



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Cuajimalpa

12 de agosto de 2021.
Dictamen C.I. 07/2021

DICTAMEN
QUE PRESENTA LA COMISIÓN DE INVESTIGACIÓN DE LA DIVISIÓN DE CIENCIAS DE LA
COMUNICACIÓN Y DISEÑO

ANTECEDENTES

- I. El Consejo Divisional de Ciencias de la Comunicación y Diseño, en la sesión 03.21, celebrada el 29 de enero de 2021, integró esta Comisión en los términos señalados en el artículo 55 de Reglamento Interno de los Órganos Colegiados Académicos.

- II. El Consejo Divisional designó para esta Comisión a los siguientes integrantes:
 - a) Órganos personales:
 - ✓ Dra. Margarita Espinosa Meneses, Jefa del Departamento de Ciencias de la Comunicación;
 - ✓ Dra. Erika Cecilia Castañeda Arredondo, Jefa del Departamento de Teoría y Procesos del Diseño;
 - ✓ Dr. Carlos Joel Rivero Moreno, Jefe del Departamento de Tecnologías de la Información.

 - b) Representantes propietarios:
 - Personal académico:
 - ✓ Departamento de Ciencias de la Comunicación;
 - ✓ Dra. Lucero Fabiola García Franco, Departamento de Teoría y Procesos del Diseño.
 - ✓ Dr. Alfredo Piero Mateos Papis, Departamento de Tecnologías de la Información.

CONSIDERACIONES

- I. La Comisión recibió, para análisis y discusión, el informe de actividades académicas desarrolladas por el **Dr. Erick De Jesús López Ornelas**, durante el disfrute del año sabático comprendido del 27 de julio de 2020 al 26 de julio de 2021.



División de Ciencias
de la Comunicación
y Diseño

Unidad Cuajimalpa

DCCD | División de Ciencias de la Comunicación y Diseño
Oficina Técnica del Consejo Divisional
Torre III, 5to. piso, Av. Vasco de Quiroga 4871,
Colonia Santa Fe Cuajimalpa, Alcaldía Cuajimalpa de Morelos,
C.P. 05348, Ciudad de México.
Tel: (+52) 55.5814.3505
<http://dccd.cua.uam.mx>



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Cuajimalpa

- II. El periodo sabático fue aprobado en la Sesión 03.20 celebrada el 27 de febrero de 2020 mediante Acuerdo DCCD.CD.15.03.20 del Consejo Divisional de Ciencias de la Comunicación y Diseño.
- III. La Comisión de Investigación sesionó vía remota el día 12 de agosto de 2021, fecha en la que concluyó su trabajo de análisis y evaluación del informe, con el presente Dictamen.
- IV. La Comisión contó, para su análisis, con los siguientes elementos:
 - Programa de actividades académicas por desarrollar durante el periodo sabático.
 - Evaluación general.
- V. La Comisión evaluó el informe de actividades académicas, las constancias y documentos que demuestran las actividades realizadas por el **Dr. Erick De Jesús López Ornelas**, durante el disfrute del año sabático comprendido del 27 de julio de 2020 al 26 de julio de 2021.

El profesor realizó una estancia de investigación en el Instituto de Matemáticas de Toulouse, Francia, estancia supervisada por el Dr. Paulo Carrillo Rouse. A pesar de la problemática provocada por el virus SARS-CoV-2, pudo trabajar de manera satisfactoria en la institución receptora, enfocándose principalmente en dos líneas de investigación relacionadas con el análisis de redes sociales y la teoría de grafos; y la visualización de datos. Derivado de este trabajo se obtuvieron tres publicaciones.

Se buscó, además, impulsar la vinculación entre la UAM y la institución receptora teniendo una relación estrecha con las oficinas de la UAM-Europa y participar en redes de investigación.

Actividades realizadas:

- Participación en la Red REHVIF.
- Participación en la Red PROF-XXI.
- Construcción del estado del arte sobre Análisis de Redes Sociales.
- Construcción del estado del arte sobre Visualización de datos.
- Publicaciones:
Abascal-Mena R., López-Ornelas E. (2020) Author detection: analyzing tweets by using a Naïve Bayes classifier. Journal of Intelligent & Fuzzy Systems. Vol. 39, no. 2, pp. 2331-2339. ISSN print 1064-1246, ISSN online 1875-8967.



División de Ciencias
de la Comunicación
y Diseño

Unidad Cuajimalpa

DCCD | División de Ciencias de la Comunicación y Diseño

Oficina Técnica del Consejo Divisional

Torre III, 5to. piso, Av. Vasco de Quiroga 4871,

Colonia Santa Fe Cuajimalpa, Alcaldía Cuajimalpa de Morelos.

C.P. 05348, Ciudad de México.

Tel.: (+52) 55 5814 3505

<http://dccd.cua.uam.mx>



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Cuajimalpa

López-Ornelas E., Abascal-Mena R. (2020) Prototipo digital para fomentar la civilidad en redes sociales. Pistas Educativas. Vol. 42. No. 137. pp. 1062-1074. ISSN: 2448-847X

López-Ornelas E., Abascal-Mena R. (2020) An interactive prototype to spot Fake News in young people. Proceedings of the VI Iberoamerican Conference of Computer Human Interaction. pp. 37-44. Vol. 2747. ISSN 1613-0073. Arequipa, Perú.

DICTAMEN

ÚNICO:

Se recomienda al Consejo Divisional dar por recibido el informe del año sabático del **Dr. Erick de Jesús López Ornelas**, conforme al plazo establecido en el artículo 231 del Reglamento de Ingreso, Promoción y Permanencia del Personal Académico y del mismo se advierte que cumplió satisfactoriamente con el programa de actividades.

VOTOS:

Integrantes	Sentido de los votos
Dra. Margarita Espinosa Meneses	A favor
Dra. Erika Cecilia Castañeda Arredondo	A favor
Dr. Carlos Joel Rivero Moreno	A favor
Dra. Lucero Fabiola García Franco	A favor
Dr. Alfredo Piero Mateos Papis	A favor
Total de los votos	5 votos a favor

Coordinadora

Dra. Gloria Angélica Martínez De la Peña
Presidenta en funciones del Consejo Divisional de
Ciencias de la Comunicación y Diseño



División de Ciencias
de la Comunicación
y Diseño

Unidad Cuajimalpa

DCCD | División de Ciencias de la Comunicación y Diseño
Oficina Técnica del Consejo Divisional
Torre III, 5to. piso, Av. Vasco de Quiroga 4871,
Colonia Santa Fe Cuajimalpa, Alcaldía Cuajimalpa de Morelos,
C.P. 05348, Ciudad de México.
Tel.: (+52) 55.5814.3505
<http://dccd.cua.uam.mx>



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA
Unidad Cuajimalpa

*Comunidad académica comprometida
con el desarrollo humano de la sociedad.*

06 de agosto de 2021

Dr. Carlos Joel Rivero Moreno

Jefe del Departamento de Tecnologías de la Información
División de Ciencias de la Comunicación y Diseño
Presente

ASUNTO: Informe finalización del período sabático

Estimado Dr. Carlos Joel Rivero Moreno,

Por este medio me permito hacer entrega del informe del período sabático realizado del **27 de julio de 2020 al 26 de julio de 2021**. Es importante mencionar que en el documento adjunto ya se encuentran todos los anexos que complementan el informe.

Con lo anterior estoy cumpliendo en tiempo y forma con lo establecido por el artículo 231 del RIPPPA el cual indica lo siguiente: *“Los miembros del personal académico que hubieren disfrutado del periodo sabático, deberán rendir al consejo divisional, un informe detallado y por escrito de las actividades académicas desarrolladas. Al informe deberán acompañarse las constancias y los documentos que demuestren dichas actividades. Este informe deberá ser presentado dentro de los dos meses siguientes a su reincorporación al trabajo”*.

Le agradezco su atención a la presente.

A t e n t a m e n t e

“Casa abierta al tiempo”

Dr. Erick de Jesús López Ornelas

Departamento de Tecnologías de la Información
Universidad Autónoma Metropolitana – Unidad Cuajimalpa

c.c.p: Dra. Gloria Angélica Martínez de la Peña – Directora en función de la DCCD

Unidad Cuajimalpa

Departamento de Tecnologías de la Información
Torre III, 5to. piso. Avenida Vasco de Quiroga 4871, Colonia Santa Fe Cuajimalpa
Delegación Cuajimalpa de Morelos, México, D.F., C.P. 05348
www.cua.uam.mx

Informe final del período sabático

Dr. Erick de Jesús López Ornelas

Fecha del Informe: 06/08/2021

Fecha de inicio y término del periodo sabático: 27/07/2020-26/07/2021

Nombre de la Institución receptora: Institut de Mathématiques de Toulouse, France

Título del proyecto de investigación: Análisis de Redes Sociales y Visualización de datos.

Descripción general:

La estancia sabática se realizó en el Instituto de Matemáticas de Toulouse en Francia. La estancia fue supervisada por el Dr. Paulo Carrillo Rouse (Maître de Conférences à l'Université Paul Sabatier).

Es importante mencionar que, pese a la problemática ligada a la pandemia provocada por el virus SARS-CoV-2, se pudo trabajar de manera satisfactoria en la institución receptora. Durante este periodo de tiempo se trabajaron principalmente dos líneas de investigación relacionadas con: 1) el análisis de redes sociales y la teoría de grafos y 2) la visualización de datos. Derivado de este trabajo se obtuvieron tres publicaciones. Se buscó, además, impulsar la vinculación entre la UAM y la institución receptora teniendo una relación estrecha con las oficinas de la UAM-Europa y participar en redes de investigación.

Actividades realizadas:

1. Participación en la Red REHVIF (Anexo 1)

Se participó en múltiples reuniones con la Red internacional Habitar las Ciudades del Futuro (REHVIF) que es una red integrada por siete plataformas nacionales de investigación multidisciplinarias en Canadá, Colombia, Costa de Marfil, Ecuador, Francia, México y Senegal, y la revista Bitácora Urbano Territorial.

Esta red es la consolidación reciente de dos colaboraciones científicas y educativas desarrolladas a lo largo de varios años. La primera es el eAtlas Francophone de l'Afrique de l'Ouest (eAtlas Francófono de África Occidental, eAtlas FAO), iniciado en 2006 y, la segunda, dos Programas Regionales Francia-América Latina-Caribe (PREFALC), iniciados en 2015. El eAtlas FAO reunió al Laboratorio Interdisciplinario Solidaridad, Sociedades, Territorios del Centro Interdisciplinario de Estudios Urbanos (LISST-CIEU) de la Universidad de Toulouse (Francia), la Universidad Cheikh-Anta-Diop (UCAD) de Dakar (Senegal) y las Universidades de Bamako (Malí), Uagadugú (Burkina Faso), Abomey-Calavi (Benín) y Bouaké (Costa de Marfil). La segunda, PREFALC, centró su interés en

torno a los estudios urbanos y acercó las maestrías asociadas a los grupos de investigación del LISST-CIEU de la Universidad de Toulouse (Francia), de la Escuela del Hábitat (CEHAP) de la Universidad Nacional de Colombia, sede Medellín (Colombia), de la FLACSO (Ecuador) y de la Universidad Autónoma Metropolitana de México (México). Durante estos cinco años, la red ha realizado intercambios de profesores entre los cuatro países, cuestionando los desafíos de la ciudad inclusiva y participativa, e indagando sobre la ciudad habitable, sostenible e inteligente.

El objeto que reúne a la red es lo urbano y su transición está sujeta a interrogantes sobre los escenarios, tendencias, eventos emergentes, incertidumbres y eventos futuros resultantes de las interacciones de tres dominios que conciernen a las ciudades: la transición urbana y las evoluciones de los modos de vida y de habitar; la transición socioecológica; y la transición digital.

2. Participación en la red PROF-XXI (Anexo 2)

Se participó en la red PROF-XXI cuyo objetivo es la construcción de capacidades para el Desarrollo de Centros de Enseñanza y Aprendizaje (CEAs) para las Instituciones Educativas Superiores (IES) de Latinoamérica en el Siglo XXI, teniendo en cuenta el contexto actual de la Región, y el contexto actual particular de las 4 IES latinoamericanas que servirán de modelo (2 IES públicas y 2 IES privadas de distintos tamaños en Guatemala y Colombia).

Este proyecto tiene su origen en la preocupación de las IES de Latinoamérica en mejorar la calidad de sus procesos de enseñanza-aprendizaje, derivados de un crecimiento muy rápido de las instituciones de educación terciaria en la Región. Una de las medidas más urgentes es garantizar que los profesionales y docentes de las IES estén capacitados y tengan apoyo suficiente para llevar a cabo procesos de innovación educativa eficaces, tanto en educación presencial, como en educación virtual, como en educación mezclada o híbrida (blended).

Dentro de los objetivos de la red está el desarrollar guías y modelos de CEAs avanzados, centrados en el desarrollo de competencias docentes en tres ejes: (1) pedagogía, (2) tecnología, y (3) gestión estratégica. Además, diseñar y desarrollar Planes de Formación Docente en los CEAs de las 4 IES latinoamericanas para la gestión de la calidad, en tres ejes: (1) metodologías activas e híbridas; (2) estrategias de monitorización y evaluación mediante analítica del aprendizaje (learning analytics); (3) investigación sobre la propia práctica docente para la mejora continua.

3. Construcción del estado del arte sobre Análisis de Redes Sociales (Anexo 3)

El método de evaluación de las redes se denomina Análisis de Redes Sociales (ARS) y en general es considerado como el estudio de la estructura social, y en un sentido más amplio se puede entender cómo un método cuantitativo por medio del cual se obtiene la estructura social a partir de las regularidades en el patrón de relaciones establecidas entre entidades sociales definidas como personas, grupos u organizaciones. El Análisis de Redes Sociales es un método o conjunto de instrumentos mediante el cual se

conecta a los actores (individuos, organizaciones, etc.) con aquellas estructuras sociales que resultan de las relaciones establecidas por dichos actores.

El reporte técnico completo sobre las diversas formas de análisis de las redes sociales a partir de la teoría matemática de grafos se puede ver en el anexo 3.

4. Construcción del estado del arte sobre Visualización de datos (Anexo 4)

La visualización de datos es la presentación gráfica de información con dos propósitos. Por un lado, la interpretación y construcción de significado a partir de los datos (es decir, el análisis); y por otro lado, la comunicación.

La visualización es una herramienta muy potente para descubrir y comprender la lógica que se encuentra detrás de un conjunto de datos, así como para compartir esta interpretación con otras personas desde un punto de vista objetivo. Como se suele decir: “una imagen vale más que mil palabras”, especialmente cuando el significado que se quiere comunicar se representa mejor gráfica que verbalmente, y esta representación gráfica está diseñada de acuerdo a los principios formales que rigen la visualización de datos.

El reporte técnico completo sobre la visualización de datos se puede ver en el anexo 4.

5. Publicaciones realizadas (Anexo 5)

Durante la estancia sabática se publicaron tres artículos relacionados con este análisis de las redes sociales y la visualización de datos los cuales enlistamos a continuación y se pueden encontrar en el anexo 5.

- Abascal-Mena R., López-Ornelas E. (2020) Author detection: analyzing tweets by using a Naïve Bayes classifier. *Journal of Intelligent & Fuzzy Systems*. Vol. 39, no. 2, pp. 2331-2339. ISSN print 1064-1246, ISSN online 1875-8967. DOI: 10.3233/JIFS-179894
- López-Ornelas E., Abascal-Mena R. (2020) Prototipo digital para fomentar la civilidad en redes sociales. *Pistas Educativas*. Vol. 42. No. 137. pp. 1062-1074. ISSN: 2448-847X
- López-Ornelas E., Abascal-Mena R. (2020) An interactive prototype to spot Fake News in young people. *Proceedings of the VI Iberoamerican Conference of Computer Human Interaction*. pp. 37-44. Vol. 2747. ISSN 1613-0073. Arequipa, Perú.

Dr Erick Lopez Ornelas
Universidad Autonoma Metropolitana (UAM)
Departamento de Tecnologías de la Información
Unidad Cuajimalpa

July 29, 2021

OBJET: Fin du séjour de recherche du Professeur Erick LOPEZ ORNELAS à l'Institut de Mathématiques de Toulouse

Je soussigné, Professeur Paulo CARRILLO ROUSE, agissant en qualité de Maître de Conférences de l'Institut de Mathématiques de Toulouse, atteste que le Prof. Erick LOPEZ ORNELAS a commencé son séjour de recherche au sein de notre laboratoire le 1 septembre 2020, ce séjour a pris fin le 23 juillet 2021.

Nous allons mettre en place un plan de travail plus détaillé dans les jours à venir afin de délimiter les activités de collaboration et de continuité de nos projets de recherche commune.

Je vous prie d'agréer, Madame, mes salutations distinguées,

**INSTITUT de MATHÉMATIQUES
de TOULOUSE**

Université Paul Sabatier - Toulouse III
31062 Toulouse Cedex 9

Dr. Paulo CARRILLO ROUSE
Maître des Conférences
Institut Mathématique de Toulouse
118, route de Narbonne
31062, Toulouse Cedex, FRANCE
E mail: carrillo@math.univ-toulouse.fr

Anexo 1



- TODO

OpenEdition Search

TOUT OPENEDITION

Portail de ressources électroniques en sciences humaines et sociales

[OpenEdition](#)

Nos plateformes

[OpenEdition Books](#) [OpenEdition Journals](#) [Hypotheses Ca enda](#)

Bibliothèques et institutions

[OpenEdition Freemium](#)

Nos services

[OpenEdition Search](#) [Lettre d'OpenEdition](#)

Suivez-nous

Habiter les villes du futur / Habitar las ciudades del futuro



SEMINARIOS / SEMINARIOS

PECHA-KUCHA: ATELIER / TALLER



CYCLE DE CONFÉRENCES / CICLO DE CONFERENCIAS

CYCLE DE CONFÉRENCES INTERNATIONALES 2020 / CICLE DE CONFERENCIAS INTERNACIONALES 2020



SEMINARIOS / SEMINARIAS

DIALOGUE CROISÉ « LES ENJEUX DE NATURE FACE AUX DÉFIS TECHNOLOGIQUES » / DIÁLOGO CRUZADO « LOS DESAFÍOS DE NATURALEZA FRENTE A LOS DESAFÍOS TECNOLÓGICOS »



REHVIF



Habiter les villes du futur

Les mondes urbains en transition et la recomposition des systèmes d'acteurs autour des projets de villes habitables, soutenables et intelligentes

REHVIF est la mise en réseau de 7 plateformes nationales de recherche multidisciplinaires : France, Canada, Colombie, Côte d'Ivoire, Équateur, Mexique et Sénégal (voir les partenaires ci-dessous) portant sur les transformations des villes contemporaines, au cœur de l'habiter dans le contexte de la transition territoriale, écologique et numérique.

Ce réseau s'appuie sur un partenariat entre le monde scientifique et divers acteurs de catégories d'acteurs qui participent à la production de la ville (collectivités locales, acteurs économiques, acteurs de la société civile). Orienté vers la production de connaissance, la formation à la recherche.



Habitar las ciudades del futuro

Mundos urbanos en transición y recomposición de sistemas de actores en territorios habitables, sostenibles e inteligentes

REHVIF es la creación de redes de 7 plataformas nacionales de investigación multidisciplinarias (Colombia, Costa de Marfil, Ecuador, México y Senegal (ver socios e incorporación) sobre las ciudades contemporáneas, en torno a la vida en el contexto de lo territorial, ecológico, y digital.

Es una red que se basa en una alianza entre el mundo científico y diversas categorías de actores de producción de la ciudad (comunidades locales, actores económicos, actores de la sociedad). Orientado hacia la producción de conocimiento, formación investigadora.

La poursuite de l'ambition de nouvelles modalités de paritarisation scientifique et technologique de disposer de nouvelles sciences de la ville » autour de trois axes :

Axe 1 : La transition territoriale : évolutions des formes urbaines et des modes de vie et d'habiter.

Les villes sont d'abord des espaces habités et elles assurent ainsi la consistance d'une forme particulière du lien social, l'urbanité, et du lien politique, la citoyenneté. Les modes de vie et d'habiter sont constitués par un ensemble de pratiques quotidiennes récurrentes et de représentations sociales associées, par un habitus, propres à un groupe ou à une classe sociale ou territoriale. Or, sous l'effet des technologies du transport et du numérique, les villes se sont transformées. Si les modes de vie et d'habiter sont en partie déterminés et construits par les formes urbaines, c'est de moins en moins le cas au vu de l'intensification de la mobilité des citoyens et de l'utilisation des outils numériques par les habitants des métropoles, qui tendent à les arracher du territoire local, avec toutes les conséquences inégalitaires.

Axe 2 : La transition socio-écologique : réorientation de pratiques pour gérer la complexité et l'incertitude.

Appliqué à la ville, le développement durable est un paradigme issu de la lente prise de conscience de l'impact des usages urbains sur l'équilibre environnemental, la conservation en termes de quantité et de qualité des ressources (eau, énergie, sol, matières premières), ainsi que sur le dérèglement climatique. La recherche d'une adéquation entre l'économie, l'environnement, le social implique de prendre en compte les enjeux liés à la prévention des risques et à l'adaptation au changement climatique (îlot de chaleur urbain, inondations, glissements de terrain, risques technologiques, risques sanitaires, sécurité alimentaire). Répondre à de tels défis nécessite l'appropriation de ce paradigme, provoquant des répercussions dans les formes et les échanges urbains et dans les modes de vie et d'habiter ; qui seront influencés par le rôle prépondérant qu'a aujourd'hui le numérique dans nos sociétés.

Axe 3 : La transition numérique : de la « ville intelligente » à l'« intelligence territoriale » et à l'appropriation sociale de technologies numériques.

