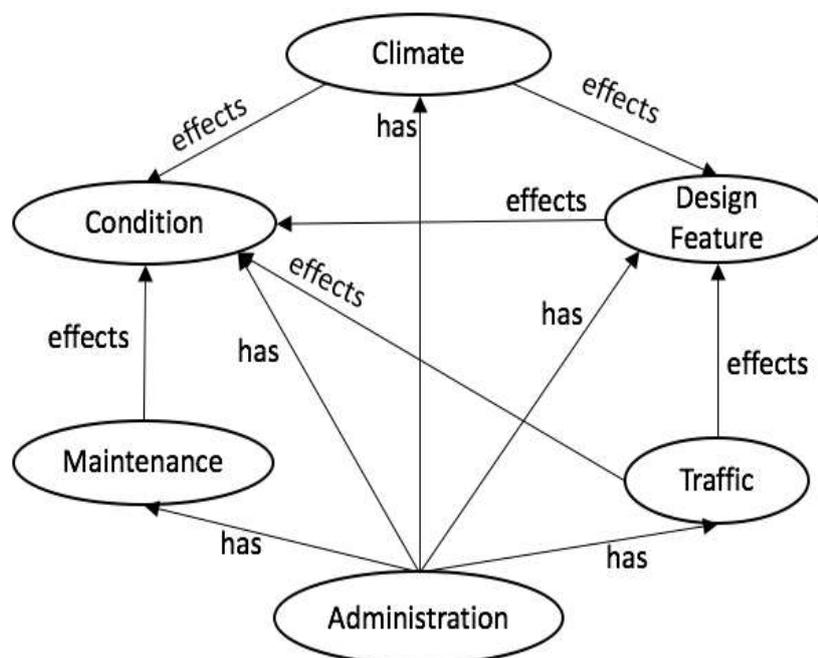


Figure 2. Highway Asset Ontology

Table 2. Attributes of the Data Classes

Admin.	Design Feature	Condition	Traffic	Climate	Maintenance
Functional System	Facility type	IRI	ADT	Climate Zone	Year of Last Improvement
Ownership	Structure Type	PSR	ADT Single Unit		Year of Last Construction
State	Surface Type	Rutting	AADT Combination		
	Base Type	Faulting	Future ADT		
	Rigid Layer	Cracking	Percent Peak Combination Trucks		
	Other 13 properties				

3.3. Hierarchies of Categorical Attributes

A data hierarchy is a detailed extension of the ontology. Each abstract concept of the ontology refers to a certain list of data attributes. As shown in Table 2, the majority of the attributes of the *administration*, *design feature*, and *climate* classes are categorical. The attributes of *condition* and *traffic* mainly take numeric values. For categorical data, their value ranges are commonly complex and are structured as a classification taxonomy. Table 3 illustrates the classification hierarchy of the possible values of the *surface type* attribute. The surface type can be classified into a three-level structure, where the components in the lower levels are a sub-type of the corresponding type in the upper level. In this study, we developed 14 different hierarchies for those categorical attributes of the concepts in the ontology.

Table 3. An Excerpt of Surface Type Taxonomy

1 st level of surface type	2 nd level surface type	3 rd level surface type
Unpaved		
Flexible	Bituminous	
Rigid	Concrete Pavement	JPCP _ Jointed Plain
Rigid	Concrete Pavement	JRCP _ Jointed Reinforced
Rigid	Concrete Pavement	CRCP _ Continuously Reinforced
Flexible	Flexible Overlay	AC over AC
Composite	AC over JCP	
Composite	Bituminous over CRCP	
Rigid	Rigid Overlay	Unbonded Jointed Concrete on PCC
Rigid	Rigid Overlay	Bonded PCC on PCC
Other		

3.4. Ontology-based Heat Tree Construction

This study adopts the concept of heat tree introduced by Foster *et al.*, (2017) to visualize taxonomic roadway asset data. A heat tree is a taxonomic tree in which the color and size of elements (nodes and links) describe the value of the interesting variables. The MetacodeR R package (Foster *et al.*, 2017) is the one that can support heat tree visualization. Although it is developed by a researcher in the biological domain, it is generic and can be applied for any type of hierarchical data. However, this package is limited for visualization and does not allow for the integration of domain knowledge that is needed for automated generation of heat tree. It requires users to specify data attributes to be plotted. As an attempt to eliminate that barrier, this study develops an algorithm in R that utilizes a domain ontology to enable computers to properly identify related data attributes and automatically generate heat trees.

