**Instituto Tecnológico y de Estudios Superiores de Occidente**

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación el 29 de noviembre de 1976.

Departamento de Electrónica, Sistemas e Informática

Maestría en Sistemas Computacionales



**ANÁLISIS Y VISUALIZACIÓN DE GRAFOS USANDO REALIDAD VIRTUAL**

Tesis Que para obtener el grado de

Maestro en Sistemas Computacionales

Presenta: David Alfonso Velasco Sedano

Asesor: Dr. Luis Fernando Gutiérrez Preciado

Tlaquepaque, Jalisco. 30 de Noviembre de 2017.

RESUMEN

Los grafos son una representación visual de un conjunto de elementos dentro de un sistema. Se construye a partir de 2 elementos principales. Primero tenemos los llamados nodos, los cuales simbolizan miembros dentro del sistema (como un rúter dentro de un sistema de conexión de red tipo WAN). El segundo elemento se le denomina como arista, el cual representa una conexión directa entre 2 nodos (este puede ser a sí mismo en algunos casos).

Utilizando está representación podemos extraer información del sistema. Está siendo, y no limitada, a distancias y caminos entre diferentes elementos. Ver que tan robusto es el sistema y su conexión. Podemos entender el costo de atravesar ciertos caminos y cuáles son las rutas óptimas para alcanzar nuestro objetivo a partir de un punto inicial deseado.

Actualmente se encuentran herramientas (tal como Gephi) que permiten desplegar y hacer análisis de los grafos, aunque estos se encuentran representados en ambientes bidimensionales. Durante la investigación se va gestionar una solución para la visualización de un grafo en ambiente virtual tridimensional.

Al inicio el trabajo se estará enfocando a la solución únicamente para la red social llamada Twitter. Aunque se estará planeando tener un alcance también académico y para investigaciones. Generando rutas para poder alcanzar mejores recursos relacionados al tema de investigación, qué tan relevante es la investigación y colaboradores/autores trabajando en tópicos similares. Tomando como base de datos inicial el Repositorio Institucional del ITESO (REI).

TABLA DE CONTENIDO

[1. INTRODUCCIÓN 7](#_Toc499634433)

[1.1. Antecedentes 7](#_Toc499634434)

[1.2. Justificación 7](#_Toc499634435)

[1.3. Problema 7](#_Toc499634436)

[1.4. Objetivos 8](#_Toc499634437)

[1.4.1. Objetivo General: 8](#_Toc499634438)

[1.4.2. Objetivos Específicos: 8](#_Toc499634439)

[1.5. Mercado meta 8](#_Toc499634440)

[1.6. Hábitos de consumo actuales 9](#_Toc499634441)

[1.7. Características funcionales 9](#_Toc499634442)

[1.8. Novedad científica, tecnológica o aportación 10](#_Toc499634443)

[2. ESTADO DEL ARTE o de la TÉCNICA 11](#_Toc499634444)

[2.1. Virtual Reality Network Visualization 11](#_Toc499634445)

[2.2. Un acercamiento inmersivo a la exploración visual geoespacial de base de datos de redes 11](#_Toc499634446)

[3. MARCO TEÓRICO/CONCEPTUAL 12](#_Toc499634447)

[3.1. Grafo 12](#_Toc499634448)

[3.2. Framework 12](#_Toc499634449)

[4. Plan del proyecto TOG 13](#_Toc499634450)

[4.1. Alcance 13](#_Toc499634451)

[4.1.1. Estructura del desglose del trabajo (WBS) 13](#_Toc499634452)

[4.1.2. Requerimientos 13](#_Toc499634453)

[4.2. Cronograma 17](#_Toc499634454)

[4.3. Riesgos 17](#_Toc499634455)

[4.3.1. Supuestos 18](#_Toc499634456)

[4.3.2. Elementos fuera de alcance 18](#_Toc499634457)

[4.3.3. Criterios de éxito 19](#_Toc499634458)

[4.4. Secciones adicionales (Costos, calidad, recursos humanos, riesgos, comunicaciones). 19](#_Toc499634459)

[5. CONCLUSIONES 20](#_Toc499634460)

[5.1. Conclusiones 20](#_Toc499634461)

[5.2. Trabajo Futuro 20](#_Toc499634462)

LISTA DE FIGURAS

[Figura 4.1 Estructura del desglose del trabajo (WBS) 16](#_Toc497321448)

[Figura 4.2 Cronograma 18](#_Toc497321449)