Le modèle des « villes intelligentes », pour autant qu'il y en ait, est en cours d'élaboration. La toute première histoire, il s'inscrit dans la généalogie de ce que l'on appelle aujourd'hui les « villes numériques » en Europe, ou « *Ciudades Digitales* » en Amérique latine (notamment au sein de la « *Red ibero-americana de las ciudades digitales* »), il prolonge à la fois les dynamiques initiées par les débats politiques autour du « développement durable » et ceux portés sur la « société de l'information ». Bien qu'imprécis, comme en témoigne le débat, ce modèle des « villes intelligentes » peut se définir par le recours systématique ou en voie de systématisation des acteurs de l'urbain, notamment l'acteur public de référence, aux technologies numériques. Ce recours vise à optimiser le fonctionnement de la ville dans plusieurs directions, en particulier dans la gouvernance et dans la production et la mise à disposition des services urbains adaptés aux besoins de la population.

Partenaires du réseau

Su ambición es sacar nuevas modalidades de asociaciones científicas y es o a través de sacar «nuevas ciencias de la ciudad» en torno a tres ejes:

Eje 1 : Transición territorial: evoluciones de las formas urbanas y de modos de vida y habi

Las ciudades son, primeramente, espacios habitados y, por lo tanto, garantizan la consistencia del vínculo social, la urbanidad, y de vínculo político, la ciudadanía. Los modos de vida y por un conjunto de prácticas cotidianas recurrentes y representaciones sociales asociadas, por grupos o clases sociales o territoriales. Ahora bien, especialmente bajo el efecto de las tecnologías digitales, las ciudades se han transformado. Si los estilos de vida y el hábitat y limitados en parte por las formas urbanas, es cada vez menos el caso dada la intensificación de las ciudades y el uso de herramientas digitales por parte de los habitantes del territorio local, aunque con importantes desigualdades.

Eje 2: Transición socioecológica: reorientación de prácticas para gestionar la complejidad

Aplicado a la ciudad, el desarrollo sostenible es un paradigma que surge de la lenta toma de conciencia de los usos urbanos sobre el equilibrio medioambiental, la conservación en términos de recursos (agua, energía, suelo, materias primas), así como sobre el cambio climático. La búsqueda de la adecuación entre la economía, el medio ambiente y la sociedad implica tener en cuenta los aspectos de prevención de riesgos y la adaptación al cambio climático (isla de calor urbana, inundaciones, riesgos tecnológicos, riesgos sanitarios, seguridad alimentaria). Responder a tales desafíos requiere la apropiación de este paradigma, provocando repercusiones en las formas e intercambios urbanos ; que serán influenciados por el papel predominante que tiene la tecnología hoy.

Eje 3: La transición digital: de la « ciudad inteligente » a la « inteligencia territorial » y la tecnologías digitales.

El modelo de « ciudades inteligentes », si existe, está en curso de elaboración. Pero, tiene raíces en la genealogía de las que hasta hace poco se llamaban « ville numérique » en Europa, o América latina (en particular en torno a la « Red iberoamericana de las ciudades digitales »), en tanto que las dinámicas establecidas por los debates políticos en torno al « desarrollo sostenible » y la « sociedad de la información ». Aunque impreciso y muy discutido, este modelo de « ciudad inteligente » puede definirse por el uso sistemático o el desarrollo sistemático de las tecnologías digitales por parte de los actores de l'urbain, especialmente el actor público de referencia, a las tecnologías digitales. Este recurso tiene como objetivo optimizar el funcionamiento de la ciudad en varias direcciones, en particular en la gobernanza y en la prestación de servicios urbanos adaptados a las necesidades de la población.

Socios de la red

Canada:	Université LAVAL Université du Québec à Montréal (UQAM)
Colombie:	Universidad Nacional de Colombia en Medellín (UNAL) Facultad de Minas (UNAL) Universidad Pontificia Bolivariana (UPB)
Côte d'Ivoire:	Université Assane Ouattara de Bouaké (UAO)
Ecuador:	Facultad Latinoamericana de Ciencias Sociales (FLACSO)
France:	Université Toulouse Jean Jaurès (UT2J)
México:	Universidad Autónoma Metropolitana (UAM)
Sénégal :	Université Cheikh Anta Diop de Dakar (UCAD)
Responsable du réseau : Gerente de la red:	Emmanuel EVENO, Université Toulouse Jean Jaurès

[Nous contacter](#)

[Contáctenos](#)

rehvif@univ-tlse2.fr

Ce site utilise Akismet pour réduire les indésirables. [En savoir plus sur comment les données de vos commentaires sont utilisées.](#)

Rechercher dans OpenEdition Search

Vous allez être redirigé vers OpenEdition Search

Expression ou mot-clé

Dans tout OpenEdition

Dans Habiter les villes du futur

RECHERCHER

Anexo 2



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

PRESENTACIÓN PROYECTO PRO-XXI A LA UNIVERSIDAD AUTÓNOMA METROPOLITANA

([HTTPS://EUROPE.UAM.MX/](https://europe.uam.mx/))



© 11 FEBRERO, 2021 [LYVIER \(HTTPS://EUROPE.UAM.MX/BLOG/AUTHOR/LYVIER/\)](https://europe.uam.mx/blog/author/lyvier/)
 □ FILED UNDER: [DOCENCIA \(HTTPS://EUROPE.UAM.MX/BLOG/CATEGORY/DOCENCIA/\)](https://europe.uam.mx/blog/category/docencia/), [EDUCACIÓN \(HTTPS://EUROPE.UAM.MX/BLOG/CATEGORY/EDUCACION/\)](https://europe.uam.mx/blog/category/educacion/), [EUROPA \(HTTPS://EUROPE.UAM.MX/BLOG/CATEGORY/EUROPA/\)](https://europe.uam.mx/blog/category/europa/), [FRANCIA \(HTTPS://EUROPE.UAM.MX/BLOG/CATEGORY/FRANCIA/\)](https://europe.uam.mx/blog/category/francia/), [INVESTIGACIÓN \(HTTPS://EUROPE.UAM.MX/BLOG/CATEGORY/INVESTIGACION/\)](https://europe.uam.mx/blog/category/investigacion/), [MÉXICO \(HTTPS://EUROPE.UAM.MX/BLOG/CATEGORY/MEXICO/\)](https://europe.uam.mx/blog/category/mexico/)

El 27 de Enero se realizó la presentación del proyecto PROF-XXI a la Oficina de Cooperación y Vinculación de la UAM en Europa.

La Dra. Mar Pérez Sanagustín, socia del proyecto PROF-XXI de la Université Paul Sabatier Toulouse III (Francia), presentó los



Profesores UAM en Francia

objetivos y el plan de trabajo enfocado en los Centros de enseñanza y aprendizaje (CEAs) en Iberoamérica.

Entre los profesores que asistieron se encuentran:

- Dr. Salomón González Arellano – Responsable de la Oficina de la UAM en Europa.
- Dra. Rocío Abascal – Departamento de Tecnologías de la Información / UAM Cuajimalpa
- Dr. Erick López – Departamento de Tecnologías de la Información / UAM Cuajimalpa

PREVIOUS POST



uameurope@correo.uam.mx
(<mailto:uameurope@correo.ec.uam.mx>).



+ 33 55 6110 0145



41 allées Jules Guesde -
31003 Toulouse Cedex -
France



[Sitio web UAM en México](http://www.uam.mx/)
(<http://www.uam.mx/>).

© UNIVERSIDAD AUTÓNOMA METROPOLITANA 2018. OFICINA DE COOPERACIÓN Y VINCULACIÓN EN EUROPA.

Anexo 3

Reporte Técnico

Análisis de redes sociales

Erick López Ornelas

Universidad Autónoma Metropolitana – Cuajimalpa

El método de evaluación de las redes se denomina Análisis de Redes Sociales (ARS) y en general es considerado como el estudio de la estructura social, y en un sentido más amplio se puede entender cómo un método cuantitativo por medio del cual se obtiene la estructura social a partir de las regularidades en el patrón de relaciones establecidas entre entidades sociales definidas como personas, grupos u organizaciones.

El Análisis de Redes Sociales es un método o conjunto de instrumentos mediante el cual se conecta a los actores —individuos, organizaciones, etc.— con aquellas estructuras sociales que resultan de las relaciones establecidas por dichos actores.

Características del análisis de redes sociales

El Análisis de Redes Sociales debe centrarse esencialmente sobre las relaciones de los actores y no sobre los atributos de los mismos, las estructuras de relaciones tienen un poder explicativo más importante que los atributos personales de los miembros que componen el sistema. Desde este principio se considera que el ARS no se centra en operaciones taxonómicas de agrupamiento de individuos de acuerdo a sus rasgos o variables en común, sino que centra su estudio en la determinación de las estructuras de relaciones que los individuos establecen. El Análisis de Redes Sociales entonces no se centra en atributos sino en las relaciones, y se debe privilegiar la interacción entre los actores de la red.

El ARS se centra en la evaluación del comportamiento real de los actores dentro de la red, se logra con ello medir y cuantificar la dinámica real de interacciones e intercambio de la estructura social, así mismo, permite la evaluación de la conducta frente al control y la normatividad de la red, identificando restricciones y oportunidades. Los intercambios están ajustados a condiciones estructurales como la posición en la red de los actores. La red no está determinada por organizaciones inflexibles ni por sistemas cerrados sino que los actores se ven incluidos en múltiples redes o círculos sociales, lo cual les inserta a su vez en numerosas y complejas relaciones de poder.

Terminología del análisis de redes sociales (ARS)

El análisis de redes ha desarrollado con el paso del tiempo su propio sistema de expresión, un conjunto de conceptos particulares ha sido utilizado regularmente. En general se pueden mencionar cuatro grandes mecanismos que permiten la valoración de las particularidades de las redes, dentro de tales mecanismos se incluyen a) la definición de las propiedades generales con que

cuenta la red, b) el análisis de las características posicionales de los actores, c) la identificación de subagrupaciones con sus respectivas características y d), el método de visualización de redes, que constituye uno de los más valiosos recursos analíticos para comprender las dinámicas relacionales y la formación de estas estructuras sociales, sin embargo, previo a la descripción de estos enfoques evaluativos es válido y necesario una delimitación de las características esenciales relacionadas con la generación de grafos de red y las propiedades que los mismos exhiben.

Grafos, redes y sus conexiones

La teoría de los grafos es esencial en el análisis de redes pues proporciona un lenguaje formal que facilita estudiar la red y su estructura. Se definen los grafos como un conjunto de puntos interconectados por una serie de líneas, los cuales reciben los nombres de nodos (node) y aristas (arcs) respectivamente. El concepto de grafo es fundamental en el ARS, puesto que la teoría de los grafos aporta importantes términos y expresiones adecuados para el estudio y descripción de las redes y sus particularidades.

Cuando un grafo representa una red social, los puntos representan a los diferentes actores sociales, los cuales son los miembros de la red diferenciados entre sí (individuos, organizaciones, usuarios de un servicio de salud, miembros de una comunidad).

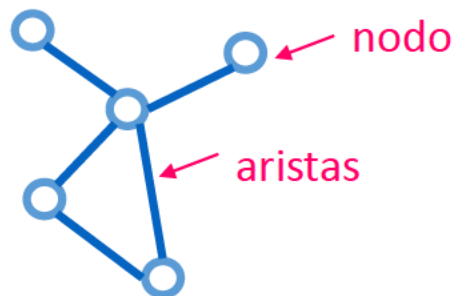


Figura 1. Ejemplos de un grafo donde se aprecian los nodos y las aristas

Las aristas de los grafos entre tanto, representan las relaciones sociales y han recibido diversos nombres como ties, links o bonds los cuales son bastante ambiguos para la traducción al español y en general se ha tendido a usar el término vínculo. Un vínculo es el lazo que existen entre dos o más nodos y como ya dijimos se representan por líneas (aristas). En resumen, cuando un grafo representa una red social, los puntos constituyen a diferentes actores sociales y las aristas, las relaciones establecidas entre ellos las cuales pueden ser expresadas matemáticamente.

Un actor puede tener múltiples lazos con otros actores, la cual es resultante de la combinación de otras relaciones; puede ser orientada o recíproca pero suele ser ponderada. Los lazos pueden ser débiles y son aquellos en los cuales existen relaciones especializadas entre dos actores sociales. Igualmente, los lazos pueden ser fuertes y contrario a los primeros indican relaciones sociales cercanas y solidarias. Otro concepto asociado es el de Camino (path), el cual consiste en una secuencia de nodos y relaciones en la cual cada nodo sólo puede ser usado una vez. Así mismo, se pueden distinguir diferentes tipos de vínculos:

Vínculo, conexión o relación orientada: relación que parte de un nodo hacia otro. Se presenta con una flecha apuntado al nodo receptor.

Vínculo, conexión o relación ponderada: relación calificada con un valor ordinal o de rango. Se opone a la relación binaria (presencia o ausencia) y permite gradaciones.

Vínculo, conexión o relación recíproca: relación idéntica para cada uno de los dos nodos. Suele representarse con una línea sin flechas.

Relación multiplexada o múltiple: relación resultante de la combinación de otras relaciones. Puede ser orientada o recíproca pero suele ser ponderada.

Otro de los conceptos asociados a la dinámica de los actores o nodos es la cohesión (cohesion), que se refiere a la interrelación de los actores en una red. Existen tres medidas comunes de cohesión:

Distancia (Distance): aquella existente entre dos actores en una red (o nodos en un grafo), cuyo cálculo se realiza sumando el número de vínculos (líneas) que existen a lo largo de la ruta más corta entre ellos.

Accesibilidad (Reachability): medida que determina si los actores dentro de una red están relacionados, directa o indirectamente, a todos los demás actores. Aquellos actores que no están conectados a los demás actores se les denominan Aislamientos.

Densidad (Density): la densidad es una de las más elementales medidas en el análisis de redes y uno de los datos más comúnmente utilizados en las nociones de epidemiología social. Se refiere al número total de vínculos relacionales dividido por el total de número posible de vínculos relacionales. Con la densidad se valora la fuerza de los vínculos, su utilidad radica en conocer la velocidad con la cual se difunde información entre los actores de la red, además, es una medida que permite conocer el capital social y el nivel de coacción social entre los actores de una determinada red.

Modelos de una red

Las redes sociales, se pueden clasificar de acuerdo a diversos criterios como el tamaño u otras propiedades, es incluso muy probable que surjan nuevas formas de clasificación con los avances que se vienen gestando en el estudio de las redes. Algunos autores, consideran que la existencia de cambios pequeños en la topología de una red puede generar una amplia posibilidad en la complejidad de relaciones, así dichos cambios afecten a un número pequeño de actores, lo cual demuestra precisamente la importancia de contar con modelos de definición y comprensión de las propiedades de las redes y la posibilidad de construir nuevos sistemas de clasificación a partir de dichos modelos. Existen en general tres modelos de redes con amplia difusión.

Redes aleatorias

Este modelo de red consiste en la generación de redes a partir de la agregación de nodos (actores) aleatoriamente a un conjunto de datos fijos. De acuerdo con este modelo, aunque los enlaces de la red se hayan distribuido al azar el sistema que resulta de la distribución es plenamente democrático, lo que conduce a que la distribución de relaciones o enlaces entre los actores de la red sean muy similar a la media.

Redes de mundo pequeño

Este tipo de redes se caracterizan por poseer un diámetro (camino más corto en los actores) pequeño y a su vez contar con un elevado nivel de agrupamiento, en otras palabras, se trata de redes con características peculiares debido a que cuentan con una conectividad particular a partir de la cual la distancia entre dos actores cualquiera es pequeña en comparación al tamaño total de la red. Uno de los primeros experimentos desarrollados frente al fenómeno lo constituyó el estudio de Milgram (1963), quien evaluó la distancia promedio que existía entre personas que constituían redes de contacto. Para su estudio empleó personas que residían entre Nebraska y Kansas en los Estados Unidos, las cuales debían contactarse con otras personas de Massachusetts señalando si conocían a estas últimas; en los casos en que no les conocieran debían reportar a un contacto que les pudieran conocer. El resultado del estudio condujo a Milgram a identificar que el número de conocidos de una persona crece de forma exponencial con el número de enlaces de una red personal, hallando una distancia promedio de seis personas, es decir, que todos estamos conectados por medio de un máximo de seis contactos intermedios. Este resultado es conocido popularmente como seis grados de separación, teoría que incluso ha servido de nombre para sitios de redes en la Internet. Las redes de mundo pequeño han sido confirmadas en distintas disciplinas caracterizando incluso a redes de mundo real como mundos pequeños.

Redes libres de escala

En este tipo de redes se identifica que algún nodo con alto grado tiende a ser conectado a otro nodo que también cuenta con alto grado, lo cual indica que los enlaces de la red se concentran en un número reducido de actores. De acuerdo con esto se afirma que en el mundo real existen redes para las cuales no es definible un patrón que determine el grado de conexión de sus actores, sino que la forma como se distribuyen este tipo de redes implica una especie de dominio de un número pequeño de actores. En las redes libres de escala la distribución entre los enlaces es más favorable en comparación a las redes aleatorias, dado que existen más nodos con pocos enlaces en lugar de muchos enlaces para pocos nodos como sucede en la red aleatoria.

El modelo de redes libres de escala demuestra avances frente a las redes aleatorias y de mundos pequeños debido a que estas últimas presentan condiciones que no se cumplen en las redes del mundo real. Inicialmente está el hecho que en las redes aleatorias se parte de la idea que un gran número de los actores se unen por azar, sin embargo, en el mundo real las redes tienden a ser mucho más abiertas y a crecer con el tiempo por la adición de nuevos miembros; en segundo lugar, la idea del patrón aleatorio de relaciones implica uniformidad en las conexiones pero realmente los actores de las redes muestran preferencias al momento de seleccionar a los individuos con quienes interactuar. Para el caso de las redes libres de escala el crecimiento de las mismas, parte de adición de nuevos nodos hacia los cuales hay un comportamiento anexo preferencial basado en la inclusión de nodos con un elevado número de enlaces. Algunos ejemplos de este tipo de redes en el mundo real lo constituyen las redes de celulares, las redes de correos electrónicos, entre otras.

La posición de los nodos

Un conjunto de datos de elevado valor dentro del análisis de redes sociales consiste en el cálculo de las medidas de centralidad y poder, las cuales se basan en la identificación de posiciones que ofrecen ventajas dentro de la red de relaciones. Desde esta idea, los individuos con mejores

posiciones en la red tiene mayor posibilidad de ejercer poder y de ser centrales (mayor nivel de importancia social). En las relaciones sociales el poder es una propiedad fundamental y desde el enfoque de las redes se le considera como una característica inherentemente relacional en la medida que el mismo es una consecuencia de los patrones de relaciones.

Los sujetos con mayor centralidad y poder social gozan de mejores posiciones sociales y cuentan con un número menor de restricciones para la integración con otros actores. Algunas medidas de centralidad y poder son las siguientes:

Centralidad y centralización.

La centralidad (centrality) es una de las ideas que orientaron los primeros estudios sobre redes y respecto a su medición existen diversas propuestas pero la más extendida y simple consiste en la centralidad del grado (degree), por la cual se considera que un grafo es central si tiene un grado alto, ello se corresponde con la idea de centralidad que asegura que un punto es central si está adecuadamente conectado con los demás puntos de su entorno. Así pues, el concepto de centralidad se asocia mayoritariamente a la cuestión de la centralidad relativa de los puntos de un grafo (lo que se conoce como centralidad de los puntos). La centralidad puede calcularse de acuerdo con diferentes medidas, que dan lugar a diferentes conceptos de centralidad, se han propuesto diversas nociones de este indicador, una de ellas, la de centralidad global, medida en términos de la cercanía (closeness) de cada punto respecto a los demás; así mismo, sobresale la medida basada en la idea de intermediación (betweenness), que determina en qué orden un punto hace de intermediario entre otros puntos por estar situado en el camino entre ellos. Con el fin de tener más claridad frente a estos conceptos se definirán de forma más amplia a continuación.

Centralidad de grado/Degree.

Se define al grado de un nodo como el número de enlaces que posee, es decir, el número de relaciones que tiene el nodo con otros nodos. Se dice que un nodo tiene un grupo relacional conformado por los nodos que están alrededor. Significa actividad o popularidad vista en muchos lazos establecidos con otros actores lo que aumenta el grado central. La centralidad de grado es un dato relacionado con aquellos actores que tienen más vínculos con otros actores, esto les puede hacer más privilegiados que otros debido a que tienen muchos lazos de interacción y diversas formas de satisfacer las necesidades que demandan en la red, y por lo tanto son menos dependientes de otras personas. Debido a las numerosas relaciones estos nodos pueden tener acceso a los recursos de la red, y a menudo son intermediarios y negociadores de los intercambios. En la Figura 1.2 se aprecia un grafo en el que el tamaño de los nodos expresa diferentes niveles de centralidad de grado.

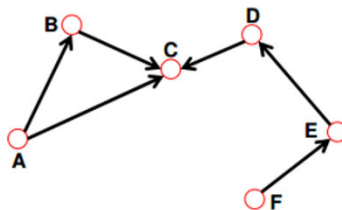


Figura 1.2. Grafo que representa la centralidad de grado de una red.

Un grafo el grado del nodo suele ser dividido en dos formas: a) el grado de entrada (in-degree), referido al total de lazos reportados por los demás nodos (actores) hacia un determinado sujeto; y b) grado de salida (out-degree), que implica los lazos referenciados por un determinado sujeto con los demás actores de la red.

Cercanía/closeness.

La cercanía se fundamenta en la definición de distancia. Determina el nivel de proximidad o distancia que tiene un actor con los demás actores de la red. Si un actor está cerca de todos los demás en la red, la distancia es de no más de uno, entonces él o ella no depende de ningún otro para llegar a todos en la red. Se considera que la cercanía o centralidad de cercanía mide los pasos que se necesitan para acceder a cada otro vértice desde un vértice dado. Al igual que con el grado, la cercanía también puede ser de entrada y salida.

Cercanía de entrada/InCloseness.

Es la percepción que tienen los individuos dentro de la red de los procesos y condiciones gregarias que se presentan, es decir, que tanta relación hay entre los actores de dicha red.

Cercanía de salida/OutCloseness.

Hace referencia a las percepciones que tienen los individuos dentro de la red, acerca de las relaciones sostenidas con los demás actores, es decir, define que tan cercano percibe a los demás, cada actor de la red.

