Galindo, J. A., Ruiz, E. G., Benavides, D.: Variabilidad en visualización de datos: retos y posibilidades. In: Pérez-Benedí, J. (ed.) Actas de las XXIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2019). Sistedes (2019). https://hdl.handle.net/11705/JISBD/2019/053

Variabilidad en visualización de datos: retos y posibilidades

Problema abierto - investigación en curso

José A. Galindo[1], Elvira G. Ruiz[1], and David Benavides[1]

Universidad de Sevilla,
Dept de lenguajes y sistemas informáticos, Sevilla, España
{jagalindo,elvira,benavides}@us.es

Resumen Los sistemas de visualización de la información nos permiten visualizar datos usando abstracciones de los mismos, por ejemplo, usando gráficos de barras o de tartas. No obstante, la diversidad de visualizaciones dificulta la correcta elección de los sistemas más apropiados para cada conjunto de datos. La ingeniería de líneas de producto y sistemas de alta variabilidad ha generado múltiples técnicas que permiten la configuración óptima de productos software dados unos requisitos o características. En este trabajo proponemos el uso y adaptación de las técnicas de configuración, derivación y análisis automático existentes en el área de lineas de producto software al contexto de la visualización, permitiendo de esta forma el guiado sobre las opciones de configuración para visualizar un conjunto de datos.

Keywords: Visualización de datos · Sistemas de alta variabilidad · Ingeniería de líneas de producto.

1. Introducción

En la sociedad de la información en la que estamos sumergidos vivimos con dispositivos que generan y almacenan una ingente cantidad de datos, como por ejemplo dispositivos móviles, redes sociales o casas inteligentes entre otros [3]. Esta cantidad y diversidad de datos hace complicada la extracción y presentación de la información relevante para el consumo de los usuarios finales.

Existen diferentes soluciones software que permiten generar una multitud de visualizaciones distintas para un mismo conjunto de datos, como pueden ser diagramas de barras, mapas de calor o gráficos de tartas, entre muchas otras. Algunas de las herramientas más conocidas son R[1] o Grafana[2]. No obstante, estas opciones no ofrecen soporte para guiar a los usuarios sobre la forma más correcta de visualizar los datos. Esto es, decidir qué visualización es la más apropiada para un conjunto de datos con ciertas propiedades. Esto hace que podamos encontrar muchas formas de visualizar los datos que en unos casos no

[1] https://www.r-project.org/
[2] https://grafana.com/

Galindo, J. A., Ruiz, E. G., Benavides, D.: Variabilidad en visualización de datos: retos y posibilidades. In: Pérez-Benedí, J. (ed.) Actas de las XXIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2019). Sistedes (2019). https://hdl.handle.net/11705/JISBD/2019/053

permitirán comunicar la información deseada de una manera eficiente y efectiva. Un ejemplo común de una visualización de datos poco recomendable es el uso de gráficos de tartas con datos de entrada que no representan una relación entre una parte y un todo, puesto que hay evidencias de percepción cognitiva que apuntan a que el ojo humano no está bien preparado para percibir las diferencias de tamaño de un ángulo.

Por otra parte, la gestión de la variabilidad es una disciplina que se focaliza en la gestión de las partes comunes y variables de un conjunto de productos. Por ejemplo, dentro de los dispositivos móviles basados en Android podemos encontrar dispositivos con distintas pantallas, opciones de conectividad o procesador aún mientras comparten el sistema operativo [6]. Para representar esta información, investigaciones previas propusieron el uso de distintos modelos que permiten la representación de las partes comunes y variables de un sistema[7]. Además, dada la enorme cantidad de opciones disponibles en un sistema complejo, se crearon técnicas asistidas por ordenador para permitir, entre otras operaciones, la configuración de dichos sistemas complejos[5,2].

