

EVALUACIÓN Y MEJORA DE NARRATIVAS DIGITALES BASADAS EN DATOS

Emanuel Irrazabal; Andrea Lezcano Airaldi; Joaquín Acevedo Duprato

Facultad de Ciencias Exactas y Naturales y Agrimensura.
Universidad Nacional del Nordeste
ISIS
{eirrazabal, alezcano}@exa.unne.edu.ar

RESUMEN

Esta línea se enmarca en el Proyecto F018-2017; una continuación de los proyectos F07-2009 y F10-2013, ambos enfocados en modelos, métodos y herramientas para la calidad del software de Universidad Nacional del Nordeste. Este proyecto aborda los temas emergentes en el área de la calidad de software, en particular, aspectos referidos a la visualización de la información aplicando técnicas narrativas o de *storytelling* y a las buenas prácticas asociadas a ella. Esta línea es apoyada por CONICET a través de la Beca Interna Doctoral, con una duración de 60 meses. Se pretende generar métodos y herramientas que permitan evaluar y mejorar la calidad de las visualizaciones en los productos software.

En particular, se está trabajando en la definición de un enfoque para la evaluación de la calidad y mejora asistida de visualizaciones aplicando técnicas narrativas. Esto ha incluido el desarrollo de un estudio de caso y de revisiones sistemáticas de la literatura.

Palabras clave: calidad de software, data storytelling, narrativa basada en datos, visualización de la información.

CONTEXTO

Las líneas de Investigación y Desarrollo presentada en este trabajo corresponden al proyecto PI-17F018 “Metodologías y herramientas emergentes para contribuir con la calidad del software”, acreditado por

la Secretaría de Ciencia y Técnica de la Universidad Nacional del Nordeste (UNNE) para el periodo 2018-2021 y a la Beca Interna Doctoral de CONICET 2021 – 2025.

1. INTRODUCCIÓN

La narrativa digital o storytelling basada en datos es una estrategia popular y efectiva para transmitir información, siendo su propósito estimular la atención y la energía de la audiencia (o engagement) a través de emociones [1]. Una de las definiciones más utilizadas de storytelling es: “el arte de comunicar ideas a través de historias”[2]. En el contexto actual, donde las posibilidades de toma de datos de distintos sistemas software ha tenido un crecimiento explosivo (Big Data), la visualización de datos proporciona un medio valioso por permitir un análisis asistido (por ejemplo, mediante controles de sumarización, división de los datos, búsquedas contextuales, o gráficos interactivos) [3]. Más aún, dado que las narrativas digitales muchas veces se utilizan para la toma de decisiones, es importante considerar las perspectivas de los usuarios finales en la construcción de dichas narrativas. Esta interrelación entre: i) las fuentes de datos que dan soporte a una narrativa digital, y ii) los intereses, preferencias y necesidades de los destinatarios de la misma, determina en gran medida la calidad de las narrativas a construir. La no consideración de estos aspectos suele ocasionar problemas relacionados con: baja calidad del

contenido visual, pobre utilización de las fuentes de datos, ineficiencia en el desarrollo del contenido y, en general, poca entrega de valor.

En 2001, Gershon y Page propusieron por primera vez el uso del *storytelling* para la visualización[4], mientras que Segel y Heer [5] (en 2010) introducen el concepto de visualizaciones narrativas basadas en datos analizando una gran cantidad de ejemplos. En el contexto actual de desarrollos ágiles, uno de los problemas recurrentes es la poca calidad del contenido visual [6]. Por ello, la implementación de las buenas prácticas de *storytelling* en la construcción de un sistema y su posterior evaluación, pueden tener un impacto significativo en la calidad del producto resultante y en la entrega de valor por parte del sistema a los usuarios finales [7][8].

Se han realizado varios estudios que incluyen las prácticas y pautas generales a seguir para crear visualizaciones efectivas [1][3], [10] - [13] o [14]. Si bien se han propuesto algunos criterios para evaluaciones parciales, no se encuentran trabajos exhaustivos, herramientas que asistan a ello o propuestas sistemáticas de modelos para evaluar las narrativas basadas en datos. Entre los criterios descritos en la literatura pueden mencionarse: la aceptación de la historia por parte de los usuarios[15], el impacto de la historia [16], y qué tanto pueden ser recordados los puntos claves [17], el mensaje y el contenido [18]. Otro criterio importante es qué tanto se mantiene la atención de la audiencia durante la entrega de una historia [3] y su grado de compromiso, aunque hay múltiples definiciones, desde ver hasta interactuar, analizar o tomar decisiones [19].