Intermediación/betweenness.

La intermediación representa la ubicación de los actores en relación a que tanto un punto actúa de intermediario con otro en la red. Existen dos tipos de centralidad de intermediación, la primera está basada en la frecuencia en la que un nodo aparece en camino entre dos nodos, es decir, las veces en que se presenta entre un nodo con trayectoria mínima, y la segunda está basada en la importancia que tiene una arista con respecto a una trayectoria mínima, es decir, las veces en que un enlace se presenta en medio de una trayectoria mínima. Se trata, entonces, de una medida de las posibilidades de control pues identifica cómo un actor que tiene alto contenido de intermediario regula el flujo de los contenidos y recursos que se conectan entre uno y otro actor. Existen redes que por su tendencia al cierre y la facilidad de los contactos entre actores no cuentan con indicador de intermediación, en este caso todos los actores cuentan con posibilidades de acceder a todos los contactos lo que hace que la medida de intermediación sea igual a cero. La figura 3.3 representa este tipo de redes.

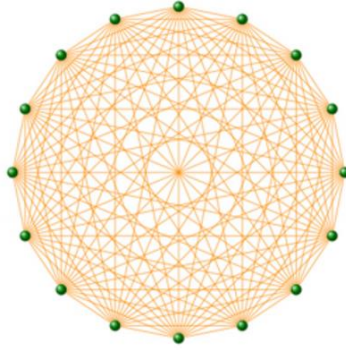


Figura 1.3. Red de intermediación cero (0) en la que no se identifican subgrupos.

De igual forma, la centralidad no solo se ocupa de los puntos centrales del grafo de una red sino que también convoca la participación de los puntos periféricos, de modo que se distingue igualmente puntos centralmente periféricos y globalmente periféricos.

Comunidades en una red y subgrupos

Por su parte, la centralidad es una medida que permite identificar hasta qué punto el grafo es o no una figura centralizada. Igualmente, existen medidas que permiten identificar si un grafo está organizado o no en relación a sus puntos más centrales; cuando sobre un conjunto de puntos recae la organización total del grafo se le denomina centro estructural (structural centre). Del mismo modo, es posible que la centralización sea aún más extendida y que el grafo se organice de acuerdo a un solo punto y no a un conjunto de ellos, en ese caso se denomina centro absoluto (absolute centre).

Como puede notarse, las medidas descritas permiten la identificación de limitaciones y oportunidades de los individuos dentro de las redes de socialización, las medidas de integración, participación y conexión social así como la manera en que las mismas pueden definir la posición social de los individuos y sus posibilidades de acceder a recursos sociales de soporte y desarrollo.

El análisis de las redes permite conocer la forma como los actores están integrados a la red facilitando el reconocimiento de las interacciones locales entre actores, dándole una especial importancia a la discusión de la dualidad individuo-estructura, de hecho, uno de los problemas más persistentes en el análisis de las redes sociales ha sido el descubrir las diversas *cliques* y comunidades con una cierta entidad propia en los que se puede dividir una red, es decir, dentro de una red es posible identificar sub-estructuras conformadas por la unión de individuos o actores cuya interacción diferencial frente al resto de la red se basa en la similaridad y los lazos preferentes.

Frente a las redes en contextos comunitarios y aún en el creciente campo de estudio de las redes virtuales, el desarrollo de comunidades o la identificación de subgrafos, — como se le denomina técnicamente en ARS— representa una de las sobresalientes modalidades de análisis, en cuanto facilita la detección de comunidades que se construyen en la red con base a un determinado criterio de afinidad. Normalmente, a la hora de formar y analizar subgrupos, de lo que se trata es de agrupar a los vértices en torno a alguna categoría que pueda resultar significativa a la hora de distinguir distintas pautas en la formación de la red. Sin embargo, los análisis basados en la formación de

cliqués y similares adoptan un punto de vista muy diferente con respecto al estudio de comunidades.

Cliqué.

Matemáticamente se define un cliqué como el conjunto de puntos en el que todos los pares de puntos son conectados por lo menos por una arista, es decir, se basa en el cálculo de relaciones conformadas por más de dos actores en las cuales todos los individuos se conectan entre sí. Esta técnica es la más frecuentemente utilizada para identificar la densidad de los subgrupos dentro de una red. En la teoría de grafos un *cliqué* es un subgrafo en el cual cada vértice está conectado a otro vértice del grafo, es decir, el subgrafo puede ser considerado como un grafo completo.

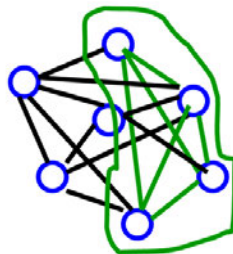


Figura 1.4. Grafo con cliqués.

En la *Cliqué*, el subgrafo que se forma, puede dividirse del grafo dado que su constitución se determina por un conjunto de actores que de manera natural se relacionan mutuamente, se basa por tanto en la identificación de la relación de todos con todos como criterio de definición de subagrupaciones.

Una medida más flexible de identificación de subgrupos la representan los Clanes o Clúster, a través de los cuales se aprovecha toda la información relacional de la red mediante un análisis de pasos que integra subconjuntos de actores acorde al nivel de similitud de los mismos. El clúster se basa en la similitud y equivalencia de las relaciones entre los actores que van formando clúster en diferentes pasos y paulatinamente se van adicionando a diferentes clúster hasta la conformación de uno que integra a todos los actores de la red. Evidentemente, se trata de un método de adición de actores con base a la detección de similitudes entre los mismos; su utilidad es amplia, en especial porque facilita la detección de secuencias de agrupamiento en donde es posible observar la manera cómo los diferentes actores se unen a subagrupaciones en la medida que comparten características vinculanticas que son atractivas para los integrantes de cada subgrupo.

Otras medidas que merecen consideración en el estudio de subagrupaciones en la red son las siguientes:

Componentes: es una parte de la red en la que todos los actores están conectados, directa o indirectamente, al menos con una relación. Se puede considerar también como el máximo subgrafo posible en cual se cumple que cada par de puntos se conecta por un camino posible.

Círculos: surgen en el curso de la interrelación y pueden no ser visibles para quienes pertenecen a ellos. La cohesión de un círculo social no radica en el nivel de contacto de sus actores sino en la existencia de cadenas de contactos que los ligan unos con otros.

Posiciones: Las posiciones sociales se refieren a aquellos lugares que ocupan los actores sociales y dentro del análisis de redes se han estudiado mediante el desarrollo del concepto de equivalencia estructural (structural equivalence), el cual identifica que dos actores son estructuralmente equivalentes cuando se conectan con otros actores de manera idéntica lo que los hace sustituibles el uno por el otro. Cuando dos actores estructuralmente equivalentes se relacionan con otros que también son equivalentes entre sí, se denomina equivalencia regular. La medición de las posiciones de acuerdo a la equivalencia se suele realizar mediante el método de modelado de bloques, el cual es una estructura simplificada que puede representar toda la red, es decir, se establece una serie de particiones de la red las cuales serían dichas estructuras simplificadas.

La visualización de las redes

La capacidad de representar datos visualmente con el propósito de generar comunicaciones de cierto tipo de información es conocida como visualización y su desarrollo se sustenta en la aplicación de metodologías basadas en el cálculo matemático de información significativa de conjuntos complejos de datos, por medio de la generación de gráficos o imágenes interactivas.

En la última década se ha presentado un considerable avance en el desarrollo de técnicas que involucran la tecnología para la visualización de información, lo cual ha permitido la interacción directa con la información visualizada gracias al desarrollo de diversas interfaces dispuestas para éste fin. La visualización de redes es una técnica de exploración de las propiedades de la estructura social cuyo valor no solo es meramente ilustrativo sino que permite incluso, bajo el cumplimiento de determinados principios metodológicos, la formulación de explicaciones a los fenómenos de interacción que dentro de las redes se desarrollan.

La calidad de las visualizaciones generadas sobre las redes puede favorecer a que se identifiquen conocimientos de elevado valor sobre las redes estudiadas conservando incluso propiedades explicativas. Las visualizaciones constituyen una importante herramienta que facilita la detección de propiedades esenciales de las redes como las características de los nodos y los lazos vinculares, así como las estructuras generales y los resultados sociales. De igual forma, se ha identificado que las visualizaciones o grafos de redes cuentan con bondades explicativas especialmente de las posiciones estructurales de los nodos.

La visualización ofrece aportes importantes para el estudio de las redes sociales en cuanto le permite al investigador la realización del tamizaje de los datos, la identificación de patrones de datos interesantes y la posibilidad de darle sentido a las observaciones formuladas sobre las redes.

Recopilación de datos

La recopilación de datos en el ARS implican retos particulares dado que trabaja con datos relacionales y no detiene su interés sobre datos que impliquen atributos como opiniones, actitudes o variables sobre hechos. Los datos relacionales expresan relaciones, vínculos, conexiones, comunicaciones entre grupos, a partir de los agentes que los componen. Estos actores o agentes se relacionan entre sí y los datos surgidos del análisis se refieren a la información obtenida acerca de dichas relaciones. La información que se obtiene (datos relacionales) son analizados tanto en sus componentes formales como en sus contenidos o aspectos sustantivos, es decir, se analizan conjuntamente el contenido y la forma de la red.

La peculiaridad de los datos relacionales dificulta el desarrollo de procesos tradicionales de muestreo para poder obtener la información deseada sobre la red, por tal razón, los procesos de selección de los actores a evaluar y la recogida de los datos de sus relaciones requiere de procedimientos diferentes. Algunos autores han sugerido que la forma de muestrear los participantes de los estudios de redes consiste en la identificación de redes parciales —de salud, religiosas, políticas— para encontrar en los mismos actores que puedan ser seleccionados mediante métodos tradicionales de la investigación por encuestas.

Otros autores consideran que el muestreo está definido por los límites establecidos para la red. De esta idea se desprenden dos enfoques, uno realista, por el que se considera que en las investigaciones se debe incluir a los actores de una red que se consideren conscientemente miembros de la misma dado que de por sí ellos tienen ese nivel de conciencia de pertenencia a una red particular.

Un segundo enfoque, el nominalista, es por su parte contrario al primero y asegura que es el investigador quien debe definir los límites de una red pues los actores no son siempre conscientes de su pertenencia a la misma. Cualquiera que fuera el procedimiento para establecer el muestreo del ARS, la delimitación que dicho muestreo hace de la red influye en los instrumentos y el proceso de recogida de los datos como tal, si bien esencialmente los datos relacionales se recogen por observación, cuestionarios, encuestas o métodos etnográficos que posibilitan procesar los datos mediante los procedimientos formales de análisis de redes sociales, estos procedimientos consisten esencialmente en la teoría de los grafos y la teoría matricial.

El análisis de la red permite evaluar las redes sociales desde dos aproximaciones: la relacional y la posicional, y según las cuales, la concepción del actor en cada una de ellas es diferente. Esencialmente, se incluyen tres categorías:

Análisis centrado en un solo actor: los análisis centrados en un actor desde la perspectiva relacional se tratan de un tipo de análisis que ha dado lugar a la evaluación de redes egocéntricas por medio de las cuales se analizan las relaciones que poseen los actores dentro de la red. Por su parte, los análisis posicionales centrados en un solo actor se han interesado más por la posición del actor en la red y por incluir las relaciones que posee tanto como las que no posee en la red.

Análisis centrado en subgrupos: los análisis de localización de subgrupos desde el enfoque relacional trabajan esencialmente con el concepto de cliqué (actores conectados uno con otros mediante lazos fuertes). Por su parte, aquellos estudios sobre subgrupos de enfoque posicional se basan en la identificación de dichos subgrupos en términos de equivalencia estructural.

Análisis centrado en las relaciones de actores/subgrupos con la red completa: desde el enfoque relacional se desarrollan estudios basados en la densidad y transitividad de la red. La perspectiva posicional por su parte se interesa en las pautas de relación que unen los actores en diferentes posiciones en la red lo que facilita identificar el grado de centralización (si todas las relaciones pasan por un actor central en la red) y jerarquización de las relaciones.

Limitaciones del análisis de redes sociales

En el análisis de redes sociales se reconoce la existencia unas limitaciones precisas asociadas a los alcances y aplicabilidad de los resultados obtenidos. Por una parte, se incluyen dentro de estas

condiciones la tendencia a analizar flujos de datos mediante mediciones transversales, lo cual implica la identificación de información estática que impide determinar de forma confiable el sistema relacional con base a los cambios que continuamente se presentan dentro de los sistemas sociales. En redes de gran tamaño y que además cuentan con una dinámica sorprendente de cambios y transformaciones relacionales como ocurre con las redes de la Web, resulta de mucha dificultad lograr el completo desarrollo de sistemas que permitan representar y comprender cantidades elevadas de información.

A pesar del creciente avance de diversas áreas interesadas en el análisis de las redes, aún resulta complejo lograr el desarrollo de la representación textual de los datos de redes con éste nivel de dinamismo, por lo que con frecuencia los sistemas de análisis incluyen entre sus debilidades la pérdida potencial de la información que ha sido almacenada. Otro de los aspectos críticos que limitan los resultados obtenidos con el empleo del análisis de redes sociales consiste en la dificultad para la producción de respuestas casuales a los fenómenos sociales, debido a que no es posible lograr que dichas explicaciones estén desprovistas de errores o limitaciones que reducen el nivel de aplicabilidad de los datos.

Esta limitación surge como resultado del sinnúmero de variables ambientales, personales y sociales (y de interacciones entre las mismas) que pueden conducir a la existencia de niveles amplios de varianza estructural. A pesar de lo anterior, existe un consenso general acerca de las bondades y los alcances del análisis de redes sociales (ARS), al considerarlo como un método que proporciona una aproximación conceptual amplia para identificar las estructuras sociales que emergen de las distintas formas de relación, pero a su vez es un conjunto específico de métodos y técnicas.

El ARS se ha venido desarrollado como herramienta de medición y análisis de las estructuras sociales que surgen de las relaciones entre actores sociales diversos (individuos, organizaciones, naciones, etc.), mostrando aplicación y utilidad en múltiples escenarios que incluyen desde las redes de tipo biológico hasta análisis semánticos, de cooperación científica y comunicaciones vía Web. Se pueden distinguir algunos elementos significativos de la aplicación del análisis de redes sociales lo cuales presentamos a manera de resumen de las ganancias asociadas al empleo de dicho método:

- a. Favorece el estudio de relaciones entre individuos y grupos así como la posibilidad de estudiar las relaciones hacia el interior de las redes, para comprender los procesos de la estructura interna de relaciones.
- b. Al basar el análisis en las interrelaciones reconoce la interdependencia entre los actores y sus implicaciones para la integración social. El abordaje de las interdependencias permite analizar todos los flujos de relaciones sin que se den sesgos por algún tipo de interacción.
- c. Permite la construcción de los sistemas sociales de interacción desde diferentes niveles que incluyen los sistemas microsociales y los marcosistemas, entendiendo la sociedad como una red de redes, desde la cual analiza los distintos niveles de integración social en una misma sociedad superando los límites de la relación dicotómica entre las partes y el todo.
- d. Además de los aspectos relacionados con la interrelación entre actores, el análisis de redes se encarga del estudio del contenido de tales relaciones.
- e. La comprensión del contenido relacional le permite al análisis de redes reconocer las diferencias relacionadas con el acceso a recursos sociales así como las variaciones en el nivel

de poder social entre actores de una unidad social; desde este conocimiento el análisis de redes puede aportar a la subsanación de las asimetrías en las relaciones sociales.

- f. El reconocimiento de dichas asimetrías sensibiliza a los investigadores al respecto de las dinámicas existentes en las redes.
- g. El análisis de redes supera barreras conceptuales e institucionales en materia del abordaje de fenómenos sociales complejos, yendo más allá de la formulación de análisis discretos o fragmentarios.
- h. Producto de su interés y capacidad de explicación de la relación, la interdependencia y la interacción, el análisis de redes puede estudiar otras formas de organización social que emergen de la interacción como las relaciones patróncliente o las coaliciones entre líderes y seguidores, es decir, permite reconocer que existen formas emergentes de relación social que permean diversos acontecimientos vitales importantes.
- i. El análisis de redes sociales introduce a las personas como actoras de los procesos de interacción social superando la noción de roles o papeles como categorías de análisis sociológico. Así mismo, reconoce la importancia de las acciones y decisiones de las personas en la forma de funcionamiento y futuro del desempeño de la red.
- j. El análisis de redes permite incluir nuevas categorías de sujetos dentro del análisis sociológico al evaluar la existencia de amigos de los amigos, es decir, incluye como unidades de análisis a individuos con los que no se tienen contacto directo en calidad de informantes.

Referencias

Ávila-Toscano, J. H., Gutiérrez, B. & Pérez, J. (2010). Indicadores Estructurales y Conglomerados de Actores en la Red Social de una Subcultura Urbana. *Revista colombiana de psicología*, 20 (2), 193-207. Barabási, A. L. (2002). *Linked: the new science of networks*. Cambridge: Perseus.

Barabási, A. L. & Bonabeau, E. (2003). Redes sin escala. *Investigación y Ciencia*, 322, 58-67.

Boissevain, J. (1979). Network Analysis: A Reappraisal. *Current Anthropology*, 20 (2), 392-394.

Brandes, U., Raab, J. & Wagner, D. (2001). Exploratory Network Visualization: Simultaneous Display of Actor Status and Connections, *Journal of Social Structure*, 2 (4). Disponible en: <http://www.library.cmu.edu:7850/JoSS/brandes/index.html>.

Brandes, U., Kenis, P. & Raab, J. (2005). La explicación a través de la visualización de redes. *REDES-Revista Hispana para el Análisis de Redes Sociales*, 9 (6). Disponible en: [http:// revista-redes.rediris.es](http://revista-redes.rediris.es)

Burt, R. S. (1980). Models of network structure. *Annual Review of Sociology*, 6, 79-141. Dürsteler, J. (2005). Visualización del contenido de la web. Disponible en: <http://www.infovis.net/printMag.php?num=175&lang=1>

Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, 1, 215-239.

Granovetter, M. (2003). La fuerza de los lazos débiles. Revisión de la teoría reticular. En F. Requena (Ed.), *Análisis de redes sociales. Orígenes, teoría y práctica*. Colección monografía, 198, Centro de estudios sociológicos (pp. 196-230). Madrid: Siglo XXI.

Hanneman, R. & Riddle, M. (2005). Introduction to social network methods. Recuperado de <http://www.faculty.ucr.edu/~hanneman/nettext/>.

Hawe, P., Webster, C. & Shiell, A. (2004). A glossary of terms for navigating the field of social network analysis. *Journal of Epidemiology Community Health*, 58, 971-975.

Heer, J., Card, S. & Landay, J. (2005). Prefuse: A toolkit for interactive information visualization. *ACM Human Factors in Computing Systems (CHI)*, 421-430.

Herrero, R. (2000). La terminología del análisis de redes. *Problemas de definición y traducción. Política y Sociedad*, 33, 199-206.

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, 7 (1), 100-107.

Lozares, C. (1996). La teoría de redes sociales. *Papers*, 48, 103-126.

Lozares, C. (2005). Bases socio-metodológicas para el Análisis de Redes Sociales, *ARS. Empiria. Revista de Metodología de Ciencias Sociales*, 10, 9-35.

Luján, M., Martig, S. & Castro, S. (2008). Aplicación de Visualización de Grafos utilizando Servicios Web. *X Workshop de Investigadores en Ciencias de la Computación —WICC—* (pp. 301-305). 5 y 6 de Mayo. La Pampa, Argentina: Universidad Nacional de La Pampa. Disponible en: http://www.cs.uns.edu.ar/~mlg/archivos/Publicaciones/WICC_2008/Wicc08_Aplicaci%3n%20de%20Vis_Grafos_ServiciosWeb.pdf

Mejía, C. (2010). Análisis de Redes Sociales a Gran Escala. (Tesis de Maestría). Centro de investigación y de estudios avanzados del Instituto Politécnico Nacional: México.

Medrano, J. F., Alonso, J. L. & Figuerola, C. G. (2010). Visualización de Grafos Web. Disponible en: http://www.visinfo.com.ar/articulos/Visualizacion_de_Grafos_Web_Articulo.pdf.

Miceli, J. (2006). La ciencia de las redes. Reseña de: Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age*. Random House: London, UK. *REDES- Revista Hispana para el Análisis de Redes Sociales*, 10 (10).

Milgram, S. (1963). Behavioral study of obedience. *Journal of Abnormal and Social Psychology*, 67 (4), 371-378.

Monsalve, M. (2008). Análisis de redes sociales: un tutorial. Universidad de Chile. Departamento de ciencias de la computación. Disponible en: <http://www.dcc.uchile.cl/~mnmonsal/SNA.pdf>

Palacio, J. & Madariaga, C. (2006a). *Psicología Social Aplicada y Análisis de Redes Sociales (ARS)*. En *Psicología social: Teoría y Práctica*. Barranquilla: Ediciones Uninorte.

Perianes-Rodríguez, A., Olmeda-Gómez, C. & de Moya-Anegón, F. (2008). Introducción al análisis de redes. *El Profesional de la Información*, 17 (6), 664-669.

Rheingans, P. & Landreth, C. (1995). Perceptual principles for effective visualizations. En: G. Grinstein & H. Levkowitz (Eds). *Perceptual Issues in Visualization*. (pp. 59-74). Springer-Verlag.

Sanz Menéndez, J. (2003). Análisis de redes sociales: o cómo representar las estructuras sociales subyacentes. *Apuntes de Ciencia y Tecnología*, 7, 20-29.

Trujillo, H., Mañas, F. & González-Cabrera, J. (2010). Evaluación de la potencia explicativa de los grafos de redes sociales clandestinas con Ucinet y Net-Draw. *Universitas Psychologica*, 9 (1), 67-78.

Tufte, E. R. (1997). *Visual Explanations. Images and Quantities, Evidence and Narrative*. Cheshire/Connecticut: Graphics Press

Verd, J. M. & Martí, J. (1999). Muestreo y recogida de datos en el análisis de redes sociales. *Questiio*, 23 (3), 507-524.