An asset heat tree is composed of nodes and links. Nodes represent a specific group of asset data. The size of nodes and links demonstrates the total quantity of mile length. The color scheme indicates the distribution of an attribute throughout a classification tree. Given an input data attribute from the end user, which is called *target* information, the process of generating a heat tree is described in the steps below.

Step 1 - Identify related data attributes. This step looks for semantically relevant concepts for the target information in the ontology. For instance, given the target information of *surface type*, which is one attribute of the *design feature* concept, there are only three groups of information categories found to have direct effects including: *traffic*, *climate*, and *administration*. This step helps the system automatically identify related

categorical data attributes to ensure the inclusion of right information on the heat tree. Those identified attributes are called *base* attributes. Each pair of the target and a base attribute corresponds to a heat tree.

Step 2 – Construct taxonomic trees of the base data. As discussed earlier, categorical data of transportation assets are commonly complex and represented as classification taxonomies. In this step, the taxonomic resource is utilized to generate tree networks for those base attributes identified in step 1. A taxonomic tree is composed of nodes and links. Nodes refer to different categories, and links represent the hierarchical relationships among categories.

Step 3 – Develop heat trees. This step is to integrate the target information into the taxonomic trees which will be converted into heat trees. A heat tree shows the visualization of the target information (*e.g.*, rutting degree) throughout a classification tree (*e.g.*, surface type). The elements of the tree are colored and sized in accordance with the data value in the data set. The size of nodes and edges indicates the total length of all road sections for each category node of the taxonomic tree. The color describes the weighted average of the target attribute per categories, where the weights are relative to roadway length.

4. Implementation

To support the implementation of the method, this study develops an R script that can automatically generate heat trees for roadway asset data. This program requires users to specify the source data set, and the target information to be visualized. The code is designed to automatically identify the related attributes of the target information and generate the corresponding heat trees.

An experiment was conducted on a 548MB HPMS data set which consists of 4,334,681 records and 31 attributes. using a personal MacBook with 1.4 GHz Intel Core i5 and a 4GB memory. The system was tested using the target information of roadway condition. The total running time for reading the data set is 210.104 seconds, and the time for data processing and generating heat trees is 94.093 seconds.

One of the heat trees generated from the case study for the roadway sections in Delaware is presented in Figure 3. This heat tree shows the distribution of the weighted mean of IRI over various categories of the surface type hierarchy. The size of nodes and edges represent the total mile length and the color presents the IRI for various roadway types in the taxonomic classification structure. As revealed by the size of tree nodes, the majority of the roadways in Delaware reported in 2015 are composite and flexible. The rigid pavement accounts for only a very small portion. With respect to roadway condition, flexible pavement tends to clearly fall behind the composite pavement and rigid pavement in IRI as the color varies from dark purple to light orange. This significant difference indicates that surface type is a key indicator and has a strong impact on the performance of the pavement. Deeper analysis may be needed to discover other influential factors hidden behind this pattern.