Desde el punto de vista de cómo medir estos criterios, existen estudios que buscan cuantificarlos a partir de los datos estadísticos en los soportes digitales de las visualizaciones, principalmente los sitios Web. Así, por ejemplo, la cantidad de clics determinaría el nivel de audiencia o el

compromiso [1]. Otros trabajos estudian y adaptan las evaluaciones para medir el tiempo de permanencia en una página o la cantidad de visitas del mismo usuario [20]. En [21], el nivel de impacto también se cuantifica a partir de la cantidad de unidades de historia recordadas.

De acuerdo con lo expuesto anteriormente, la hipótesis de trabajo plantea que: “la definición de un modelo de evaluación para narrativas digitales basadas en datos que asimile los esfuerzos parciales existentes y los materialice en un entorno semiautomático de asistencia, puede mejorar el valor entregado de las narrativas respecto sus usuarios finales”. En este caso, el valor entregado se relaciona directamente con el soporte para la toma de decisiones por parte de los usuarios. Se prevé una implementación del modelo vía métricas y refactorizaciones en función de dichas métricas, tomando analogías de otros dominios de desarrollo de software, tales como: requerimientos [22] diseño detallado [23] y documentación [6].

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El objetivo principal de la investigación es “la definición de un enfoque para la evaluación de la calidad y mejora asistida de narrativas digital basadas en datos”. Este objetivo se descompone en tres partes, atendiendo tanto aspectos metodológicos como tecnológicos:

1. Identificar un conjunto de acciones (o patrones) comúnmente utilizados en el diseño de narrativas digitales efectivas.
2. Elaborar un modelo de evaluación de la calidad de una narrativa digital, y particularmente sus formas de visualización en el marco de sitios Web.
3. Desarrollar herramientas para asistir en la aplicación del modelo de evaluación, y adicionalmente asistir en la provisión (semiautomática) de recomendaciones de mejora para el

equipo encargado del diseño de las visualizaciones.

La contribución esperada del trabajo consiste en la provisión de guías accionables (es decir, mediante un entorno semiautomatizado) para evaluar la calidad de las narrativas digitales basadas en datos y orientadas a la toma de decisiones, y por ende mejorar el proceso de construcción de estas.

3. RESULTADOS OBTENIDOS/ESPERADOS

En el marco de este proyecto y respecto de la línea de gestión de procesos se llevó a cabo un estudio de caso comparando con el objetivo de fue determinar los beneficios de incluir las buenas prácticas de visualización de datos en el desarrollo de un sistema realizado en el contexto actual de crisis sanitaria con tiempos límite restringidos. Para ello, se presentó el estudio de caso realizado sobre 16 gráficos de un sistema de información que gestiona el seguimiento del aislamiento y los permisos de circulación en la cuarentena por la pandemia de COVID-9.

Los resultados del estudio evidencian que el uso de técnicas de narrativa basada en datos contribuye a facilitar el proceso de toma de decisiones mediante la implementación de buenas prácticas de visualización que aumentaron la comprensión y la memorabilidad de los gráficos. Asimismo, se comprueba que la causa por la cual no se aplicaron las buenas prácticas de visualización fue la falta de conocimiento, no la falta de tiempo. En ausencia de estas habilidades, se utilizan los ajustes predeterminados de las herramientas de desarrollo disminuyendo su comprensión.

Para este estudio se llevó a cabo el test de alfabetización visual [24] en el cual los participantes obtuvieron puntajes significativamente mayores cuando se trataba de visualizaciones que implementaban buenas prácticas,

demostrando una mayor comprensión con este tipo de gráficos.

Por otro lado, se está realizando una revisión sistemática de la literatura con el objetivo de caracterizar los modelos de evaluación de visualizaciones reportados al momento. En este contexto, se analizaron una serie de estudios secundarios sobre problemáticas similares, resumidos en la Tabla 1.

Tabla I: Estudios secundarios analizados por categoría

Categoría	# de RSL
Criterios de evaluación	3
Estrategias de evaluación	6
Problemas asociados al uso de visualizaciones	6
Dominios de aplicación de las visualizaciones	4

Para esta línea de trabajo se espera desarrollar un modelo de medición de calidad de visualizaciones a partir de la evaluación de las buenas prácticas implementadas.

4. FORMACIÓN DE RECURSOS HUMANOS

En esta línea de trabajo del Grupo de Investigación sobre Calidad de Software (GICS) están involucrados 3 docentes investigadores, 1 becario interno CONICET y 1 becario de investigación de pregrado.