Wellman, B. (1999). *El análisis estructural: del método y la metáfora a la teoría y la sustancia*. Universidad de Toronto.

Anexo 4

Reporte Técnico

Visualización de datos

Erick López Ornelas

Universidad Autónoma Metropolitana – Cuajimalpa

En la actualidad nuestro consumo de información se ha multiplicado de manera exponencial debido a dos factores: cada vez se produce más información (redes sociales, dispositivos, etc.) y cada vez tenemos más capacidad de acceso a dicha información, especialmente a través de Internet y de la Web. La capacidad de sacar partido y entender la información bruta está íntimamente ligada a nuestra capacidad para explotarla y transformarla en algo más que puro dato: los datos adquieren significado.

No obstante, los datos, entendidos como registros aislados, no aportan un significado concreto. Sólo cuando nos acercamos a ellos y le aplicamos una interpretación, estos cobran sentido y se transforman en conocimiento. En el ámbito de la tecnología, la explotación de datos ha evolucionado en las últimas décadas para diseñar mecanismos de interpretación cada vez más robustos y asequibles. Y, entre estos mecanismos de explotación, el más importante es la visualización de datos.

El uso de datos

La visualización de datos es la presentación gráfica de información con dos propósitos. Por un lado, la interpretación y construcción de significado a partir de los datos (es decir, el análisis); y por otro lado, la comunicación.

La visualización es una herramienta muy potente para descubrir y comprender la lógica que se encuentra detrás de un conjunto de datos, así como para compartir esta interpretación con otras personas desde un punto de vista objetivo. Como se suele decir: “una imagen vale más que mil palabras”, especialmente cuando el significado que se quiere comunicar se representa mejor gráfica que verbalmente, y esta representación gráfica está diseñada de acuerdo a los principios formales que rigen la visualización de datos.

Podemos estar durante horas ante una tabla de números y no visualizar lo que es inmediatamente obvio cuando esta información se presenta gráficamente. Lo que los números no pueden comunicar cuando son presentados en una tabla, se vuelve visible e inteligible cuando se comunica visualmente. Este es el “poder” de la visualización de datos.

Es importante tener en cuenta que, aunque la visualización de datos se utiliza para representar generalmente variables cuantitativas y relaciones entre ellas, también puede utilizarse para representar relaciones entre entidades de naturaleza cualitativa. Por ejemplo, las relaciones entre

las personas de una determinada red social, que pueden ser además “tipadas” de acuerdo a la naturaleza de esta relación: amistad, familia, trabajo, etc. Estas visualizaciones de representación de entidades y propiedades de tipo relacional se sustentan sobre la tipología de la estructura que se quiere representar y utilizan gráficos basados en grafos (nodos y arcos).

Hoy en día, existen multitud de herramientas para la generación de gráficos y mapas de maneras asistida, que permiten construir valor sobre los datos. Aunque la visualización puede seguir siendo un proceso artesanal basado en el diseño y la elaboración manual, como son algunos ejemplos de infografías, en la mayor parte de aplicaciones se utiliza software específico para la construcción de estas representaciones gráficas.

Tipos de visualización

Existen multitud de técnicas y aproximaciones para la visualización según sea la naturaleza del dato de la información. Desde el punto de vista de los datos, especialmente datos estructurados (o semi-estructurados) y su explotación visual, podemos establecer grosso modo la siguiente clasificación de tipos de visualización según complejidad y elaboración de la información.

1. Elementos básicos de representación de datos.

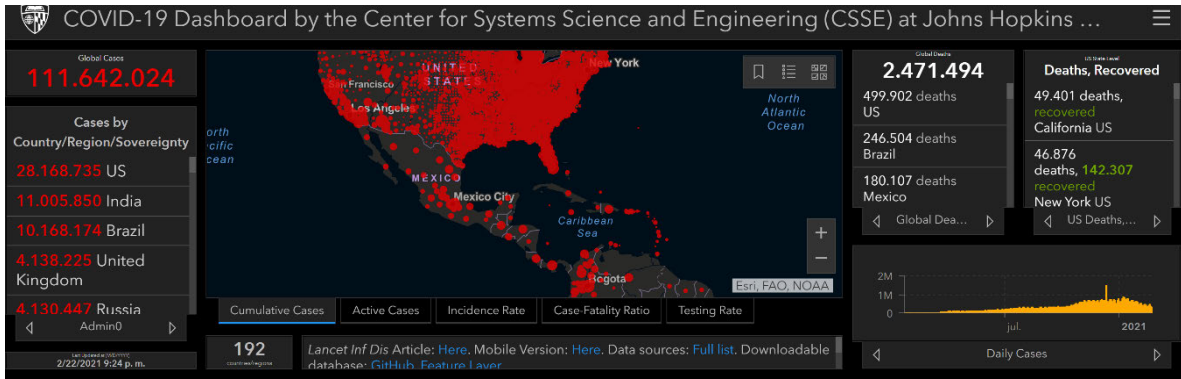
Es el caso más sencillo. Un elemento básico de visualización puede ser un gráfico, un mapa, un KPI, tablas de datos, un grafo etc. A continuación se señalan algunos tipos de visualizaciones básicas.

- Gráficas: barras, líneas, columnas, puntos, “tree maps”, pie, semi-pie, etc.
- Mapas: burbujas, coropletas (o mapa temático), mapa de calor.
- Tablas: con anidación, dinámicas, de drill-down, de transiciones, etc.



2. Panel de control.

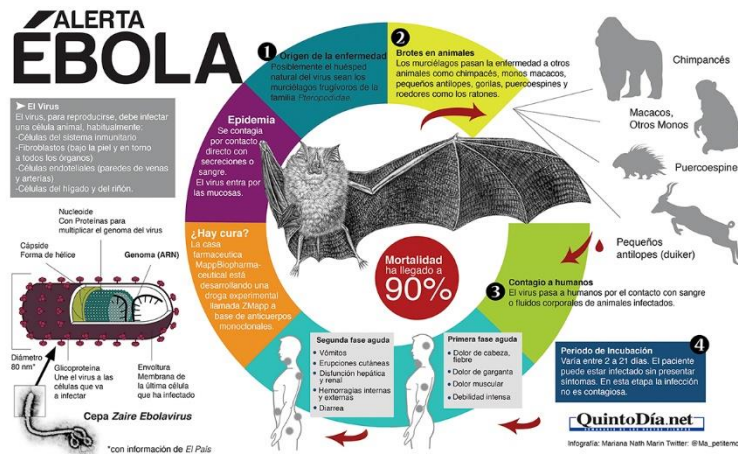
Un panel de control es una composición compleja de visualizaciones individuales que guardan una coherencia y una relación temática entre ellas. Son ampliamente utilizados en las organizaciones para análisis de conjuntos de variables y toma de decisiones.



3. Infografías

Una infografía también es una composición elaborada de visualizaciones que construye, a partir de diferentes elementos, un significado complejo para el usuario. Sin embargo, las infografías no están tan destinadas al análisis de variables sino a la construcción de narrativas a partir de los datos; es decir, las infografías se utilizan para contar “historias”. Esta narrativa no se construye a través de texto, sino mediante la disposición de la información en la que las visualizaciones se combinan con otros elementos como: símbolos, leyendas, dibujos, imágenes sintéticas, etc.

Una infografía tiene un componente de elaboración manual y de diseño en su construcción muy elevado.



Los datos abiertos

El movimiento de apertura de datos dentro de las administraciones se ha focalizado en los últimos años en los aspectos de su publicación. Por un lado, se han definido políticas y normativas para empujar a las administraciones a abrir sus datos a la sociedad, tanto por compromisos de transparencia como en la búsqueda de un retorno en la reutilización de la información a través de otras organizaciones y de desarrollo de un tejido productivo a su alrededor.

Por otro lado, se han definido formatos de publicación y buenas prácticas para que las administraciones supiesen no sólo qué publicar, sino cómo y cuándo, de manera que los procesos de búsqueda, intercambio y acceso de la información obedezcan a principios de diseño estándares que faciliten la automatización de la reutilización de los datos abiertos.

En este sentido, podemos destacar la catalogación de los diferentes niveles de publicación de datos y la definición de buenas prácticas en cuanto a reutilización.

Sin embargo, una de las críticas más habituales en torno al open data es que el esfuerzo actual se concentra en la publicación de datos y no en su usabilidad; es decir, en cómo esos datos se van a consumir por el usuario final. Muchos proyectos open data se han concentrado excesivamente en estas cuestiones técnicas: formatos, end-points, etc., ajenas al usuario potencial, y no han prestado atención en cómo esos datos pueden utilizarse y se les puede extraer valor. Esto ha provocado que muchos portales se hayan convertido en meros repositorios de datos, con un tráfico marginal y con un impacto social muy bajo.

Así, es fundamental entender que es natural que los primeros pasos en el movimiento de datos abiertos se hayan concentrado precisamente en la definición de qué es un dato abierto y cuál es su procedimiento de publicación. Ahora mismo, sin embargo, tanto la tecnología disponible y la madurez del movimiento como la demanda ciudadana, obligan a dar un paso más allá y evolucionar el concepto open data a una dimensión más amplia en la que no sólo se tengan en cuenta cuestiones referentes a la publicación y la reutilización desde el punto de vista técnico.

La siguiente tarea es proporcionar herramientas para que los usuarios puedan consumir y explotar los datos de manera autónoma, haciendo realidad el objetivo inicial de que la información de las administraciones tenga un impacto positivo y real en nuestra sociedad. En este punto, la visualización, como hemos comentado antes, es la herramienta más potente para acercar los datos a cualquier usuario.

Actualmente la visualización forma parte de cualquier proceso estándar de gestión y análisis de datos en el mundo empresarial, especialmente cuando hablamos de cuestiones relacionadas con el análisis estadístico y la “Inteligencia de negocio”, aunque no de manera exclusiva. En cualquier caso, podemos definir estos procesos como el conjunto de técnicas y herramientas para la adquisición, tratamiento y transformación de datos brutos en información y conocimiento útiles para un propósito de análisis de determinado.

En el ámbito de la empresa, el propósito suele estar ligado a procesos y aspectos de negocio, pero la aplicación de las técnicas y herramientas es genérico respecto de la intención del análisis. En estos casos, la visualización forma parte del ciclo de vida de los datos, en los que después de su preparación y transformación, la información se explota por parte del usuario final a través de representaciones gráficas que le permiten interpretar el sentido de los datos.

Herramientas de visualización de datos

A lo largo de esta sección, se describen los estándares web que se han ido desarrollando en los últimos años para la evolución de las aplicaciones web, base fundamental para creación de visualizaciones web basadas en datos. A su vez, también se describen diferentes bibliotecas

JavaScript que utilizan estos estándares para la creación de las partes fundamentales de la visualización.

Estándares web

En los últimos años, el desarrollo de contenidos y aplicaciones web ha sufrido una revolución motivada por los nuevos estándares web. Desde que en 2014 el W3C publicó la versión definitiva del estándar HTML5 se han ido sumando nuevos estándares que permiten a los desarrolladores crear, no solo visualizaciones estáticas, sino potentes aplicaciones de visualización que incluyen un alto grado de dinamismo y personalización estética. Estándares como HTML5 o la nueva versión del estándar de hojas de estilo en cascada (CSS3) junto con la evolución de los navegadores modernos permiten a los desarrolladores crear aplicaciones web multidispositivo sin tener que preocuparse, como solía ocurrir, de qué navegador o dispositivo se use para ver la visualización.

HTML5 y Canvas

El W3C publicaba en 2014 la versión 5 del estándar más importante para la web: HTML5, terminando así con 6 años de desarrollo en colaboración con las principales empresas tecnológicas. HTML es el lenguaje básico sobre el que se construyen todos los contenidos. Esta nueva versión supuso un salto de calidad en este estándar incluyendo nuevas API para la interacción con el navegador y el dispositivo, y lo más importante, obligando a todos los navegadores web a actualizarse para soportar esta nueva versión.

Esto trajo como consecuencia que las compañías responsables de cada uno de los navegadores modernos estuviesen obligadas a mejorar la compatibilidad de sus productos con esta nueva versión. Se facilitó la homogeneización de los contenidos web y se terminó, al menos parcialmente, con los “hacks” de código necesarios para visualizar los contenidos correctamente en cada uno de ellos.

Dentro de esta nueva versión HTML5 se definió un elemento importante para la visualización de contenidos, el Canvas. Este elemento HTML es usado para dibujar gráficos, normalmente 2D aunque también puede ser utilizado junto con WebGL para visualizar gráficos 3D, usando scripts, normalmente escritos en JavaScript. Canvas puede ser utilizado para dibujar gráficos, realizar composición de fotos y animaciones. Existen múltiples bibliotecas JavaScript que facilitan la creación de gráficos para este elemento.

CSS3

Las hojas de estilo en cascada o CSS (Cascading Style Sheets) es un lenguaje que permite definir la presentación de un documento estructurado como HTML o CSS. Este lenguaje es un estándar del consorcio W3C que permite diferenciar el contenido de las páginas web de la presentación de este contenido. Tras muchos años de desarrollo, en 2011 se publicó la versión 3 de este lenguaje que supone un gran avance en cuanto a potencia y funcionalidad.

Debido a su definición modular, no todos los elementos del lenguaje CSS3 tienen el mismo nivel de madurez tecnológica. Respecto a la visualización de datos, los módulos más relevantes son los siguientes:

Fondos y colores: Añadida la posibilidad de añadir múltiples fondos de pantalla, así como nuevos elementos que facilitan controlar mucho mejor los tamaños y posiciones de estos.

Animaciones: Evolución de las animaciones que se pueden describir para los diferentes elementos del lenguaje.

Posicionamiento: Este módulo es una evolución del posicionamiento clásico que permite distribuir los elementos de forma sencilla y mucho más flexible.

SVG

Los Gráficos Vectoriales Redimensionables o SVG son una especificación estándar para describir gráficos vectoriales bidimensionales, tanto estáticos como animados (estos últimos con ayuda de SMIL), en formato XML. El W3C lo convirtió en recomendación en 2001 pasando a ser uno de los estándares más utilizados a la hora de crear gráficos en 2D. Este estándar nos permite definir conceptos básicos como punto, línea, polígono para después combinarlos y formar el gráfico vectorial complejo. Resulta especialmente útil para la creación de componentes de gráficos estadísticos que permiten la visualización y el análisis de series de datos.

Actualmente la versión 2 de SVG está en desarrollo y promete añadir un amplio conjunto de características que faciliten y amplíen el uso de las nuevas características de los navegadores y estándares web modernos.

WebGL

WebGL es un estándar que nace de la mano de Mozilla pero que actualmente incluye a muchas de las grandes empresas tecnológicas. Este estándar permite definir gráficos y animaciones 3D haciendo uso del componente de HTML5, Canvas. WebGL está basado en OpenGL ES 2.0 y proporciona una API para gráficos 3D. Actualmente está ampliamente soportado por los navegadores pasando a ser el lenguaje más utilizado para las visualizaciones web en 3D.

Como WebGL es una tecnología diseñada para trabajar directamente con el procesador gráfico del dispositivo resulta bastante difícil de codificar en comparación con otros estándares web más accesibles. Para resolver esto han ido surgiendo múltiples bibliotecas JavaScript que facilitan el uso de esta especificación.

Bibliotecas y plataformas

Existen múltiples bibliotecas que facilitan la creación de gráficos 2D para visualizaciones de datos. Dentro de este grupo podemos diferenciar las que están basadas en Canvas, y las que están basadas en SVG.

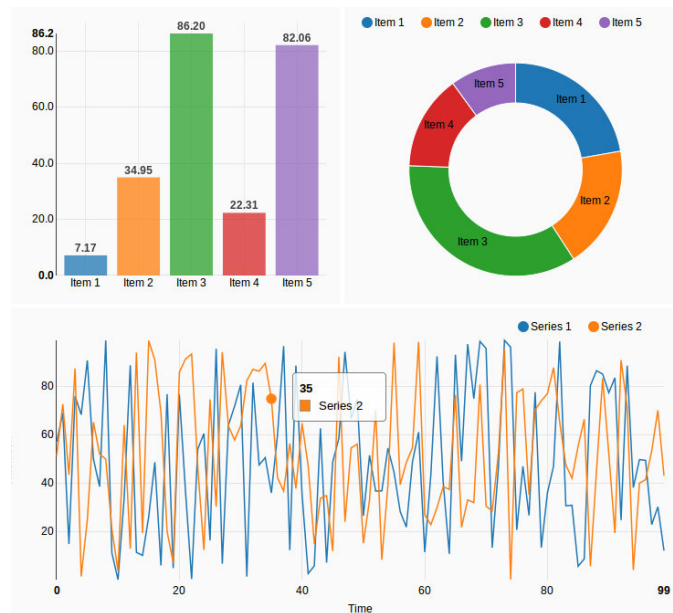
D3.js

D3.js es una biblioteca JavaScript para la manipulación de documentos basados en datos. D3 ayuda a dar vida a los datos usando HTML, SVG y CSS. El punto clave de D3 es la utilización de estos estándares web para que el editor/desarrollador pueda explotar todas las nuevas capacidades de los navegadores web modernos, facilitando la manipulación del DOM para la creación de poderosas visualizaciones basadas en datos.

D3 nos da una capa de abstracción por encima de SVG ofreciendo una serie de utilidades para la manipulación de datos y su posterior visualización. De este modo, D3, nos ofrece la creación de gráficos interactivos que pueden ser personalizados utilizando estándares con CSS3.

Por todo esto, D3 se ha convertido en una de las bibliotecas más utilizadas no solo para la creación de visualizaciones sino como base de un gran conjunto de bibliotecas más específicas basadas en esta biblioteca.

Algunos ejemplos son n3-charts, NVD311, etc. Este tipo de herramientas minimiza el principal problema que encontramos en D3, la genericidad. Al tratarse de una librería genérica para gráficos 2D, resulta complicada su utilización. Por ello las bibliotecas como NVD3 nos ofrecen clases Javascript para los principales tipos de gráficas, como por ejemplo: grafico de barras, de líneas, de área, de burbujas, etc., facilitando la creación de este tipo de visualizaciones.



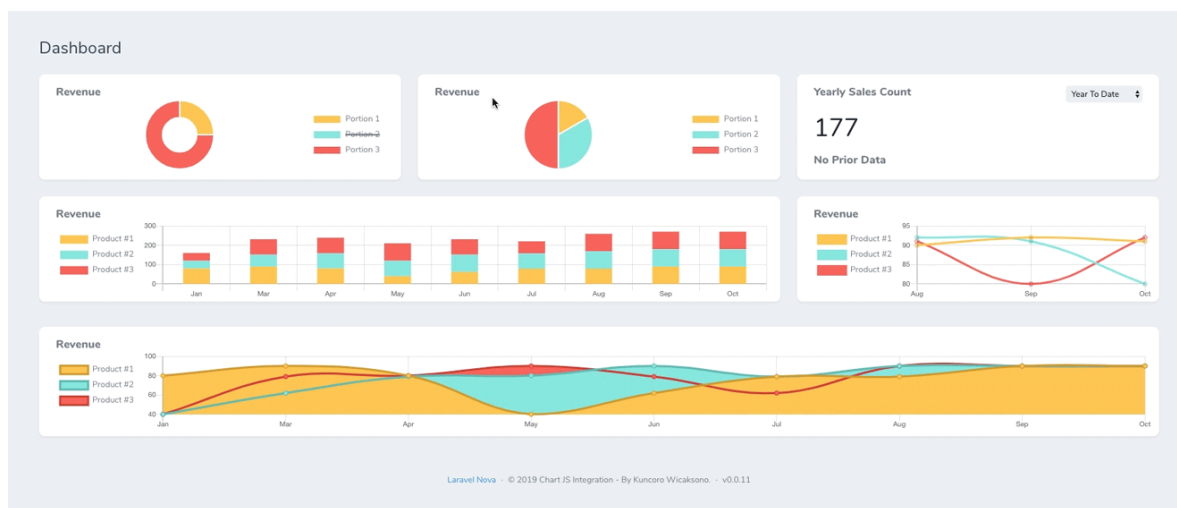
Google Charts

Google ha apostado fuerte por la visualización de datos y ha creado su propia biblioteca para la creación de visualizaciones. Esta biblioteca también está basada en SVG y dispone de un amplio número de tipos de gráficos y mapas. Dispone también de lo que han denominado datos dinámicos, que nos permiten hacer consultas SQL sobre datos almacenados en Google Spreadsheets, Google Fusion Tables o Salesforce y visualizarlos. Google ofrece también, aunque de forma limitada, algunos componentes para creación de dashboards interactivos.



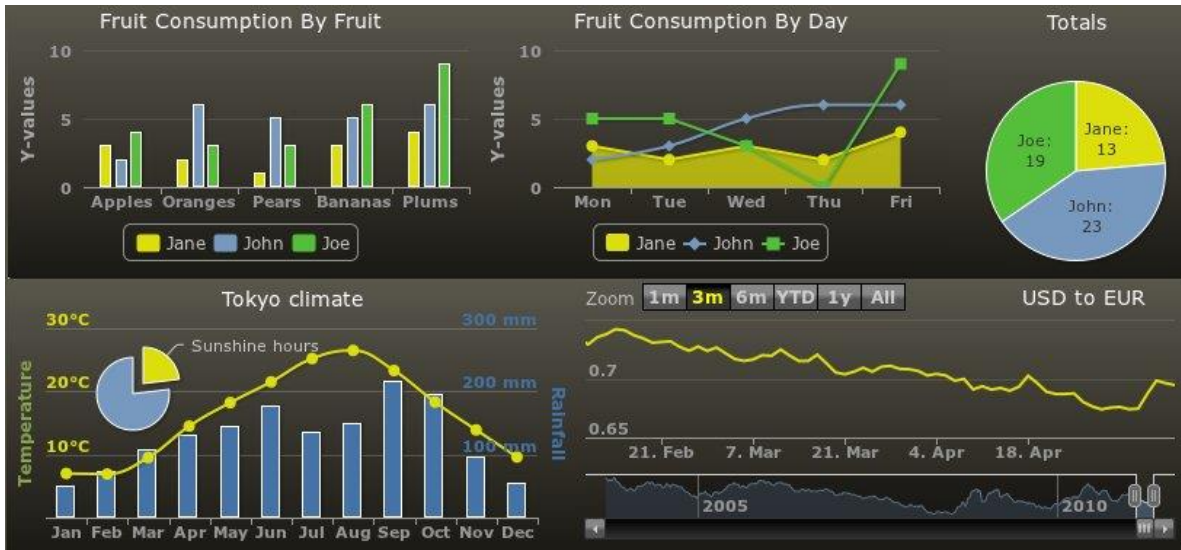
Chart.JS

Chart.JS es el primero de los ejemplos incluidos que utiliza Canvas en lugar de SVG. Este tipo de bibliotecas nos ofrecen clases JavaScript para dibujar el gráfico dentro del nuevo elemento de HTML5 Canvas. El uso de Canvas en lugar de SVG limita la interoperabilidad de los gráficos creados con estándares como CSS3 pero tiene la ventaja de que necesitan menos recursos computacionales para su para que el navegador web los represente.