El objetivo de este trabajo es proponer el uso de mecanismos de gestión de la variabilidad a los motores software de visualización de datos para permitir el guiado de los usuarios cuando hacen uso de estas herramientas. Esto permitiría reducir la probabilidad de generar visualizaciones incorrectas o poco adecuadas a los datos de entrada. Asimismo, estos motores deberán ser lo suficientemente flexibles para permitir que el usuario elija entre un catálogo de posibles visualizaciones para que, de esta forma, pueda elegir, la que muestre de mejor manera los datos en los que está interesado.

Para conseguir este objetivo nos planteamos dos subobjetivos. El primero consiste en la caracterización de los distintos tipos de datos y visualizaciones, así como las dependencias funcionales existentes entre ellas. De esta forma podríamos modelar la variabilidad existente en las visualizaciones. Esto, permitiría el análisis e identificación de las visualizaciones más adecuadas para un cierto tipo de datos. En segundo lugar, nos proponemos habilitar técnicas que permitan asistir a los usuarios cuando estén generando las distintas visualizaciones. Para esto, evaluaremos distintas opciones para dotar de mecanismos inteligentes al sistema de guiado.

2. Motivación y retos

La visualización de estos datos tiene un alto impacto en cómo los asimilamos y qué conclusiones extraemos de ellos. Para dar soporte a la generación de estas visualizaciones se han desarrollado multitud de herramientas software que permiten generar una amalgama de alternativas para representar o visualizar datos.

Estas distintas visualizaciones de datos las encontramos en nuestro día a día, como, por ejemplo, en los análisis de datos biométricos que nos ofrecen las pulseras de monitoreo de actividad deportiva, en los análisis de encuestas de intención de voto o en las aplicaciones móviles para consultar el clima. Este

Galindo, J. A., Ruiz, E. G., Benavides, D.: Variabilidad en visualización de datos: retos y posibilidades. In: Pérez-Benedí, J. (ed.) Actas de las XXIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2019). Sistedes (2019). https://hdl.handle.net/11705/JISBD/2019/053

contacto cercano con los distintos tipos de visualización hace que la importancia de ofrecer visualizaciones correctas sea aún mayor que la que teniendo en campos tan diversos como la economía o el periodismo.

De cara a ofrecer unas representaciones de datos claras y útiles, investigadores y usuarios han propuesto durante años distintos mecanismos de visualización como pueden ser los gráficos de barras, los gráficos de frecuencia o los mapas de calor. No obstante, no todas las infografías son adecuadas para cualquier tipo de datos. Por ejemplo, los gráficos de barras no son recomendados cuando vamos a representar variables continuas. Más aún, puede haber casos en el que los mismos datos se puedan representar de forma diferente de tal forma que se ofrezca una perspectiva distinta de la realidad que representan.

Es aquí donde se hace notar la ausencia de sistemas de guiado en la toma de decisiones de para generar las distintas visualizaciones, es decir, dada la diversidad de opciones disponibles cuando el usuario desea crear una nueva visualización es común que acabe generando visualizaciones no recomendables o incluso no válidas. La solución existente en la actualidad requiere, por lo tanto, la experiencia y conocimiento experto en el área de visualización para generar la "gráfica" más adecuada para un conjunto de datos.

Reto 1: Caracterización de los datos para visualización. Como primer paso para conseguir los objetivos, es necesario poder caracterizar las distintas propiedades de los datos de entrada para permitir la configuración de las posibles visualizaciones a generar. Por ejemplo, una variable discreta se puede visualizar como un gráfico de barras de una manea más intuitiva que una continua, la cual requeriría de gráficos de área o densidad entre otros. Para poder analizar los tipos de datos de entrada y caracterizarlos podemos vislumbrar técnicas basadas en análisis de regresión, machine learning, o algoritmos evolutivos de forma similar a la ingeniería inversa de modelos de características [8].

Reto 2: Asistencia en el proceso de visualización. Una vez caracterizados los datos a representar, sería necesario identificar las necesidades de los consumidores de las visualizaciones. Por ejemplo, si el público objetivo de nuestra visualización tiene problemas para visualizar colores, deberíamos optar por visualizaciones que permitan evitarlos. Asimismo, podemos implementar las dependencias entre los tipos de entrada de datos y sus posibles visualizaciones recomendadas en este paso.