5. REFERENCIAS

- [1] Nussbaumer Knaflic, C. *Storytelling with Data: A data visualization guide for business professionals*. Hoboken, New Jersey: John Wiley & Sons, Ltd., 2015.
- [2] Schreyögg, C. *Knowledge Management and Narratives: Organizational Effectiveness Through Storytelling* (9783503090297). 2006.
- [3] Kosara, R., Mackinlay, J. *Storytelling: The Next Step for Visualization*.

- Computer, Vol 46, Issue 5, 44-50 (2013)
- [4] Gershon, N., Page, W. What storytelling can do for information visualization. *Communications of the ACM*, 44(8):31-37, 2001.
- [5] Segel E., Heer. J. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139 – 1148, Dec. 2010.
- [6] Nicoletti, M., Schiaffino, S., Díaz Pace, J. A.: An optimization-based tool to support the cost-effective production of software architecture documentation. *J. Softw. Evol. Process.* 27(9): 674-699 (2015).
- [7] Baker, R. *Agile UX Storytelling*. 2017.
- [8] Zimmerman, B. Applying Tufte's principles of information design to creating effective Web sites. *ACM SIGDOC Annu. Int. Conf. Comput. Doc. Proc.*, pp. 309–317, 1997, doi: 10.1145/263367.263406.
- [9] Krausz, R. M., et al. E-Health in crisis management-emergency response to COVID 19 in Canada: Platform Development and Implementation. *JMIR Public Health and Surveillance* (2020).
- [10] Kosara, R. An Argument Structure for Data Stories. *Proc. Eurographics/IEEE VGTC Symposium on Visualization (EuroVis)*, 2017.
- [11] Tong, C. et al., "Storytelling and visualization: A survey," *VISIGRAPP 2018 - Proc. 13th Int. Jt. Conf. Comput. Vision, Imaging Comput. Graph. Theory Appl.*, vol. 3, pp. 212 224, 2018, doi: 10.5220/0006601102120224.
- [12] Boy, J., Detienne, F., Fekete, J. Storytelling in information visualizations: Does it engage users to explore data?. *Conf. Hum. Factors Comput. Syst. - Proc.*, vol. 2015-April, pp. 1449–1458, 2015.
- [13] Tufte, E. R. *The Visual Display of Quantitative Information*. 1983.
- [14] Ram D., et al. Case Report: Rapid Development of Visualization Dashboards to Enhance Situation Awareness of COVID-19 Telehealth Initiatives at a Multi-Hospital Healthcare System. *Journal of the American Medical Informatics Association* (2020).
- [15] Lee, B., Kazi, R., Smith, G. SketchStory: Telling more engaging stories with data through freeform sketching. *IEEE Trans. on Visualization and Computer Graphics (Proc. of InfoVis)*, 19(12):2416–2425, 2013.
- [16] Borkin, M., Vo, A., Bylinskii, A., Isola, P., Sunkavalli, S., Oliva, A., Pfster, H. What makes a visualization memorable? *IEEE Trans. on Visualization and Computer Graphics (Proc. of InfoVis)*, 19(12):2306–2315, 2013.
- [17] Kim, K., Boy J., Lee S., Yi, J., Elmqvist, N. Towards an open visualization literacy testing platform, 2014.
- [18] Bateman, S., Mandryk, R., Gutwin, C., Genest, A., McDine, D., Brooks, C. Useful junk?: The effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, pp. 2573–2582, 2010.
- [19] Mahyar, S., Kim, S., Kwon, B. Towards a taxonomy for evaluating user engagement in information visualization. In *Proc. of the IEEE VIS 2015 Workshop Personal Visualization*, 2015.
- [20] Saraiya, P., North, C., Lam, V., Duca, K. An insight-based longitudinal study of visual analytics. *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, 12(6):1511–1522, 2006.

- [21] Borkin, M., Bylinskii, Z., Kim, N., Bainbridge, C., Yeh, C., Borkin, D., Pfster, H., Oliva, A. Beyond memorability: Visualization recognition and recall. *IEEE Trans. on Visualization and Computer Graphics*.
- [22] Rago, A., Marcos, C., Diaz-Pace, J. A.: Identifying duplicate functionality in textual use cases by aligning semantic actions. *Software and Systems Modeling* 15(2): 579-603 (2016).
- [23] Vidal, S., Berra, I, Zulliani, S., Marcos, C., Díaz Pace, J. A.: Assessing the Refactoring of Brain Methods. *ACM Trans. Softw. Eng. Methodol.* 27(1): 2:1-2:43 (2018)
- [24] S. Lee, S. H. Kim, and B. C. Kwon, "VLAT: Development of a Visualization Literacy Assessment Test," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 1, pp. 551–560, 2017, doi: 10.1109/TVCG.2016.2598920.