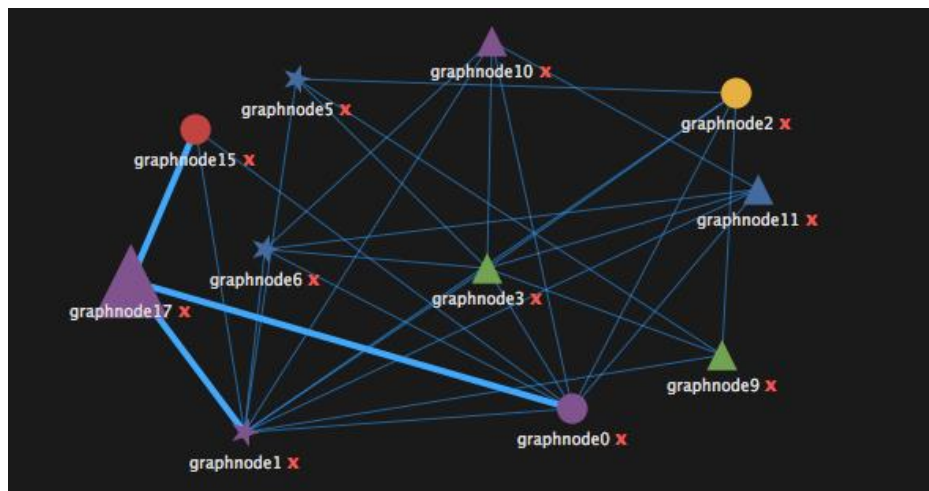
HighchartsJS

HighchartsJS es una biblioteca comercial, gratuita para uso no comercial, que cuenta un gran número de seguidores. Esta biblioteca también utiliza el estándar SVG para sus gráficos. Como principal punto a su favor, dispone de un gran número de componentes y una documentación muy extensa y cuidada. La principal desventaja es la licencia, que hay que adquirir si se quiere usar para fines comerciales.



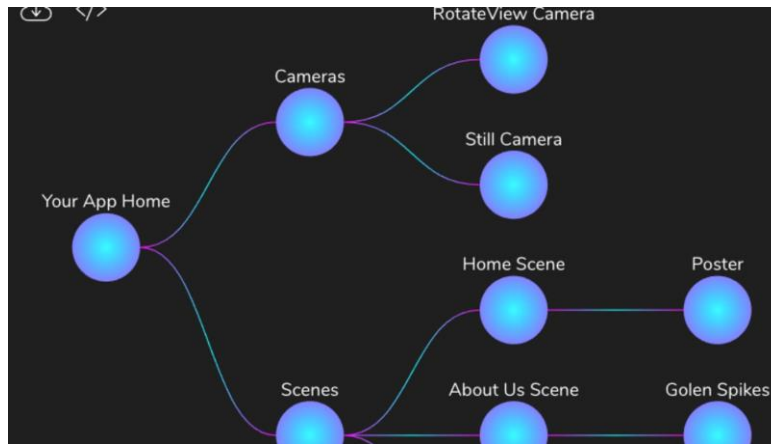
JavaScript InfoVis Toolkit

Ejemplo de biblioteca de gráficos que utiliza estándar WebGL. Esta biblioteca está desarrollada por Nicolás García Belmonte bajo copyright de Sencha Company y ofrece la incorporación de ciertos aspectos 3D a los tipos de gráficos más utilizados. Este tipo de características permite crear mapas interactivos ofreciendo espectaculares animaciones 3D.



Three.js

Three.js es una librería para construir escenas 3D con WebGL. Podemos verlo con un D3 para WebGL, es decir, una biblioteca que elimina la complejidad de escribir una visualización en WebGL. Ofrece todas las clases necesarias para aspectos comunes de las escenas 3D como las luces, cámaras, materiales y objetos, etc.



BabylonJS

Framework para la creación de juegos 3D con HTML5 y WebGL que también puede ser utilizado para crear espectaculares visualizaciones. Como ocurre con Three.js ofrece las clases necesarias de Cámara, Luz y Figura, de modo que nos abstraer de toda la complejidad de WebGL.



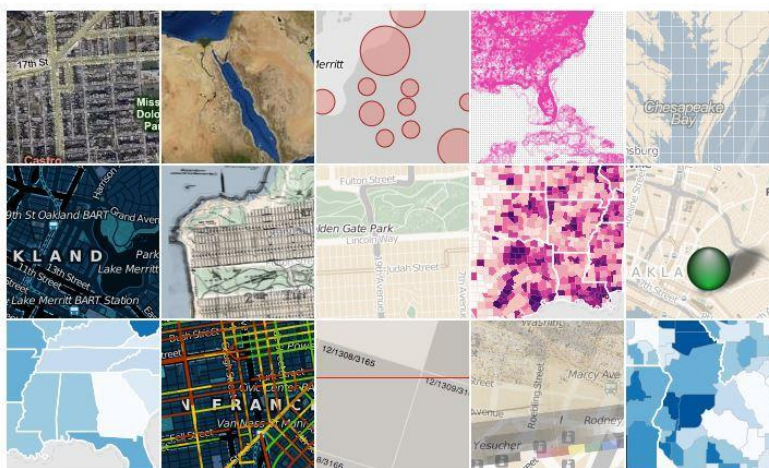
Con el auge de los servicios de mapas como Google Maps, OpenStreetMap y los sistemas GIS (Sistemas de Información Geográfica), han ido apareciendo múltiples bibliotecas para la visualización y el análisis de datos georreferenciados. Muchos de estas bibliotecas y clientes se basan en los estándares definidos por el Open Geospatial Consortium (OGC)¹⁸, como los servicios WMS (Web Map Service) o WFS (Web Feature Service).

Cabe destacar el uso del vocabulario GeoJSON como estándar de facto para la descripción de capas y objetos georreferenciados en el entorno web. Este vocabulario define los conceptos básicos de punto, ruta y polígono y permitiendo su combinación para la descripción de capas de visualización.

Polymaps.js

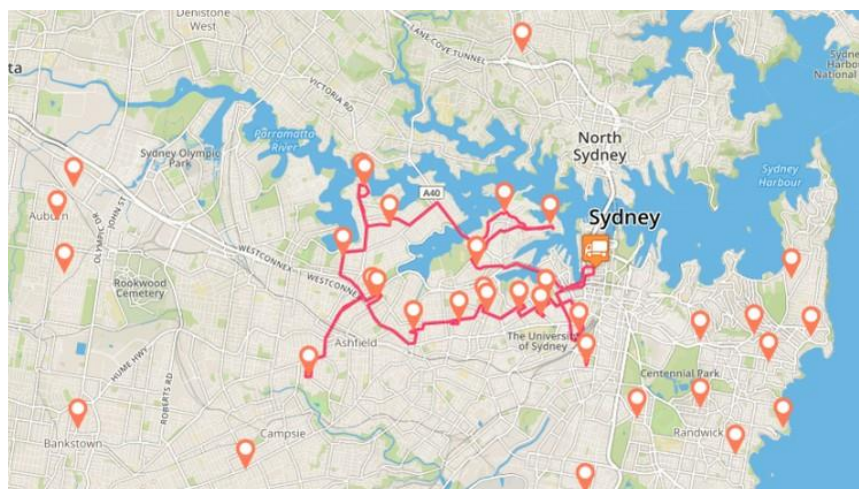
Polymaps es una biblioteca JavaScript para la creación de mapas dinámicos e interactivos para la web. Proporciona mecanismos para visualizar capas vectoriales de datos sobre las cartografías más utilizadas, como OpenStreetMap, CloudMade, Bing, etc.

Polymaps utiliza el estándar SVG para la creación de estas capas a partir de, por ejemplo, datos definidos con GeoJSON, permitiendo también su posterior personalización a través de la definición de reglas de visualización escritas en CSS.



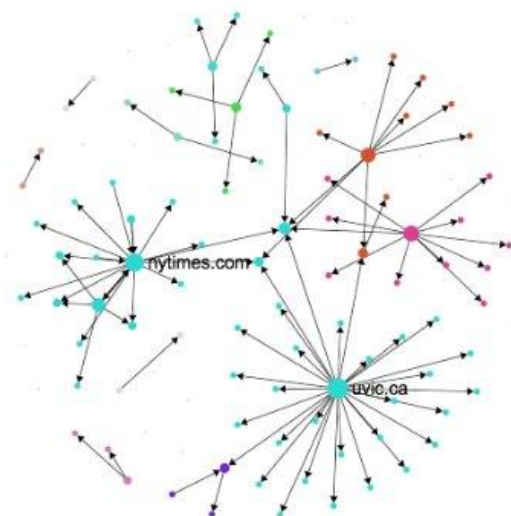
Leaflet

Leaflet es una de las bibliotecas de visualización y creación de mapas interactivos más usadas en la actualidad. Permite la creación de capas animadas a partir de múltiples formatos de datos, entre los que se encuentran GeoJSON ó WMS. Además, ofrece un gran conjunto de clases JavaScript para la creación y modificación de dichas capas así como para la visualización e interacción con ellas.



Sigma.js

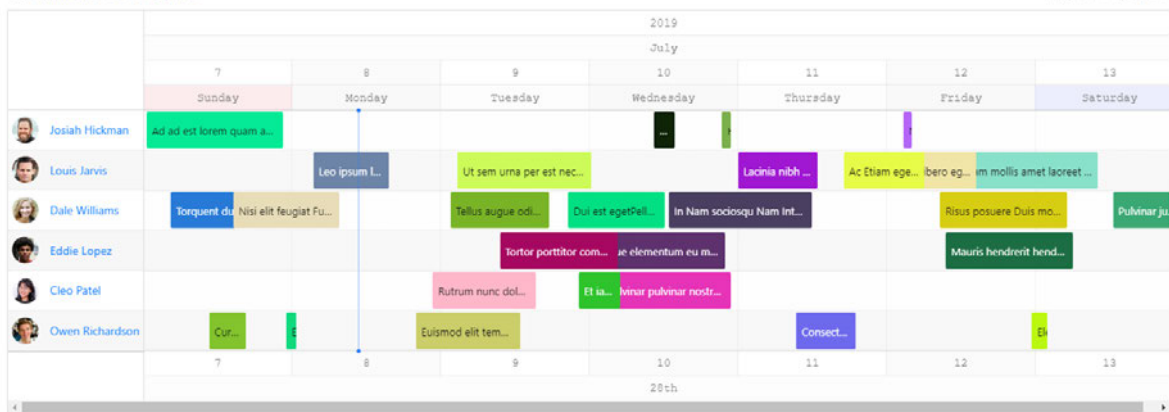
Sigma.js es una biblioteca JavaScript para la visualización de grafos para su utilización en entornos web. Esta biblioteca utiliza el elemento HTML5 Canvas para la visualización y ofrece todo un conjunto de utilidades para la visualización y el análisis de los elementos de grafo.



TimelineJS

Uno de los componentes recurrentes en las visualizaciones son las líneas de tiempo, es decir, en la localización de un conjunto de evento dentro de una línea temporal. TimelineJS ofrece de manera sencilla la creación de un componente de línea de tiempo que nos permite navegar entre los eventos representados.

Members Schedule



© MAGIC METHODS 2019

7/7/2019 - 7/14/2019

Conclusiones

Una vez se ha avanzado de manera significativa en la publicación de datos abiertos: legislación, formatos, buenas prácticas, definición de vocabularios estándar, etc., el siguiente paso es facilitar la

usabilidad y el acceso a los datos por parte de cualquier usuario. La visualización de datos es uno de los mecanismos más potentes para explotar y analizar de manera autónoma el significado implícito en los datos, independientemente del grado de conocimiento tecnológico del usuario. La visualización nos permite construir significado sobre los datos y la creación de narrativas basadas en la representación gráfica.

La visualización de datos, desde un punto de vista formal y científico, es una disciplina madura. Además, su grado de implantación en la industria es muy amplio, desde el ámbito industrial y de proceso, hasta aspectos más ligados al reporting financiero y la gestión de negocio. Su aplicación es transversal a cualquier proceso en el que existan datos, como es el caso del movimiento open data.

Esto supone una gran ventaja desde el punto de vista de introducción de herramientas de explotación de datos basados en visualización. Se conocen perfectamente los principios metodológicos y de usabilidad. Además existe actualmente una gran variedad de software disponible para el desarrollo de proyectos de visualización dentro de los portales open data.

En este reporte hemos centrado nuestro análisis en el bloque de Tecnología de visualización. Frameworks y bibliotecas de programación que permiten la construcción de aplicaciones y servicios basados en datos. El informe se centra exclusivamente en tecnología web de última generación y especialmente en aquella que se construye sobre estándares, como es el caso de HTML5 (Canvas), SVG y WebGL.

Las tecnologías y técnicas relacionadas con la visualización de datos siguen evolucionando de manera significativa. Este informe es un estado del arte que refleja qué alternativas más relevantes se encuentran en el mercado para trabajar en esta disciplina. Desde el punto de vista de la innovación y líneas futuras, existen varios caminos que se están siguiendo tanto en la industria del software como desde la parte más académica, entre las que destacan:

Visualización y grandes volúmenes de información

Un problema relacionado con la visualización es cuando se trabaja con grandes volúmenes de datos. Este problema no existe sólo a nivel de back-end, donde está más vinculado al análisis y tratamiento de grandes volúmenes de información estructurada y no estructurada (ie., Big Data), si no con la capacidad del navegador para soportar grandes conjuntos de datos (por ejemplo, cientos de miles de registros). Las restricciones de computación de los clientes en los que se ejecuta la visualización pueden ser una barrera para la construcción de ciertas visualizaciones. Este es un campo de investigación en el que se está trabajando actualmente.

Visualizaciones generadas por los usuarios.

Muchas herramientas no son accesibles más que para perfiles técnicos, ya sean a nivel informático o de tratamiento y análisis de datos. Esta barrera de la tecnología es uno de los grandes desafíos actuales: conseguir que el usuario que entiende la información tenga la autonomía suficiente para construir a través de asistentes intuitivos (WYIWYG) sus propias visualizaciones y explotar la información de manera dinámica.

Visualizaciones 3D

En los últimos años han aparecido, como se ha identificado en el informe, tecnología de aceleración de gráficos 3D para el entorno web. La aplicación de esta tecnología en la visualización de datos todavía está en un estado incipiente. La mayor parte de las herramientas siguen trabajando con perspectiva 2D. Es de esperar que según la tecnología alcance un estado de madurez más sólido, aparezcan nuevas técnicas de visualización integradas con las herramientas de usuario.

Visualización de datos y realidad aumentada

Un ámbito todavía poco explorado en la actualidad, pero que seguro supondrá una ruptura en la forma en cómo se consumen los datos, es la tecnología de realidad aumentada, que a través de dispositivos combina información real y física en un entorno interactivo para el usuario. Actualmente el estado del arte de la visualización se concentra en un consumo clásico y multidispositivo de los datos: ya sean aplicaciones web o aplicaciones móviles de propósito específico. La tecnología de realidad aumentada, con sus posibilidades de representación 3D y en tiempo real, es un territorio desconocido en este momento sobre sus potenciales aplicaciones e impacto en la disciplina de la visualización. Sin embargo, abre un vasto espacio para la imaginación y para diseñar nuevas soluciones en un entorno hasta hace poco “futurista”.

Referencias

Bach, B., Shi, C., Heulot, N., Madhyastha, T., Grabowski, T., Dragicevic, P.: Time curves: folding time to visualize patterns of temporal evolution in data. *IEEE Trans. Vis. Comput. Graph.* 22(1), 559–568 (2016)

Chen, W., Guo, F., Wang, F.-Y.: A survey of traffic data visualization. *IEEE Trans. Intell. Transport. Syst.* 16(6), 2970–2984 (2015)

Ferreira, N., Poco, J., Vo, H.T., Freire, J., Silva, C.T.: Visual exploration of big spatio-temporal urban data: a study of New York city taxi trips. *IEEE Trans. Vis. Comput. Graph.* 19(12), 2149–2158 (2013)

Liu, S., Cui, W., Wu, Y., Liu, M.: A survey on information visualization: recent advances and challenges. *Visual Comput.* 30(12), 1373–1393 (2014)

Olshannikova, E., Ometov, A., Koucheryavy, Y., Olsson, T.: Visualizing big data with augmented and virtual reality: challenges and research agenda. *J. Big Data* 2(1), 22 (2015)

Tsai, C.W., Lai, C.-F., Chao, H.-C., Vasilokos, A.V.: Big data analytics: a survey. *J. Big Data* 2(1), 21 (2015)

Zhou, H., Xu, P., Yuan, X., Qu, H.: Edge bundling in information visualization. *Tsinghua Sci. Technol.* 18(2), 145–156 (2013)

Zhu, Y.: Introducing Google chart tools and Google maps API in data visualization courses. *IEEE Comput. Graph. Appl.* 32(6), 6–9 (2012)

Anexo 5

Author detection: analyzing tweets by using a Naïve Bayes classifier

Abstract. In the context of digital social media, where users have multiple ways to obtain information, it is important to have tools to detect the authorship within a corpus supposedly created by a single author. With the tremendous amount of information coming from social networks there is a lot of research concerning author profiling, but there is a lack of research about the authorship identification. In order to detect the author of a group of tweets, a Naïve Bayes classifier is proposed which is an automatic algorithm based on Bayes' theorem. The main objective is to determine if a particular tweet was made by a specific user or not, based on its content. The data used correspond to a simple data set, obtained with the Twitter API, composed of four political accounts accompanied by their username and tweet identifier as it is mixed with multiple user tweets. To describe the performance of the classification model and interpret the obtained results, a confusion matrix is used as it contains values like accuracy, sensitivity, specificity, Kappa measure, the positive predictive and negative predictive value. These results show that the prediction model, after several cases of use, have acceptable values against the observed probabilities.

Keywords: Naïve Bayes classifier, authorship detection, social network analysis, Twitter, confusion matrix.

1. Introduction

Social media has changed the way people communicate and receive information about what happens daily. In this way, research areas like computer science evolve to have all the necessary methodological tools in order to understand social, political and economical news as quick as they are required by the modern society. Particularly, social media like Twitter has become the go-to place for latest developments [3].

It is undeniable that as a communication platform, Twitter has increasingly infused itself into daily life. So, one of the greatest tasks for academe is to understand why, how, when and with whom does the society communicates daily. In order to answer the last question, it is important to consider the tremendous amount of information, that occurs in social media, which is viable to be collected, in real time, directly from users; so, data can be analyzed by composing a corpus and applying linguistic research. However, it is difficult to identify the author of a post in order to hierarchize it

according to what or who the user wants to read first. Even though it is possible to know the authorship by looking at the username, however in public accounts, like politicians, the reality is that a group of people is behind these posts. It is important to identify in what cases the owner of the account is the same person that is publishing. This can be addressed like a classification problem where given a set of classes the objective is to seek to which of them a given post or tweet, in the case of social networks, belongs to.

Among the most popular machine learning algorithms, for implementing text classification models, the followings are found: Naïve Bayes algorithm's family, Support Vector Machines and deep learning. In this paper, an approach for the detection of the authorship, of a group of tweets, is based on the use of a Naïve Bayes classifier which is an automatic algorithm based on Bayes' theorem. The main objective is to determine if a particular tweet was made by a specific user or not, based on its content. The data used correspond to a simple data set, obtained with the Twitter API, which belong to four political accounts

accompanied by their username and tweet identifier as it is mixed with multiple user tweets. To describe the performance of the classification model and interpret the obtained results, a confusion matrix is used as it contains values like sensitivity, specificity, Cohen's Kappa measure, the positive predictive value and negative predictive value. The utility of using Naïve Bayes to classify authors according to the tweet is demonstrated as the classification model it is improved.

The rest of this paper is organized as following: section 2 gives a focused overview of related work. In section 3, the classification of tweets is explained by following four main steps. Section 4 is dedicated to the case of use while section 5 is devoted to the evaluation and interpretation of the obtained results. Finally, section 6 gives some conclusions about the work and opens several perspectives.

2. Related work

The authorship recognition or author identification is a subfield of Natural Language Processing (NLP) that uses machine learning techniques to be able to identify the author of a text based on characteristics such as vocabulary and frequency of terms. Generally, machine learning techniques have been used in the study of large texts. However, the short length of Twitter messages present a rarely examined challenge that have been also studied in a small amount of recent research. Short length creates new challenges like:

- consist of terse text, Twitter has a limitation of length;
- is an informal way of communication, which lacks syntactical forms of formal writing;
- content is inconsistent and topics may vary over time and
- contain highly unstructured data as a combination of weblinks, hashtags, emoticons, special symbols and images are present along with texts [6].

Concerning Twitter, there is a lot of research work in the analysis of tweets like hashtag recommendations, sentiment analysis, opinion mining and text classification. The classification of text, in this case tweets, is often done according to general themes or categories like health, education or politics. However, there is a lack of studies concerning the veritable identity of the user behind the tweet. Some studies have been oriented to the identification of anonymous users in Twitter, like in [1] where the objective is to uncover them using only linguistic stylometry.