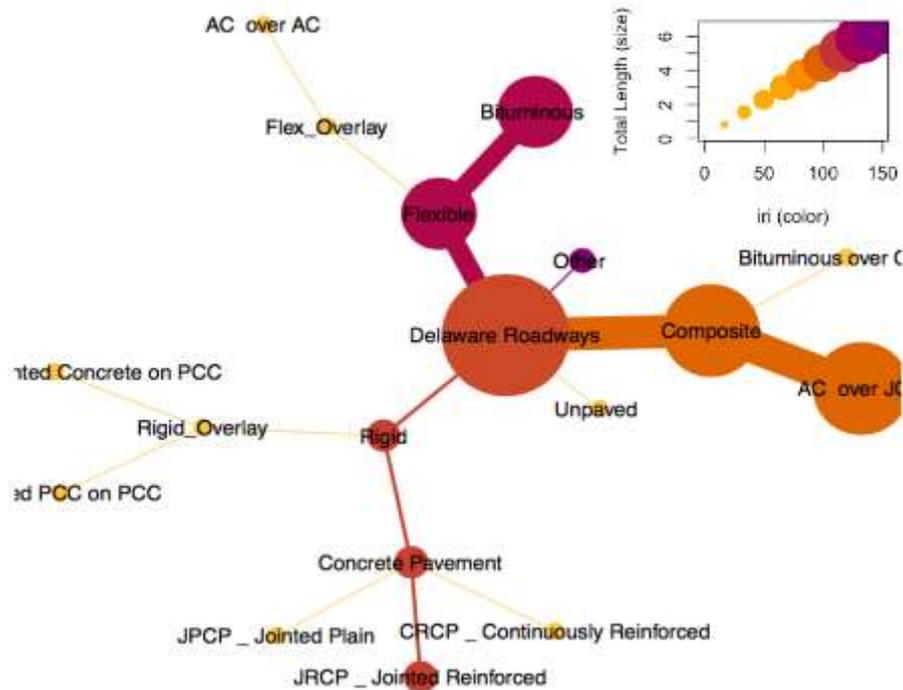


Figure 3. Roadway Condition in IRI Heat Tree

5. Implications and limitations

A primary contribution of this study is a novel method for generating heat trees that can support visual exploratory of hierarchical transportation asset data. The visualization technique helps to derive the summary of the data and allows data users to quickly see overall trends. This initial observation is the key to further statistical analyses especially in choosing an appropriate testing hypothesis. Another key advantage is that the method can automatically identify related data to be included in the visualized graph. The removal of unnecessary data attributes would enable the effective reduction of processing time of big data sets. Using the visualization system developed in this study, less effort is needed for visualizing large-scale transportation asset data when the system is able to automatically identify what data attributes are required for the multidimensional visualization plot. In addition, this study offers an inexpensive way to analyze data and eliminates the reliance on costly commercial software packages. The designed R script is flexible, extendable, and can be customized for other areas of interest.

The present study has some limitations. First, the current method is compiled in an open source R script which requires users to read and understand the code. Further research is needed to develop an open-source R package with various functions so that the end user can directly apply without a need for reading the source code. Second, the current system does not support automated generation of temporal heat trees which are useful to keep track of the status of the highway asset over time. Finally, future studies are needed to expand the domain ontology to other transportation assets such as bridges, tunnels, and other minor assets such as culverts, guardrails, and signs. Once the relations among those domains of knowledge are formulated, the systems can serve as an integrated transportation visualization system for interconnected civil infrastructure asset management.

6. Conclusions

Despite the rapid growth of transportation data, a majority of the data is not effectively leveraged for decision making due to the lack of efficient data analytics tools. This study offers a new method for visualizing transportation asset data. The method employs the

ontology approach and the heat tree method to support automated visualization of taxonomic asset data. The study includes the development of a domain ontology and many data taxonomies for highway assets. The system also includes the development of an algorithm for generating a heat tree. A heat tree can visually compare asset condition indicators or other asset statistics between different categories. The advantage of this tool compared with the existing data visualization ones is the integration of domain knowledge into the visualization algorithm. This helps the system to automatically identify related data attributes for the comparison.

The system is developed in R program. An open source R script is also developed to support the automated generation of heat trees. Although the program initially targets the highway transportation asset data, it is designed to be extendable to other areas of interest such as safety analysis.

The study is expected to offer a simple and straightforward tool for the highway professionals of Federal, State DOTs and Local agencies to effectively explore their transportation asset data. It is expected to help them quickly capture the main trend of asset performance and to improve the effectiveness of maintenance activities.