3. Antecedentes y Conclusiones

Los sistemas altamente configurables son aquellos sistemas software que pueden ser personalizados para distintos escenarios de uso. Estos sistemas hacen un uso intensivo de técnicas de gestión de variabilidad para un funcionamiento eficaz. En este área encontramos resultados centrados en el aspecto multimedia en la generación guiada de vídeos sintéticos para probar algoritmos de seguimiento de objetos [4,1] o la generación de elementos artísticos[9].

Este contenido multimedia tiene peculiaridades, como por ejemplo un mayor uso de datos continuos para codificarlos, que hacen que tradicionalmente no se

Galindo, J. A., Ruiz, E. G., Benavides, D.: Variabilidad en visualización de datos: retos y posibilidades. In: Pérez-Benedí, J. (ed.) Actas de las XXIV Jornadas de Ingeniería del Software y Bases de Datos (JISBD 2019). Sistedes (2019). https://hdl.handle.net/11705/JISBD/2019/053

hayan aplicado técnicas de ingeniería del software para su mejora. Desde ese punto de vista, se propone generalizar y aplicar resultados procedentes de la experiencia en líneas de producto a la visualización de datos.

En este artículo de prospección hemos evaluado la necesidad y beneficios de aunar las soluciones de ingeniería del software para el análisis y guiado de la generación de visualizaciones de datos.

Acknowledgements

Este trabajo ha sido financiado por el programa FEDER el proyecto del MINECO OPHELIA (RTI2018-101204-B-C22), el programa Juan de la Cierva; la red TASOVA (MCIU-AEI TIN2017-90644-REDT); y la Junta de Andalucia con el proyecto METAMORFOSIS.

Referencias

1. Mauricio Alférez, Mathieu Acher, José A. Galindo, Benoit Baudry, and David Benavides. Modeling variability in the video domain: language and experience report. *Software Quality Journal*, 27(1):307–347, 2019.
2. David Benavides, Sergio Segura, and Antonio Ruiz Cortés. Automated analysis of feature models 20 years later: A literature review. *Inf. Syst.*, 35(6):615–636, 2010.
3. Cisco. Data networking index. https://www.cisco.com/c/en/us/solutions/service-provider/visual-networking-index-vni/index.html, 2019. [Online; accessed 9-April-2019].
4. José A. Galindo, Mauricio Alférez, Mathieu Acher, Benoit Baudry, and David Benavides. A variability-based testing approach for synthesizing video sequences. In *International Symposium on Software Testing and Analysis, ISSTA '14, San Jose, CA, USA - July 21 - 26, 2014*, pages 293–303, 2014.
5. José A. Galindo, David Benavides, Pablo Trinidad, Antonio-Manuel Gutiérrez-Fernández, and Antonio Ruiz-Cortés. Automated analysis of feature models: Quo vadis? *Computing*, Aug 2018.
6. José A. Galindo, Hamilton A. Turner, David Benavides, and Jules White. Testing variability-intensive systems using automated analysis: an application to android. *Software Quality Journal*, 24(2):365–405, 2016.
7. Kyo C Kang, Sholom G Cohen, James A Hess, William E Novak, and A Spencer Peterson. Feature-oriented domain analysis (FODA) feasibility study. Technical report, DTIC Document, 1990.
8. Roberto Erick Lopez-Herrejon, Lukas Linsbauer, José A. Galindo, José Antonio Parejo, David Benavides, Sergio Segura, and Alexander Egyed. An assessment of search-based techniques for reverse engineering feature models. *Journal of Systems and Software*, 103:353–369, 2015.
9. Jabier Martinez, Gabriele Rossi, Tewfik Ziadi, Tegawendé François D. Assise Bissyandé, Jacques Klein, and Yves Le Traon. Estimating and predicting average likability on computer-generated artwork variants. In *Genetic and Evolutionary Computation Conference, GECCO 2015, Madrid, Spain, July 11-15, 2015, Companion Material Proceedings*, pages 1431–1432, 2015.