In [8], the authors examine potential avenues of author identification, in tweets, by using supervised learning methods for data classification as Support Vector Machines (SVM). In this case, Bag-of-Words (BOW) and Style Marker feature sets were extracted and evaluated through a series of experiments where the Style Marker feature sets were found to be significantly more useful than BOW and are therefore suggested for potential applications in future research [8]. The same approach is followed in [6] where a BOW and style-based markers have been used for the identification of tweets authorship obtaining a recognition rate of 97% on a database of 70 Twitter users, which validates the superiority of using social interactive data compared to traditional linguistic profiles.

By using a set of stylistic markers, including personal and idiosyncratic edition options, an automatic authorship analysis was conducted in Twitter messages among three authors [9]. In this study, SVM classifiers were used demonstrating that they can be useful to attribute authorship in combination with a group of content-agnostic features. Some works in analyzing Twitter are focused in the identification of the gender and language of the author, as in [4, 5] where n-grams or BOW model are used. In [7] a variety of algorithms, coming from WEKA, were used as some machine learning classifiers were studied in order to compare their performance in the identification of troll profiles based on their published tweets. In this case, Random Forest, J48, K-Nearest Neighbor (KNN), Sequential Minimal Optimization (SMO) and Bayes Theorem-based algorithms were implemented resulting that SMO and decision trees were the most appropriated for this case of use.

Classification methods using traditional approaches like BOW have some limitations, so in [2], a multi-label classifier is used based on a small set of domain-specific features extracted from the author's profile and the text in order to classify authors into a predefined set of generic classes [2]. Other approaches to classify text concern the use of external information to add metadata to each of the tokens. This is the case of [11] that uses data repositories like Wikipedia to improve the performance of clustering algorithms. However, this approach is time consuming when the reality is that the analysis of social media has to be done in "real time".

In [10] a difference is made between the concepts of author attribution and author profiling. The first one concerns the identification of authorship of an anonymous piece of writing while the second one is attributed to the study of certain linguistic features that

vary according to the author. So, in [10] the authors are classified according to their characteristics.

After reviewing some recent studies in the area of text classification and authorship attribution, at this state of our study, no works are found focused on the identification of the author within a corpus that *a priori* comes from a single author and it is already tagged.

In this paper, an authorship recognition method in tweets by using Naïve Bayes algorithm is presented. The proposed method is validated by using a confusion matrix in 4 political public profiles mixed with more users. The next section describes each of the four steps of the process starting with the retrieval of tweets to compose a corpus.

3. Experimental Approach

A framework for experimentation, based on Naïve Bayes, was developed in order to meet the goal outlined in the introduction. With Naïve Bayes, any vector that represents a tweet will have to contain information about the probabilities of appearance of the words of the text within the tweets, of a given public profile, so that the algorithm can compute the likelihood of that tweet belonging to the author.

Our approach consists of the following four steps: 1) Data import, 2) Data processing, 3) Prediction model and 4) Evaluate Results. For all the process, R language (<https://www.r-project.org/>) was used because of the facilities given by the some of their packages and functions.

3.1. Data import

The data import concerns the recovery of the tweets by specifying a user profile. R, and its library *twitter*,

has been used in order to connect to the Twitter platform. By specifying keys provided by the Twitter developer page it is possible to access the Application Program Interface (API).

The main structure for managing tweets is by creating a corpus which is a set of tweets that contains similar characteristics. In this case, all the tweets that compose the corpus were retrieved the same day as 4 of the accounts belong to politicians.

One of the previous steps before storing the tweets is the preprocessing step which allows the elimination of data that is not of interest. In this case, all the URL were removed because they have been shortened and so they don't provide any significance.

3.2. Data processing.

To classify text using Naïve Bayes, the data need to be storage in a data frame which is used for storing data tables. It is a list of vectors of equal length. Also, this data must have a specific structure:

- each row must correspond to a specific tweet,
- each column must correspond to a specific word,
- each cell must indicate if a word appears in a specific tweet.

All the data, that is storage in the data frame, has to be converted into a sparse matrix so it is necessary to:

- segment each tweet by words,
- create one new column for each word,
- count how many times each word appears in each tweet,
- translate the matrix from a “high” format to a “wide” format, this means to have as many columns as words.

R provides some functions that help to convert the data frame (see Figure 1) into a sparse matrix (see Figure 2). These functions are *tidytext*, *dplyr* and *tidyr*.

```
# A tibble: 9,394 x 3
  id                screenName      text
  <chr>             <chr>          <chr>
1 x2897441          EPN            #MiBandera @TerminalNavega https://t.co/...
2 x166669818       VicenteFoxQue  "\"Avanzando hacia una crisis...."
3 "Vemos las gigantescas perdidas de... ""
4 Hacia mucho tiempo no pasaba esto!! ""
5 "Muchos frentes equivocados: ""
6 Queremos que acabe con la corrupci... " sin empleo y sin Pa<U+00E... ""
7 x82119937        lopezobrador_ <U+00A1>Miren cu<U+00E1>nta gente partici...
8 x166669818       VicenteFoxQue  Situaci<U+00F3>n que a<U+00FA>n no ocurre...
9 x82119937        lopezobrador_ "\"Inauguraci<U+00F3>n del estadio de b<U...
10 " desde Ciudad de M<U+00E9>xico. h... ""
# ... with 9,384 more rows
```

Fig. 1. Initial data frame composed by 9394 rows and 3 columns (*id*, *screenName* and *text*).

```

# A tibble: 863 x 11,011
  screenName id   `_deepakanand` `_iga` `_paolimeow` `_sigure` `_v` `_0` `_0.2` `_0.6` `_00` `_000`
  <chr>      <chr>      <int> <int>      <int>      <int> <int> <int> <int> <int> <int>
1 " @lopez... " cu...      NA    NA        NA        NA    NA    NA    NA    NA    NA
2 " arriba... " l...      NA    NA        NA        NA    NA    NA    NA    NA    NA
3 " deja d... L<U+...      NA    NA        NA        NA    NA    NA    NA    NA    NA
4 " DESHON... CORR...      NA    NA        NA        NA    NA    NA    NA    NA    NA
5 " especi... " an...      NA    NA        NA        NA    NA    NA    NA    NA    NA
6 " no es ... Gobe...      NA    NA        NA        NA    NA    NA    NA    NA    NA
7 " peores... " la...      NA    NA        NA        NA    NA    NA    NA    NA    NA
8 " podr<U... Si e...      NA    NA        NA        NA    NA    NA    NA    NA    NA
9 " todos ... Enti...      NA    NA        NA        NA    NA    NA    NA    NA    NA
10 " @CeciTe... " @C...      NA    NA        NA        NA    NA    NA    NA    NA    NA
# ... with 853 more rows, and 10,999 more variables: `0001f1e6` <int>, `0001f1e7` <int>, `0001f1e8` <int>,
# `0001f1ea` <int>, `0001f1f1` <int>, `0001f1f2` <int>, `0001f1f7` <int>, `0001f1f8` <int>,
# `0001f1fa` <int>, `0001f1fb` <int>, `0001f1fd` <int>, `0001f399` <int>, `0001f3ad` <int>,
# `0001f3eb` <int>, `0001f3fb` <int>, `0001f3fc` <int>, `0001f3fd` <int>, `0001f41f` <int>,
# `0001f444` <int>, `0001f447` <int>, `0001f449` <int>, `0001f44b` <int>, `0001f44c` <int>,
# `0001f44d` <int>, `0001f44f` <int>, `0001f468` <int>, `0001f469` <int>, `0001f49a` <int>,
# `0001f4a1` <int>, `0001f4a5` <int>, `0001f4ba` <int>, `0001f4c9` <int>, `0001f4e1` <int>,

```

Fig. 2. Sparse matrix converted from the initial data frame, with 863 rows (one for each tweet) and 11011 columns (one for each word).

3.3. Prediction model.

In machine learning, there is a big interest in selecting the best hypothesis given certain data. So, in a classification problem the hypothesis may be the class to assign for a new data instance. The Bayes' Theorem provides a way to calculate the probability of a hypothesis given prior knowledge. Naïve Bayes can be used for multi-class classification problems, however in this case the classification was started with a binary (two-class).

In the creation of a classification model a diagnosis is needed of how well it is doing its job. So, to achieve this purpose the data have to be divided into two sets, one of training and one of test. With the training set the model will be adjusted, in this case, by determining conditional probabilities of each word, for each category. Then, this model is applied in the test set to be able to analyze how many of the cases were correctly classified.

For the experimental approach, the data was divided, 70% in the training set and the rest in the test one. A random sample of the data is used. A list of probabilities is stored to a file for a learned Naïve Bayes model. This includes:

- class probabilities: the probabilities of each class in the training dataset, which are the frequency of words that belong to each author (class) divided by the total number of instances.
- conditional probabilities: the conditional probabilities of each input value given each class value. These probabilities are the frequency of each at-

tribute value for a given class value (profile author) divided by the frequency of instances (words) with that class value.

When using the Naïve Bayes function, with the train set, it asks for the data that is going to be used. Also, the objective variable is an argument that have to be given in order to proceed to the classification. In this case, see Figure 1, the objective or dependent variable corresponds to the *screenName*. All the other variables will be the predictors or independent variables. The model allows to make predictions by using the function *predict()* that is contained in the R library named *base*. For this the test set is used.

3.4. Evaluate results.

The evaluation of the results is done by using a confusion matrix which is a table that describes the performance of a classification model on a set of test data for which the true values are known.

The *caret* library for machine learning, in R, can calculate a confusion matrix with the function *confusionMatrix()*. This matrix allows to analyze the accuracy of the predictions and some adjustment measures. The metrics shown in the confusion matrix are:

- Accuracy: represents how often is the classifier correct, is the fraction of predictions the model got right.
- 95% CI: confidence interval with a degree of confidence of 95%.
- P-Value: one-sided test to see if the accuracy is better than the "no information rate," which is

taken to be the largest class percentage in the data.

- Kappa: measure of how much the model improves a prediction against the observed probabilities.
- McNemar's Test P-Value: captures the errors made by both models (training and test).
- Sensitivity: proportion of positive results out of the number of samples which were actually positive.
- Specificity: proportion of truly negative cases that were classified as negative; it is a measure of how well the classifier identifies negative cases.
- Positive Predictive Value: the percent of predicted positives that are actually positive.
- Negative Predictive Value: the percent of negative positives that are actually negative.
- Prevalence: the ratio of actual yes to total number of instances.
- Detection Rate: the rate of true events also predicted to be events.
- Detection Prevalence: the prevalence of predicted events.
- Balanced Accuracy: the measure of how accurate is the overall performance of a model considering both positive and negative classes without worrying about the imbalance of a data set.

The confusion matrix has to be analyzed in order to interpret it in a correct way. Some metrics could be more useful to understand how the model is working.

In the next section is presented a case of use based on a corpus mainly composed of 4 different public profiles of politicians in Mexico and others users extracted from the hashtag *#UniformeNeutro*.

4. Case of use and results

The case of use presented in this article concerns the analysis of 4 public profiles of politicians in Twitter in order to classify them according to their tweets. The main interest is to discover in what cases a tweet is attributed to the politician according to the words used and in what other cases it is classified as *other* so it is possible to affirm that the tweet could have been written by someone else.

The classification model was implemented in R because of the facilities given through the use of their libraries like: *twitteR*, *tidyverse*, *tidytext*, *naivebayes* or *caret*.

For the case of use, the corpus is composed of tweets from 4 Mexican political public profiles (*@lopezobrador_*, *@EPN*, *@FelipeCalderon* and *@VicenteFoxQue*) as other users not concerning the political field.

To be able to analyze the tweets it was important to follow all the steps explained in Section 3. In the data import step, 4185 tweets were retrieved by specifying each of the profiles and then the hashtag *#UniformeNeutro*, an initiative of the Mexican government to allow girls and boys to choose between pants or skirts to use in the school. Obviously due to the controversy, the tweets are variated with messages addressing the government. The data import concerns, also, the replacement of badly downloaded characters so the tweets can be processed without problems. After that, the table that contains the tweets is transformed into a data frame by specifying that only it will be retrieved the columns concerning the *id*, *screen name* and *tweet*. From the data frame, it is possible to create a sparse matrix where each row stands for a tweet and each column for a word (see Figures 1 and 2). For all the political profiles the number of tweets are exactly the same so each matrix is composed of 675 rows and 9485 variables representing each word. These matrices tend to get very big, so the case of use is only an example in order to be able to manipulate the data and analyze it correctly.

The prediction model aims to estimate when a tweet belongs to a specific user. For this test, the model tries to predict whether a tweet was made by one of the 4 political authors accounts or not. Since the others profiles are not of interest they are tagged with the label “*other*”.

As explained in the previous section, to be able to adjust the model a training and a test set is used. Also, the function *naivebayes()* is applied with the training set. With this model, the predictions are made. The prediction model is applied to the 4 accounts. In Figure 3, there is an example of the vector obtained for the first 65 tweets, when using a prediction for the public profile *@VicenteFoxQue*.


```

[1] VicenteFoxQue Other      VicenteFoxQue VicenteFoxQue Other      VicenteFoxQue VicenteFoxQue
[8] VicenteFoxQue VicenteFoxQue VicenteFoxQue Other      Other      VicenteFoxQue VicenteFoxQue
[15] VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue Other
[22] VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue Other      Other      Other
[29] VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue Other      Other      VicenteFoxQue
[36] VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue
[43] VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue Other      VicenteFoxQue
[50] VicenteFoxQue Other      VicenteFoxQue VicenteFoxQue Other      Other      VicenteFoxQue
[57] Other      VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue VicenteFoxQue Other
[64] Other      VicenteFoxQue
Levels: Other VicenteFoxQue

```

Fig. 3. Vector obtained with the *screenName* values that have been predicted by the model.

To evaluate the success of the prediction model, the function *confusionmatrix()* receives the vector with the predictions and the actual values of *screenName* (the name of the account). This process was made for the 4 profiles.

To analyze the results obtained, the figures (4, 5, 6 and 7) and tables (1, 2, 3, and 4) will be used. The initial prediction model was used but it was adjusted after several tests.

4.1. Results for @lopezobrador_.

From 200 cases identified for the account @lopezobrador_ in the test set, 194 were correctly classified (Figure 4). This means, a sensitivity of 97% (Table 1), only 6 were classified as *other*. For the accounts classified as *other*, of a total of 3985, only 727 were correctly classified, that is, 18% of specificity.

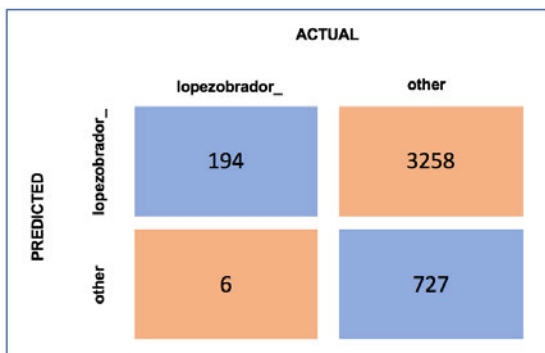


Fig. 4. Visualization of the confusion matrix for the public account @lopezobrador_.

Table 1 shows the metrics obtained for @lopezobrador_ with a precision (accuracy) of 73%.

Another useful measure corresponds to the Kappa statistic which gives a measure of how much the model improves a prediction against the observed

probabilities. The closer to 1 is the Kappa value it means that the model is better than the expected probability. For this analysis, the values that are considered ideals are situated above 0.6. In this case of use, the prediction model was ameliorated and also the Kappa value, this is shown as the different examples are presented in the next subsections. The Kappa value for the prediction model used for @lopezobrador_ got a value of 0.1987 which was still very low.

The positive predictive value (Pos Pred Value), as explained in the previous section, indicates the probability that a data that has been predicted as belonging to the positive category, really belongs to it ('Positive' class: @lopezobrador_ in this example). In this case, @lopezobrador_ the probability is of 73.43%. By complement, the negative predictive value (Neg Pred Value) indicates the probability that a data predicted as belonging to the negative category (*other*), in effects corresponds to that. The negative predictive value obtained was of 74.8%.

Finally, the balanced precision indicates how well the model predicts both positive and negative categories. This is very important with data, like the presented here, where the classes are unbalanced, that is, one is more abundant and more likely to appear than the other. In data sets like these, it is easy to obtain a high accuracy (73%) for the most probable class, even if there is low for the less probable class. The balanced accuracy is 57.80% which have to be improved.

Table 1

Metrics obtained in each confusion matrix for the public account of @lopezobrador_.

	@lopezobrador
Accuracy:	0.7353
95% CI:	(0.7272, 0.7433)
No Information Rate:	0.6989
P-Value [Acc > NIR]:	<2.2e-16
Kappa:	0.1987
Mcnemar's Test P-Value:	<2.2e-16

Sensitivity:	0.9735
Specificity:	0.1824
Pos Pred Value:	0.7343
Neg Pred Value:	0.7480
Prevalence:	0.6989
Detection Rate:	0.6804
Detection Prevalence:	0.9266
Balanced Accuracy:	0.5780
'Positive' Class:	@lopezobrador

4.2. Results for @VicenteFoxQue.

Similar to the results of @lopezobrador_, with @VicenteFoxQue 194 were correctly classified (Figure 5) obtaining a sensitivity of 97% (Table 2), only 6 were classified as other. For the accounts classified as other, 777 were correctly classified, which means that 19% corresponds to specificity.

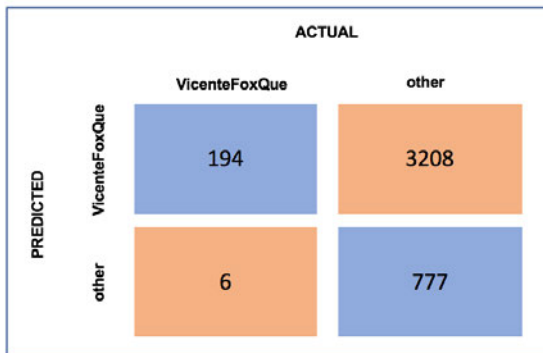


Fig. 5. Visualization of the confusion matrix for the public account @VicenteFoxQue.

The Kappa value for the prediction model used for @VicenteFoxQue is better than the one obtained in @lopezobrador_ test, with a value of 0.2261. However, this value was still low.

The positive predictive value and the negative predictive value are 89% ('Positive' class: @VicenteFoxQue) and 48% respectively.

The accuracy obtained is of 88% while the balanced accuracy is 58.37% which have to be improved.

Table 2

Metrics obtained in each confusion matrix for the public account of @VicenteFoxQue

	@VicenteFoxQue
Accuracy:	0.8804
95% CI:	(0.8748, 0.8858)
No Information Rate:	0.8816
P-Value [Acc > NIR]:	0.6705

Kappa:	0.2261
Mcnemar's Test P-Value:	<2e-16
Sensitivity:	0.9725
Specificity:	0.1949
Pos Pred Value:	0.8999
Neg Pred Value:	0.4875
Prevalence:	0.8816
Detection Rate:	0.8573
Detection Prevalence:	0.9527
Balanced Accuracy:	0.5837
'Positive' Class:	@VicenteFoxQue

4.3. Results for @EPN.

The results for @EPN are kindly different from the ones obtained with @lopezobrador_ and @VicenteFoxQue. In this case, 176 were correctly classified (Figure 6) with a sensitivity of 88% (Table 3), having 24 classified as other. For other, 1974 were correctly classified, which means that the specificity is almost of 50%.

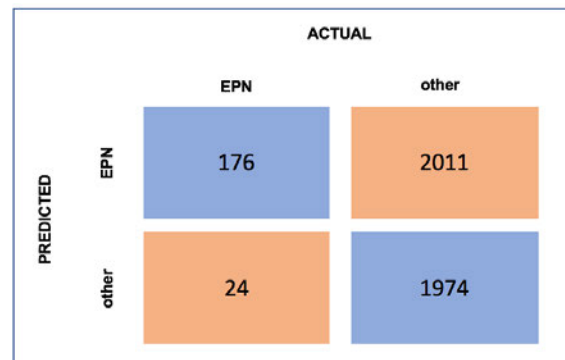


Fig. 6. Visualization of the confusion matrix for the public account @EPN.

Unlike the two Kappa values presented above, the obtained for @EPN shows a considerable improvement with a value of 0.4024 (see Table 3).

The positive predictive value and the negative predictive value are 81% and 61% respectively. However, in this example, the 'Positive' class corresponds to other.

The accuracy obtained is of 77% while the balanced accuracy is 68.89% which is bigger than the two previous examples.

Table 3

Metrics obtained in each confusion matrix for the public account of @EPN

	@EPN
Accuracy:	0.775

95% CI:	(0.7309, 0.815)
No Information Rate:	0.7225
P-Value [Acc > NIR]:	0.00994
Kappa:	0.4024
Mcnemar's Test P-Value:	0.02686
Sensitivity:	0.8824
Specificity:	0.4955
Pos Pred Value:	0.8199
Neg Pred Value:	0.6180
Prevalence:	0.7225
Detection Rate:	0.6375
Detection Prevalence:	0.7775
Balanced Accuracy:	0.6889
'Positive' Class:	Other

4.4. Results for @FelipeCalderon.

After some training and adjustment of the prediction model, it shows a better performance comparing it with the results obtained in the 3 previous cases. So, for @FelipeCalderon 199 were correctly classified (Figure 7) with a sensitivity of 99% (Table 4) and only one classified as *other*. For *other*, 2103 were correctly classified, which means that the specificity is of 52%.

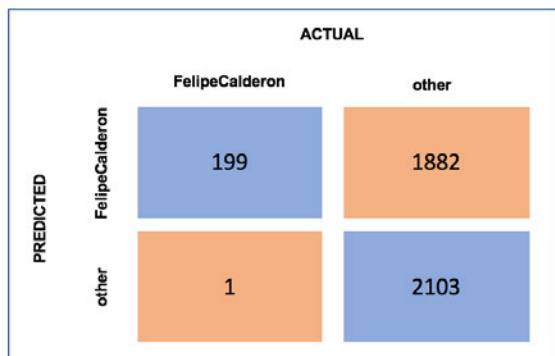


Fig. 7. Visualization of the confusion matrix for the public account @FelipeCalderon.