References

- [1] S. M. Ali, N. Gupta, G. Krishna Nayak and R. Kumar Lenka, "Big data visualization: Tools and challenges", In Contemporary Computing and Informatics (IC3I), IEEE, (2016), pp. 656-660.
- [2] J. Chaolong, W. Hanning and W. Lili, "Research on Visualization of Multi-Dimensional Real-Time Traffic Data Stream Based on Cloud Computing", Procedia Engineering, vol. 137, (2016), pp. 709-718.
- [3] W. Chen, F. Guo and F.-Y. Wang, "A survey of traffic data visualization", IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 6, (2015), pp. 2970-2984.
- [4] M. Darter, T. Lasky and B. Ravani, "Transportation asset management and visualization using semantic models and Google Earth", Transportation Research Record: Journal of the Transportation Research Board 2024, (2008), 27-34.
- [5] N. Ferreira, J. Poco, H. T. Vo, J. Freire and C. T. Silva, "Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips", IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 12, (2013), pp. 2149-2158.
- [6] Z. SL Foster, T. J. Sharpton and N. J. Grünwald, "Metacoder: An R package for visualization and manipulation of community taxonomic diversity data", PLoS computational biology, vol. 13, no. 2, (2017).
- [7] M. Khan and S. Shah Khan, "Data and information visualization methods, and interactive mechanisms: A survey", International Journal of Computer Applications, vol. 34, no. 1, (2011), pp. 1-14.
- [8] InfoPave. "Introduction to InfoPave LTPP", Accessed June 13, 2017 at https://infopave.fhwa.dot.gov/Page/Index/INTRO_TO_LTPP, (2017).
- [9] J. Kim, N. Gucunski, T. H. Duong and K. Dinh, "Three-Dimensional Visualization and Presentation of Bridge Deck Condition Based on Multiple NDE Data", Journal of Infrastructure Systems, (2016): B4016012.
- [10] Midwest Transportation Center (MTC). About the Midwest Transportation Center, (2017).
- [11] N. Higgins, R. Basile, S. Van Hecke, J. Zissman and S. Gilkeson, "Data Visualization Methods for Transportation Agencies", Washington, DC: The National Academies Press, (2017). doi:<https://doi.org/10.17226/24755>, (2017).
- [12] M. L. Pack, K. Wongsuphasawat, M. VanDaniker and D. Filippova, "ICE--visual analytics for transportation incident datasets", In Information Reuse & Integration, IEEE, (2009), pp. 200-205.
- [13] SAS, Visualization, Data, "Making Big Data Approachable and Valuable", Whitepaper, Source: IDG Research Services, (2012).
- [14] X. Wang, W. Dou, S.-E. Chen, W. Ribarsky and R. Chang, "An interactive visual analytics system for bridge management", In Computer Graphics Forum, Blackwell Publishing Ltd., vol. 29, no. 3, (2010), pp. 1033-1042.
- [15] K. Wongsuphasawat, M. Pack, D. Filippova, M. VanDaniker and A. Olea, "Visual analytics for transportation incident data sets", Transportation Research Record: Journal of the Transportation Research Board 2138, (2009), pp. 135-145.
- [16] W. Zeng, F. Chi-Wing, S. Müller Arisona and H. Qu, "Visualizing interchange patterns in massive movement data", In Computer Graphics Forum, Blackwell Publishing Ltd, vol. 32, no. 3pt3, (2013), pp. 271-280.

Authors



Dr. Tuyen Le is a Postdoctoral Research Associate in the Department of Civil, Construction, and Environmental Engineering at Iowa State University. He obtained his doctoral degree in Construction Engineering and Management from Iowa State University. His research is primarily focused on developing new automated methods and computational infrastructure for reducing manual and duplicate efforts in data collection, processing, and analytics for the construction industry. Particularly, his research interest includes Building & Civil Information Modeling (BIM/CIM), digital data exchange, natural language processing, intelligent systems, and human-data interaction.



Chau Le is a PhD Student in the Department of Civil, Construction, and Environmental Engineering at Iowa State University. His research interest includes digital data analytics, machine learning, business analytics & intelligence, and Building & Civil Modelling (BIM/CIM).



Dr. H. David Jeong is an associate professor in the department of Civil, Construction, and Environmental Engineering at Iowa State University (ISU). He is the Director of the Data Analytics Laboratory for Project and Infrastructure Management (DALPIM) at ISU. He has expertise in data analytics and data mining for effective decision making for project management and infrastructure asset management, big data applications to the construction industry, and data-information-decision relationships. Dr. Jeong has published more than 80 technical journal and conference papers for the past ten years.



Dr. Chuck Jahren is the W. A. Klinger Teaching Professor and the associate chair for Iowa State University's Construction Engineering Program, housed by the university's civil, construction and environmental engineering department. He earned his bachelor's degree in civil engineering (1977) and his master's in business administration (1982) from the University of Minnesota. He went on to earn his doctorate in civil engineering (1987) from Purdue University. He has more than six years of industrial experience as a bridge construction project engineer and as a research engineer. His research interests include the use of advance technologies to improve learning, electronic data exchange in design and construction, construction process improvement and road and maintenance and rehabilitation. His teaching interests include construction process design, construction equipment and cost estimating.