In both cases of the predictive value, the positive and the negative, the values are close to the 97%. The 'Positive' class corresponds to @FelipeCalderon.

The accuracy obtained is of 97% while the balanced accuracy is 76.36% which is a very good value in the last case of use by using a model improved thanks to the above results.

Finally, the prediction model can be considered as acceptable, or ideal, the Kappa value is of 0.6746 (see Table 4).

Table 4

Metrics obtained in each confusion matrix for the public account of @FelipeCalderon

	@FelipeCalderon
Accuracy:	0.9787
95% CI:	(0.9704, 0.9851)
No Information Rate:	0.9561
P-Value [Acc > NIR]:	6.456e-07
Kappa:	0.6746
Mcnemar's Test P-Value:	6.33e-08
Sensitivity:	0.9994
Specificity:	0.5278
Pos Pred Value:	0.9788
Neg Pred Value:	0.9744
Prevalence:	0.9561
Detection Rate:	0.9555
Detection Prevalence:	0.9762
Balanced Accuracy:	0.7636
'Positive' Class:	@FelipeCalderon

5. Conclusions and further work

Text classification on Twitter recently attracted research interest in politics using Information Retrieval and NLP. However, the vast majority of work related to the automatic classification of text has been focused on long texts. Twitter imposes new challenges in working with unstructured data. The objectives of text classification are very different, some of them tend to classify text into classes or categories in order to identify main themes.

In this article, text classification is based on the analysis of tweets because, unlike other information sources, Twitter is up-to-date and reflects the news and events occurring daily all over the world. The interest of analyzing tweets concerns the authorship identification regardless that, in politician accounts, the author is already known. However, its well known that these accounts are managed by a group of people and not always is the same person who creates the tweet. So, the present work is interested in finding some clues to identify authorship and the number of people working behind the scenes after applying a classification model based on Naïve Bayes' theorem.

In this article, a framework for experimentation based on Naïve Bayes was implemented by using R. Given a public profile, of a politician, the tweets are retrieved, processed and classified according to the author or *other*. The analysis and interpretation of the results are possible thanks to a confusion matrix, with different metrics, created for each profile. Experimental results show that with only a small set of features, the classifier achieves a significant improvement in accuracy. However, there is an opportunity to continue the work in testing the classifier accuracy and performance at larger scales, as well as in several other

areas. While the results with four authors are promising, a Kappa value of 67%, accuracy of 97% and balanced accuracy of 76%, real-world applications require thousands of authors. Also, it would be interesting to determine when accuracy falls when using a bigger corpus.

References

- [1] A. Castro, and B. Lindauer, Author Identification on Twitter, 2012.
- [2] B. Sriram, Short text classification in twitter to improve information filtering. Doctoral dissertation, The Ohio State University, 2010.
- [3] D. Murthy, Twitter: Social Communication in the Twitter Age. Cambridge, UK: Polity Press, 2013, 193 pp.
- [4] E. AlSukhni and Q. Alequr. Investigating the use of machine learning algorithms in detecting gender of the Arabic tweet author. *International Journal of Advanced Computer Science & Applications*, 1(7), 2016, pp. 319–328.
- [5] J. D. Burger, J. Henderson, G. Kim G. Zarrella, Discriminating Gender on Twitter. *Proceedings of the Conference on Empirical Methods in Natural Language Processing. EMNLP '11*, Association for Computational Linguistics, Stroudsburg, PA, USA, 2011. pp. 1301–1309.
- [6] M. Sultana, P. Polash, and M. Gavrilova, Authorship recognition of tweets: A comparison between social behavior and linguistic profiles. *IEEE International Conference on Systems, Man, and Cybernetics (SMC) 2017*, pp. 471-476.
- [7] P. Galán-García, J. G. D. L. Puerta, C. L. Gómez, I. Santos, and P. G. Bringas, Supervised machine learning for the detection of troll profiles in twitter social network: Application to a real case of cyberbullying. *Logic Journal of the IGPL*, 24(1), 2016. pp. 42-53.
- [8] R. Green and J. Sheppard, Comparing frequency- and style-based features for twitter author identification, *Proceedings of the International Florida Artificial Intelligence Research Society Conference (FLAIRS)*, 2013.
- [9] R. Sousa Silva, G. Laboreiro, L. Sarmento, T. Grant, E. Oliveira and B. Maia, 'twazn me!!! ;(' Automatic authorship analysis of micro-blogging messages. Muñoz R., Montoyo A., Métais E. (eds) *Natural Language Processing and Information Systems, NLDB 2011. Lecture Notes in Computer Science*, vol 6716. Springer, Berlin, Hidelberg, 2011. pp. 161-168.
- [10] S. Mechti, M. Jaoua, L. H. Belguith and R. Faiz, Machine learning for classifying authors of anonymous tweets, blogs, reviews and social media. *Proceedings of the PAN@ CLEF*, Sheffield, England, 2014.
- [11] X. H. Phan, L. M. Nguyen and S. Horiguchi, Learning to classify short and sparse text & web with hidden topics from large-scale data collections, *Proceedings of the World Wide Web*, ACM, Beijing, China, 2008, pp. 91-100.

PROTOTIPO DIGITAL PARA FOMENTAR LA CIVILIDAD EN REDES SOCIALES

DIGITAL PROTOTYPE TO FOSTER CIVILITY IN SOCIAL NETWORKS

Erick López Ornelas

Universidad Autónoma Metropolitana, Unidad Cuajimalpa, México
elopez@cua.uam.mx

Rocío Abascal Mena

Universidad Autónoma Metropolitana, Unidad Cuajimalpa, México
mabascal@cua.uam.mx

Recepción: 30/octubre/2020

Aceptación: 27/noviembre/2020

Resumen

En este artículo se describe la propuesta de un prototipo basado en una aplicación móvil diseñada para fomentar la civilidad digital dentro de la red social Facebook, llamada Facebook Civility, con la cual se propone moderar las conversaciones derivadas de comentarios generados dentro de la red social. El objetivo del prototipo es que las personas reflexionen sobre los tipos y niveles de violencia que ejercen a partir de sus comentarios en redes sociales. A través de varios ejercicios, los usuarios podrán experimentar las distintas maneras de expresar su opinión sin tener que recurrir a lenguaje ofensivo. En este artículo se describen las diferentes etapas del diseño y la evaluación realizada sobre el prototipo.

Palabras Clave: Civilidad digital, polarización de opinión, redes sociales.

Abstract

This article describes the proposal of a prototype based on a mobile application designed to promote digital civility within the Facebook social network, called Facebook Civility, with which it is proposed to moderate the conversations derived from comments generated within the social network. The objective of the prototype is for people to reflect on the types and levels of violence they exercise based on

their comments on social networks. Through various exercises, users will be able to experience different ways of expressing their opinion without resorting to offensive language. This article describes the different stages of the design and the evaluation carried out on the prototype.

Keywords: *Digital civility, mixed opinions, social network.*

1. Introducción

La interacción de los usuarios a través de redes sociales está plagada de violencia y frecuentemente esta violencia se traspasa del plano virtual al físico y viceversa.

En México, a partir de las elecciones presidenciales del año 2018, en redes sociales se evidenció el problema de la polarización social y política, por la diversidad de opiniones en relación con la permanencia del Partido Revolucionario Institucional (PRI) y el triunfo del populismo, representado por el candidato del partido Movimiento de Regeneración Nacional, Andrés Manuel López Obrador [Domínguez, 2018].

Si bien el objetivo de las redes sociales está ligado a la libertad de expresión también existe la corresponsabilidad que la red social Facebook estaría en posibilidades de asumir, para fomentar espacios libres de violencia en sus distintas manifestaciones. La propuesta que se plantea como solución en el presente artículo, Facebook Civility, se enfoca en dar solución al problema de violencia causado por polarización política, sin suprimir los derechos del usuario [Belaunzarán, 2019], [Tecnósfera, 2019]. El proyecto plantea que la funcionalidad de la interfaz utilice el procesamiento del lenguaje natural (Natural Language Processing [NLP]). Este método permite comprender y extraer información significativa de los mensajes y comentarios que los usuarios envían, denominados entidades; posteriormente, se utiliza la intención de estas entidades para automatizar algunas de sus respuestas, dirigir la conversación a una charla humana y recopilar datos de la audiencia [Chowdhary, 2020]. Por diversos hechos de violencia generados entre usuarios de Facebook alrededor del mundo, Mark Zuckerberg, CEO de esta empresa, tuvo que tomar acciones para mediar y buscar combatir la violencia en esta red social. Para ello, en

2018 anunció que se emplearía a 20 mil personas para vigilar los discursos de odio y mensajes reportados como violentos. Sin embargo, el mismo Zuckerberg ha aceptado que esto no es suficiente y se encuentran ante un reto difícil ya que Facebook está en más de 200 países, con su propia cultura y lenguas [Wu, 2019]. Otra red social que busca moderar los mensajes agresivos es Twitter [Kursuncu, 2019]. El 27 de junio de 2019 la compañía anunció que restará visibilidad a los mensajes que no cumplan con las normas de la red social, específicamente de las figuras públicas, como el presidente de los Estados Unidos, Donald Trump. A los usuarios les aparecerá una alerta que les avisará que están violando las reglas y le restará visibilidad para que lleguen a un menor número de usuarios en Twitter. La empresa justificó estos cambios en su algoritmo para equilibrar entre la responsabilidad de cada usuario y la libertad de expresión.

Más allá de Facebook, existen proyectos que buscan disminuir la violencia, como es el caso de ReThink [Re-Think, 2019]. Esta aplicación móvil invita a los usuarios, en su mayoría adolescentes, a pensar antes de publicar un mensaje con lenguaje ofensivo. La iniciativa de ReThink es que los jóvenes tengan una segunda oportunidad antes de emitir alguna ofensa y con esto combatir los casos más graves de cyberbullying.

Otro ejemplo de aplicación móvil es la desarrollada por la Fundación en movimiento llamada Bullysemáforo, con la cual se pretende prevenir, detectar e incluso resolver casos de acoso escolar [Evans, 2019].

A partir de esta indagación se encontró que no existen interfaces de habla hispana destinadas a combatir el problema de violencia en redes sociales, aunque esté identificado y categorizado como una alerta social en el país.

2. Métodos

El método aplicado se llevó a cabo a través de un proceso conformado por el planteamiento del problema o *Initial design Testing*, la definición y creación de usuarios con la técnica *Personas*, desarrollo de casos de uso con el uso de *Storyboard*, la conceptualización del video y finalmente la realización del prototipo.

Initial design testing

Como primera etapa se definió problema relacionado con la civilidad digital que se pudiera resolver a través de medios digitales e Internet.

Una vez que se tuvo la delimitación del tema a la polarización de opiniones en Facebook, se procedió a diseñar una entrevista para poder obtener información sobre cómo la gente vive la violencia en redes sociales y cuáles son sus puntos de vista.

Se realizaron un conjunto de entrevistas para obtener información cualitativa de la situación. Con los resultados se realizó un análisis para identificar las principales necesidades de los entrevistados y los perfiles generales de estos. Se generó una lista de necesidades, de la cual se eligió el crear espacios de tolerancia y respeto en redes sociales. Esta etapa fue fundamental para la detección de la problemática a abordar: la civilidad en las redes sociales.

Personas

Se utilizó la técnica Personas de Allan Cooper [Haag, 2019] que consiste en crear personajes ficticios para representar al usuario de la manera más realista posible, se define su edad, algunas características, sus metas, estudios, ocupación, destrezas tecnológicas, habilidades y frustraciones.

A partir de los perfiles identificados previamente, se resumieron en dos que estuvieran implicados en el problema planteado. Este método es muy importante ya que sirve para generar empatía con los usuarios y poder determinar de manera adecuada el perfil de los usuarios con los que se trabajará.

La Persona 1 se refiere al usuario que se autocensura, no comentando su opinión o no comentando en lo absoluto, para evitar confrontaciones ya sea por temor a discutir con sus amigos o a recibir insultos (figura 1).

El segundo perfil identificado se refiere a la Persona 2, que son los que emiten su opinión, e incluso responden a otros comentarios sin importar las repercusiones que esto podría causar, tiene una postura y la defiende, los que estén en contra están mal para esta persona (figura 2).

Persona que se autocensura		Metas
	<p>Ramón</p> <p>Tiene 25 años Vive con sus papás Es de clase media baja Fue la primera vez que pudo emitir su voto, y este fue a favor de MORENA</p>	Graduarse
Estudios y ocupación	Destrezas tecnológicas	Frustraciones
<p>Estudia Sociología en la facultad de Ciencias Políticas y Sociales</p> <p>Es becado, los recursos de su familia no eran suficientes para estudiar</p>	<p>Usa redes sociales como Twitter, Facebook, Instagram, Snapchat y whatsapp</p> <p>A pesar de estar informado, a veces no revisa las fuentes de información</p>	<p>Prefiere no opinar en redes sociales puesto que en su licenciatura las opiniones son muy polarizadas y no quiere perder amigos o que lo tachen por sus opiniones</p> <p>La corrupción que se vive en el país</p>

Figura 1 Primer perfil de usuario definido por el método de *Personas*.

Persona que confronta		Metas
	<p>Laura</p> <p>Tiene 35 años Tiene 1 hijo y vive en unión libre Vive en el centro del país Se considera anti Amlo y siempre ha votado por el PRI Es clase media</p>	<p>Hacer crecer su negocio</p> <p>Que su hijo tenga una buena educación</p>
Estudios y ocupación	Destrezas tecnológicas	Frustraciones
<p>Estudió la licenciatura en administración de empresas</p> <p>Tiene su propio negocio de venta de ropa para dama</p>	<p>Usa las aplicaciones más comunes y de moda como Uber.</p> <p>Prefiere Facebook e Instagram.</p>	<p>La situación de violencia en el país</p> <p>No tener suficiente tiempo para ella</p> <p>No le gustan las noticias, pero trata de mantenerse informada.</p>

Figura 2 Segundo perfil de usuario definido por el método de *Personas*.

Storyboard

El Storyboard es una técnica de diseño que se utilizó para realizar una primera propuesta de solución de la aplicación. El storyboard es muy útil para identificar la satisfacción final del usuario, además ayuda en la conceptualización de la aplicación. Se generaron 3 propuestas de casos de uso visualizadas a través de un storyboard que consistían en:

- Añadir un filtro de privacidad similar al que existe en las publicaciones, pero dirigido a los comentarios, con el fin de que éste solo pueda ser visualizado por ciertos usuarios definidos por quien está comentando.

- Detectar y bloquear las cuentas de los usuarios que emitan mensajes violentos.
- Una interfaz que suavice el lenguaje ofensivo y/o cambie el discurso de odio por uno amigable y de esta manera crear disrupción en el usuario.

Estas propuestas se evaluaron y analizaron por su conveniencia, eficiencia y originalidad, bajo estos criterios se eligió la propuesta número tres. El caso de uso consistía en un usuario que, al observar un comentario con el que no está de acuerdo, responde de manera ofensiva por lo que la interfaz cambia su discurso con palabras ofensivas por uno amable, por lo que la conversación que se pudo desarrollar de manera agresiva y ofensiva se convierte en una conversación civilizada donde ambas partes quedan satisfechas (figura 3).



Figura 3 Storyboard utilizado para implementar el escenario con mayor civilidad.

Prototipo en papel

Para el primer boceto de la interfaz, se creó un prototipo de papel (figura 4) para conocer diferentes formas de conceptualizar la interfaz. Se seleccionaron elementos de cada idea para conformar una sola interfaz. A partir de esta propuesta conjunta se simuló el flujo de pantallas y los posibles caminos que integrarían el caso de uso.

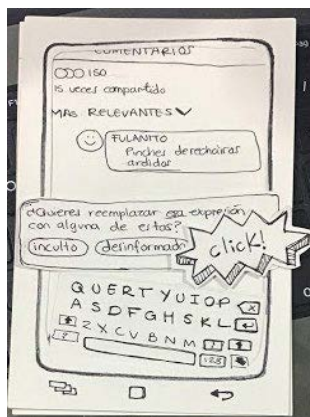


Figura 4 Prototipo en papel utilizado para la conceptualización de las ideas.

Se definieron posteriormente, dos tipos de insultos que deberían de ser detectados por la aplicación. Estos insultos tienen que ver con el uso de palabras altisonantes e insultos compuestos. La idea es que estos insultos puedan ser intercambiados por sinónimos menos agresivos y que inciten a una mejor convivencia e intercambio de mensajes de forma civilizada entre los usuarios (tabla 1).

Tabla 1 Palabras altisonantes e insultos compuestos.

Palabras altisonantes	Insulto compuesto
Pendejo	Puta madre
Puto	Pinche madre
Derechairo	Hijo de puta
Pejzombie	Pinche pendejo
Chingar	Su puta madre
Ignorante	Pinche mamón
Alv	Chingue a su madre
Imbécil	Pinche resentido
Idiota	A la chingada

Prototipo digital

Para crear el prototipo interactivo se utilizaron dos herramientas. Para el diseño gráfico de la interfaz se utilizó el programa de ilustración vectorial Illustrator y para diagramar la funcionalidad de la interfaz la plataforma de diseño de productos digitales Marvel (<http://www.marvelapp.com>), figuras 5 y 6).

Se siguieron las métricas de estilo establecidas en Facebook, para que esta se mimetizara con la identidad gráfica de la red social. La configuración y los textos de introducción para la aplicación móvil está inspirada en la aplicación Facebook

Parejas, puesto que la intención es que Facebook Civility funcione de la misma manera forma, como un apéndice o add-on de Facebook.

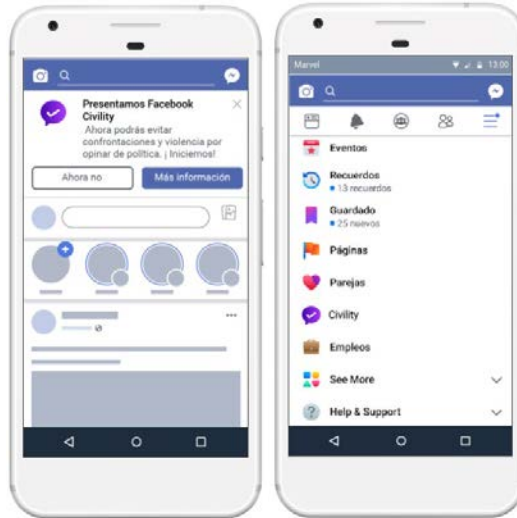


Figura 5 Prototipo digital utilizado para la conceptualización de la aplicación.

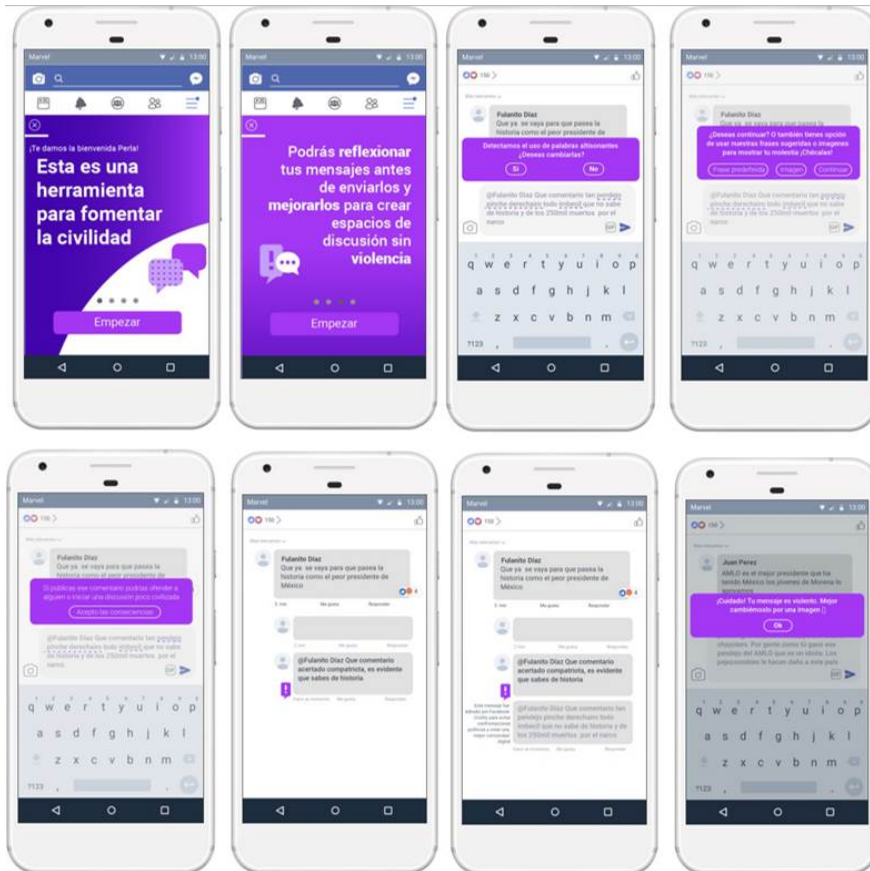


Figura 6 Prototipo digital donde se ejemplifican las funcionalidades implementadas.

El funcionamiento del diagrama inicia con la instalación de la aplicación, para esta acción existen dos caminos, el primero se realiza a través de la configuración general de Facebook y el otro a través de una sugerencia para activar la aplicación (figura 5). El siguiente paso es, que al encontrar una noticia y dirigirse a la sección de comentarios, selecciona uno al cual responder por qué no coincide con sus ideales y quiere demostrar su enojo o disgusto, por lo que la interfaz ofrece una primera opción cambiar las palabras identificadas como agresivas.

Si el usuario acepta cambiar las palabras, se cambian las altisonantes por sinónimos más amables y con connotación menos ofensiva, mientras que las palabras compuestas se eliminan, dejando un mensaje con el mismo sentido del discurso, pero menos violento. Si el usuario no acepta se despliegan tres opciones más, donde intenta hacer recapacitar al usuario ofreciendo otras opciones: usar una imagen o GIF de la base de imágenes creada, con las cuales demostrar principalmente enojo o desacuerdo, sin tener connotaciones políticas u ofensivas ya que se eligieron cuidadosamente memes populares en la cultura mexicana que fueran neutros, no se prestaran a dobles sentidos o sarcasmo y su significado fuera obvio; usar una frase predefinida por el sistema para expresar el disgusto del usuario de manera educada y como última opción, continuar para publicar su comentario. Si el usuario rechaza todas las oportunidades de cambiar su comentario y decide publicarlo con las palabras ofensivas, se cambian las palabras detectadas como insultos con antónimos, cambiando de esta manera el sentido del discurso del usuario, aunque también ofrece la opción de ver el comentario original para evitar caer en la censura. Si cualquier comentario tiene errores gramaticales o más de 10 palabras identificadas como insulto, la interfaz automáticamente omite las opciones de cambio de palabra y ofrece al usuario la opción de publicar una imagen o GIF que exprese su opinión.

3. Resultados

Cuando se tuvo un primer prototipo digital se realizó la primera evaluación, la cual fue una Evaluación Heurística (EH). Una EH es un método de inspección de la usabilidad sin usuarios (Nielsen, 1994). Esta, consiste en examinar la calidad de

uso de una interfaz por parte de varios evaluadores expertos, a partir del cumplimiento de unos principios reconocidos de usabilidad: los heurísticos. Esta evaluación nos ayudó a verificar los principios de usabilidad y qué heurísticas no se estaban cumpliendo, por lo que algunos de los cambios que se realizaron fueron:

- Control y libertad: Se tenían solo dos opciones de mensajes, por lo que se replanteó el funcionamiento y se añadieron más opciones de cambio de palabra como frases predefinidas e imágenes.
- Prevención de errores: Se añadió una introducción de la aplicación, para explicar al usuario que es Facebook Civility y cómo funciona, así mismo se cambió su funcionamiento de aplicación externa a un complemento add-on dentro de Facebook.

Como segunda etapa de evaluación se realizó una prueba de usabilidad con un usuario real, la cual consistió en mostrarle el prototipo y pedirle que, siguiendo las métricas del protocolo Think-Aloud (McDonald, 2013), describiera su interacción con la interfaz, al igual que dudas o ideas que surgieran durante el proceso. Al mismo tiempo, se registró presencialmente la interacción, mientras que un grupo de expertos, lo observaba de manera remota a través de dos cámaras colocadas en la sala [Rubin, 2008]. Una de las cámaras fue posicionada directamente sobre el Smartphone para conocer el desempeño de la interfaz; y otra permaneció dirigida hacia el usuario, para observar sus gestos y actitud hacia la interfaz.

Los hallazgos detectados en esta fase de evaluación fueron:

- La falta de algún botón para regresar a la pantalla inicial. Como el prototipo muestra varias posibles opciones y rutas, se volvía tedioso regresar al inicio sólo con el botón de ir hacia atrás.
- Evidenciar las rutas activas del prototipo. Por el mimetismo del prototipo con Facebook, el usuario quería interactuar con él como lo hace con la aplicación real y recurría a dar clic en todos los elementos hasta encontrar el que si tenía funcionamiento programado.
- La barra de botones colocada para simular el teléfono móvil confundía al usuario, ya que, por inercia, intentaba interactuar con esta.

En general, se obtuvieron resultados positivos y las funcionalidades implementadas resultaron de agrado y benéficas para los usuarios, los cuales comentaron que si les será de gran ayuda, sobre todo cuando existe un intercambio de mensajes entre usuarios.

4. Discusión

El hablar de civilidad en las redes sociales, es un elemento complejo debido a que las redes sociales brindan un cierto anonimato y una forma de ser desinhibido en la interacción que se realiza. Es por eso que al interactuar se debe buscar siempre tener cierta empatía y amabilidad en cada interacción, y tratar a todos con los que se conecta en línea con dignidad y respeto.

De igual forma hay que tomar en cuenta que todos somos diferentes por lo que se deben de considerar las diversas perspectivas y cuando surjan desacuerdos, participar cuidadosamente y evitar insultos y ataques personales.

Es importante hacer una pausa y reflexionar antes de responder a cosas con las que no está de acuerdo, y no publicar ni enviar nada que pueda dañar a otra persona, afectar una reputación o amenazar la seguridad de alguien. De igual manera apoyar a los demás y denunciar actividades amenazantes y preservando evidencia de comportamiento inapropiado o inseguro. Estos son algunos de los elementos que, en este artículo, se intenta reflexionar. Al ser un prototipo no todas las funcionalidades han sido implementadas, pero si deja ver el alcance y la forma en que éste ayudará al usuario y fomentará la civilidad en las redes sociales.

5. Conclusiones

En esta sección resulta importante resaltar el proceso de diseño que se siguió para la generación de la propuesta, desde la búsqueda de necesidades y poder identificar la problemática, hasta la aplicación de un conjunto de técnicas como el storyboard y la utilización de prototipos para la generación de la propuesta. Se aplicaron de igual forma dos tipos de evaluaciones, la verificación de heurísticas y la pruebas con usuarios reales, se recibió una crítica positiva respecto a la interfaz y la premisa en la que está basada.

El objetivo de crear disrupción en el usuario se cumple y genera en él una reflexión para no ser tan agresivo, al mismo tiempo que brinda opciones de cómo mejorar sus comentarios para evitar ofender y de esta manera ser ofendido en redes sociales.

Facebook civility, es una propuesta que busca fomentar la civilidad en las interacciones realizadas en Facebook y donde los usuarios tengan la capacidad de discernir entre un buen comentario amable y un comentario agresivo. Esto por supuesto fomenta una mejor interacción entre los usuarios, el intercambio de ideas y aumenta la civilidad que debería siempre existir en el mundo virtual, donde las redes sociales nos han invadido en la actualidad.

Es importante mencionar que, en el presente artículo, se busca encontrar una solución viable al problema de la civilidad en las redes sociales. En el artículo no detalla la manera de identificación automática de mensajes ni las técnicas de tratamiento de lenguaje natural utilizadas.

Como trabajo a futuro, es interesante el poder realizar la implementación real de la aplicación, donde forzosamente se deberá de contar con una base de datos además de la implementación de algoritmos de procesamiento de lenguaje natural para poder realizar la verificación y el intercambio de conceptos que se viertan por parte de los usuarios de la aplicación.

6. Bibliografía y Referencias

- [1] Belaunzarán F., (2019). México polarizado. Excelsior: <https://www.excelsior.com.mx/opinion/opinion-del-experto-nacional/mexico-polarizado/1303797>.
- [2] Chowdhary K. R., (2020). Natural Language Processing. In: *Fundamentals of Artificial Intelligence*. Springer, New Delhi.
- [3] Domínguez González G., (2018). Elección, polarización y hartazgo social en México. Nueva sociedad: <https://nuso.org/articulo/eleccion-polarizacion-y-hartazgo-social-en-mexico/>.
- [4] Evans C. B. R., Smokowski, P. R., Rose R. A. et al., (2019). Cumulative Bullying Experiences, Adolescent Behavioral and Mental Health, and Academic Achievement: An Integrative Model of Perpetration, Victimization,

- and Bystander Behavior *J. Child Fam Stud* 28, 2415–2428. <https://doi.org/10.1007/s10826-018-1078-4>.
- [5] Haag M., Marsden N., (2019). Exploring personas as a method to foster empathy in student IT design teams. *Int J Technol Des Educ* 29, pp. 565–582. <https://doi.org/10.1007/s10798-018-9452-5>.
- [6] Kursuncu U., Gaur M., Lokala U., Thirunarayan K., Sheth A., Arpinar I. B., (2019). Predictive Analysis on Twitter: Techniques and Applications. In: Agarwal N., Dokoochaki N., Tokdemir S., (eds) *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*. Lecture Notes in Social Networks. Springer, Cham.
- [7] McDonald S., McGarry K., Willis L.M., (2013). Thinking-aloud about web navigation: the relationship between think-aloud instructions, task difficulty and performance. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 57, pp. 2037–2041. SAGE Publications, Los Angeles.
- [8] Nielsen J., (1994). Heuristic evaluation. In Nielsen J., and Mack R.L. (Eds.), *Usability Inspection Methods*, John Wiley & Sons, New York, NY.
- [9] Re-Think Illinois, (2019). Estados Unidos: <http://www.rethinkwords.com>
- [10] Rubin J., (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley Publishing Inc.
- [11] Tecnósfera, (2019). Cambio en Twitter podría reducir visibilidad de Trump. El Tiempo: <https://www.eltiempo.com/tecnosfera/novedades-tecnologia/el-cambio-de-politicas-de-twitter-que-puede-quitar-visibilidad-a-trump-381554>.
- [12] Wu, C. et al., (2019). Machine Learning at Facebook: Understanding Inference at the Edge, 2019 IEEE International Symposium on High Performance Computer Architecture (HPCA), Washington, DC, USA, pp. 331-344, doi: 10.1109/HPCA.2019.00048.

An interactive prototype to spot Fake News in young people

Erick López Ornelas¹, Rocío Abascal Mena¹

¹Universidad Autónoma Metropolitana - Cuajimalpa
{elopez,mabascal}@cua.uam.mx

Abstract. Do you know how to identify Fake News? This project intended to ensure that young people under twenty years with access to the internet and social networks can identify important elements of news to distinguish from fake ones. The creation of an interactive prototype will be guided taking into account the importance of the user experience (UX) for the development of the interactive prototype. This paper explains the main steps in the methodology in order to recommend some issues to spot Fake News in young people.

Keywords: Fake News, User Center Design, User Research Methods, User Experience

1 Introduction

The term “fake news” (FN) was officially ushered into the lexicon when the Oxford Dictionary added the term in 2017 [1]. While the term is frequently used and definitions vary, the problem of deceptive data is serious and exposes a profound and underlying flaw in information and network security models. This flaw is trust in entities without verification of the content that they exchange.

“Trust but verify” [2] is an old proverb that, until recently, resulted in trust at the expense of verification. Fake news has become a social phenomenon due to the confusion they cause in society. Every day, individuals through social networks are surrounded by large amounts of information of all kinds, however, more information is not synonymous with having reliable data, on the contrary, increasingly more sites and publications dedicated to disseminate proliferate distorted information or, worst, totally fake [3]. False information is very harmful to society, as it can escalate to unexpected levels and manipulate the process of making important decisions such as political elections [4]. Malicious processes of this nature directly affect our lives, since they involve our environment and our person. The purpose of this study, is to contribute to the right identification of the different characteristics of a news item, in order to improve the users' criteria when browsing through social networks (Facebook in particular) and internet pages. We focus on young people, because they are the

ones who naturally experience digital technology also, they receive and spread large amount of data and news on internet.

2 Background

Detecting fake news in social media has been an extremely important, yet technically very challenging problem. In one study, human judges, by a rough measure of comparison, achieved only 50- 63 % success rates in identifying fake news [5]. The most of fake news detection algorithms try to linguistic cues [6]. Several successful studies on fake news detection have demonstrated the effectiveness of linguistic cue identification, as the language of truth news is known to differ from that of fake news [7]. For example, deceivers are likely to use more sentiment words, more sense-based words (e.g., seeing, touching), and other-oriented pronouns, but less self-oriented pronouns.

Compared to real news, fake news shows lower cognitive complexity and uses more negative emotion words. However, the linguistic indicators of fake news across different topics and media platforms are not well understood. Rubin, points out that there are many types of fake news, each with different potential textual indicators [5]. This indicates that using linguistic features is not only laborious but also topic/media dependent domain knowledge, thus limiting the scalability of these solutions.

In addition to lexical features, speaker profile information can be useful [8]. Speaker profiles, including party affiliations, job title of speaker, as well as topical information which can also be used to indicate the credibility of a piece of news. For the study use profile information, [9] proposes a hybrid CNN model to detect fake news, which uses speaker profiles as a part of the input data. The Long-Short memory network (LSTM), as a neural network model, is proven to work better for long sentences [10]. Attention models are also proposed to weigh the importance of different words in their context. Current attention models are either based on local semantic attentions [11] or user attentions [12].

The difficulty comes partly from the fact that even human beings may have difficulty identifying between real news and fake news. Even worse, young people are not aware of the information they receive and replicate. This is the main reason of using User Center Design in order to identify the real experience in our target user.

For the development of this study, the User Experience (UX) will be considered as a core point in its elaboration, in order that the product really covers the specific needs of the user and understands their interaction with the environment.

3 User Research Methods

The UX process consists of several stages, in which the main objective is to identify the problem to be treated.

The first stage was to generate a brainstorm that would lead to a theme that represents a real problem in the environment. The result was the theme of Fake News.

The second stage consisted of using the "Persona" methodology, which is a very useful method to define the different user profiles for the study [13]. After analyzing the different profiles, we decided to analyze young people (16-20-year-old), since young people are the largest consumers of Social Networks, where the Fake News usually circulates (Fig. 1).



Fig. 1. Selected user based on “Persona” methodology.

It was important to collect information directly from the user, for this step some interviews were conducted [5]. In the first phase some interviews were applied in one group (16-20 years old) with different educational levels.

With this interviews, individuals were questioned about their Internet consumption and how they search for information and specially how they react with news. Through these interviews, relevant data were found such as:

- Young people trust, especially in news with a character of urgent (natural disasters, for example) in traditional media, especially on television and radio.

After that, some surveys were conducted in order to deepen the information obtained. With this surveys we discover the capacity of young people to identify Fake News and the reliability they have in the information they find in their day to day on social networks and websites. Similarly, it was verified if the user relied on traditional media and what verification strategy they used to identify fake news.

From the data obtained in the surveys, a series of premises similar to those of the interviews was obtained, but with a greater degree of precision, such as:

- Young people use the Internet not only for entertainment, it is also one of their main sources to conduct school researches and a way to obtain information of all kinds (tutorials, films, images, etc.).

- They do not trust completely in what they find in social networks, they have many doubts and prefer to check what was discovered in another type of media (mainly traditional).
- They try to consult multiple sources. They do not usually share information immediately, many of them prefer to verify the veracity of it before doing so.
- They tend to give importance only to the things they like or get their attention.

Finally, the needs of the users were analyzed and our target user was identified. These young people of school age were selected due to their high Internet consumption for their various tasks and their adaptation to technology. In addition, they have had greater contact with social networks.

Once knowing our user, we had to check if the findings were true, that is why the following study was designed.

The study

Once the users were identified and knowing their consumption, an interactive prototype was used to verify the true reaction of the users when they interact with a fake news (see Fig. 2).

The prototype developed is an interactive application where the user interacts and try to identify some fake news from the good ones. It can be seen like an interactive game because the user has a reduced time to decide if a news is fake or not. [4].

The interactive prototype was originally planned to have two modes, one individual and one in a competition of maximum four people.

In the first case, the individual starts almost immediately and interact with the prototype. In the first step, the instructions are given. Then, the interaction starts with five questions that were classified as intermediate (some news that are not so easy to identify if they are fake or not), once those five questions have been answered the prototype detects the level of expertise of the participant and continues with five more questions (either of a lower or higher difficulty) in order to have a complete round of ten questions.



Fig. 2. Example of the news displayed.

Finally, the prototype based on your answers, shows the level of real experience you have to identify fake news. Based on this level of experience, the prototype shows some simple recommendations to verify fake news (see Fig. 3).

In the case of the multiplayer, each one would have the opportunity to scan through a QR code an extension of the interface to run the prototype using their smartphone. In this case, the prototype screen would only serve as a dashboard that would show, in real-time, the questions, the response options, the progress of the participants, the highest and lowest scores and the problems of each user to identify fake news.

Both prototype modes would allow at the end to know the level that the user has to identify fake news, a feedback is shown in all cases with the recommendations strategies to identify fake news. Once this process is finished, an infographic is generated in order to print it or save it with the recommendations proposed by the prototype (see Fig. 4).

These recommendations [14] that are shown to the user can be summarized in:

- Consider the source. The user needs to investigate the site, its mission and its contact info.
- Read beyond. Headlines can be outrageous in an effort to get likes, the user need to read the whole story.
- Check the author. The user need to verify the author and their credibility.
- Supporting sources. The user need to determine if the info given actually supports the story.
- Check the date. Reposting old news stories doesn't mean they are relevant to current events
- Is it a joke? The user needs to research the site and author.
- Ask the experts. Ask a professor or a librarian.



Fig. 3. First feedback of the prototype

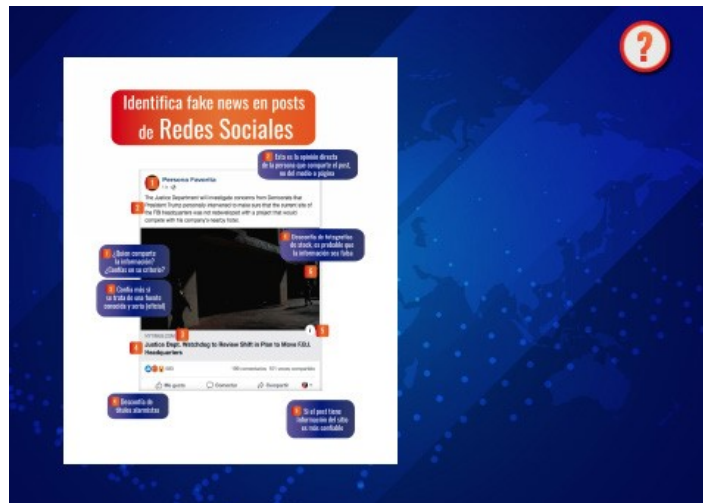


Fig. 4. Infographics generated as a result of the interaction with the prototype

User Testing and evaluation

For this phase, two evaluation studies were performed. First a set of heuristic evaluations were performed, where the viability of the prototype was verified. This

study was conducted by 30 users between 16 and 20 years old and the items that were verified were: simplicity, consistency, feedback, affordance, flexibility, perceptibility and ease of use.

After that a real user testing was performed. The study was conducted with 5 real users, the result was recorded in order to be analyze and improve the prototype (Fig. 5).

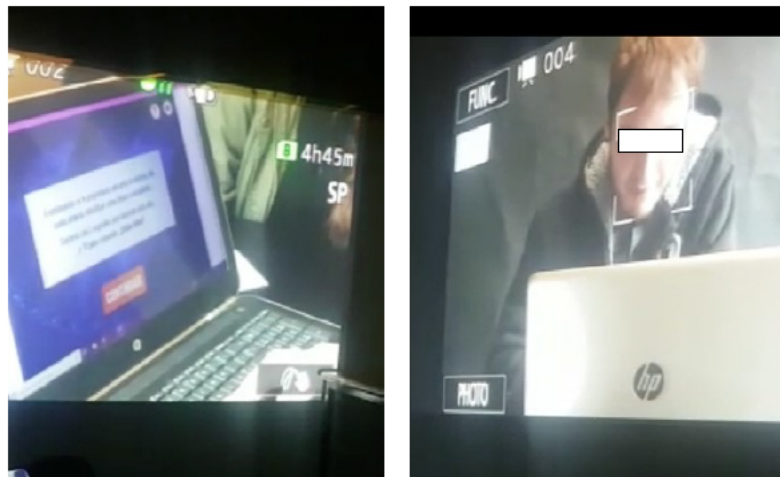


Fig. 5. User testing

Conclusions

This paper proposes a prototype that help the user to spot fake news. This is a complete methodology (User Center Design) to propose a solution of the real problem. The project starts with the user identification and its problematic. Then a prototype implementation process is made. Finally, a complete evaluation was conducted in order to verify the viability of the proposal.

Young people with their great interaction with technology and in particular with social networks, must have tools that help them make better decisions. Fake news circulates freely and daily on social networks and it is important to make young people aware that not all news we receive is real and that we must learn to identify fake news. The prototype proposal tries to raise awareness about this problem so that young users can navigate safely and reliably on the networks.

References

1. Oxford dictionaries, website,
2. <https://www.oxforddictionaries.com/press/news/2016/12/11/WOTY-16>. 2017.
3. Collins dictionary, website, <https://www.collinsdictionary.com/woty>. 2017.
4. Conroy, N. J., Rubin, V. L. and Chen, Y.: Automatic deception detection: Methods for finding fake news. *Proc. Assoc. Info. Sci. Tech.*, (2015) 52: 1-4. DOI:10.1002/pra2.2015.145052010082
5. Allcott, H., Gentzkow, M.: Social Media and Fake News in the 2016 election. *Journal of Economic Perspectives*. 31 (2015) (2): 211–236. doi:10.1257/jep.31.2.211.
6. Rubin, V., Chen, Y., Conroy, N.: Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, 52(1): (2015) 1–4.
7. Ruchansky, N., Atali, S., Liu, Y.: Csi: A hybrid deep model for fake news. arXiv preprint arXiv:1703.06959 (2017).
8. Larcker, D., Zakolyukina, A., Detecting deceptive discussions in conference calls. *Journal of Accounting Research*, 50(2) (2012):495–540.
9. Long, Y., Xiao, Y., Li, M., Huang, H.: Domain-specific user preference prediction based on multiple user activities. In *Big Data (Big Data)*, IEEE International Conference on, pages 3913–3921. IEEE (2016).
10. Wang, W., “Liar, liar pants on fire”: A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648 (2017).
11. Tang, D., Qin, B., Liu, T.: Document modeling with gated recurrent neural network for sentiment classification. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1422–1432, (2015).
12. Yang, Z., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2016).
13. Chen, H., Tu, C., Lin, Y., Liu, Z.: Neural sentiment classification with user and product attention. *EMNLP* (2016).
14. Nielsen, L. *Engaging Personas and Narrative Scenarios*. Samfundslitteratur, PhD-Series. (2004)
15. Gibson, C., Jacobson, T.: Informing and Extending the Draft ACRL Information Literacy Framework for Higher Education: An Overview and Avenues for Research. *College and Research Libraries* 75, no. 3 (2014): 250–4.