LISTA DE TABLAS

[Tabla 4.1 Tabla de Requerimientos 17](#_Toc499634417)

[Tabla 4.2 Tabla de riesgos 18](#_Toc499634418)

[Tabla 4.3 Tabla de supuestos 18](#_Toc499634419)

[Tabla 4.4 Tabla de elementos fuera de alcance 19](#_Toc499634420)

[Tabla 4.5 Tabla de criterios de éxito 19](#_Toc499634421)

LISTA DE ACRÓNIMOS Y ABREVIATURAS

|  |  |  |
| --- | --- | --- |
| REI |  | Repositorio Institucional del ITESO |
| ITESO |  | Instituto Tecnológico y de Estudios Superiores de Occidente |
|  |  |  |
|  |  |  |

1. INTRODUCCIÓN
   1. Antecedentes

El estado actual del arte con respecto al despliegue de cualquier grafo en un ambiente virtual tridimensional está limitado a proyectos estudiantiles. Mucho son prototipos que utilizan redes sociales como su fuente de información. Estos ligan al despliegue de los mensajes enviados por usuarios al igual que la creación de caminos para alcanzar a cualquier individuo en una red social.

También se encuentran herramientas de modelado de grafo y su interpretación. Aunque por el momento estas están limitadas a un ambiente bidimensional. Para mencionar algunas de estas, podemos encontrar a Graphviz [1] y Gephi [2]. Aunque estos carecen de un despliegue del grafo de forma tridimensional.

* 1. Justificación

Se debe hacer más eficiente la manera en que se consigue, analiza y se despliega información basado en grafos. Una representación de esto es en las redes sociales. Debe de generarse una forma de fácil acceso, digerir y entregar información relacionado a un usuario o un concepto.

* 1. Problema

Las redes sociales vistas como un grafo tienen consigo un gran desafío en su tiempo de ejecución y análisis. Esto a su vez conlleva al desafió de poder resolver la petición hecha por el usuario en un tiempo eficiente y entregar un resultado lo más cercano posible a lo deseado.

A la hora de hacer investigaciones (por ejemplo, aquellas relacionadas al área de mercadotecnia), se dispone de mucha información, aunque esta se entrega de una manera pobre y difícil de digerir. Lo cual conlleva a múltiples iteraciones de filtrado del contenido hasta obtener la información que se requería desde la primera consulta.

* 1. Objetivos
     1. Objetivo General:

Se estará desarrollando lo que se denomina un framework, el cual es una solución pre-hecha y lista para ser usada por otros desarrolladores. Un framework se puede ver como una solución completa o semi-completa de software de la cual puedes utilizar sus funciones y atributos de acceso público. En este escenario, nuestros elementos de acceso público estarán enfocados al filtrado y la generación de grafos tridimensionales. Aunque se puede extender a tener implementaciones abstractas las cuales el usuario final (como un investigador) puede definir para obtener una mejor respuesta basada en sus necesidades.

Dentro de nuestro escenario, se va a construir el framework especializado en la red social Twitter [3]. La cual debe tener la capacidad de generar los grafos tridimensionales en base a la meta-data de los comentarios, basado en las imágenes, entre otras opciones. Las opciones serán los requerimientos del usuario entregados a partir de un conjunto de filtros.

* + 1. Objetivos Específicos:

De forma colectiva se espera, y no está limitado, a los siguientes objetivos:

* Hacer una representación 3D del grafo usando datos obtenidos de la página Twitter. La complejidad se encuentra en la evaluación y lógica para distribuir en un espacio virtual los nodos de una manera óptima.
* Generar y gestionar filtros. Estos se deben representar como sub-grafos.
* Se debe disponer de una lógica de identificación de nodos principales. A su vez, estos se deben priorizar en su despliegue.
* Tener un sistema funcional de navegación del ambiente virtual donde el usuario pueda explorar.

De manera personal, se espera alcanzar por lo menos los siguientes objetivos:

* Expandir mi conocimiento de generación de ambientes tridimensionales usando Unity.
* Generar conocimiento de cómo proponer y crear algoritmos eficientes para solucionar problemas cuya respuesta tiene una representación visual.
* Ampliar mi conocimiento del análisis de grafos en el escenario de redes sociales, usando meta-datos como nodos y la probabilidad de éxito en tener un sub-siguiente meta-dato o imagen asociado a esté como arista.
  1. Mercado meta

El trabajo está dirigido a cualquier mercadólogo, sociólogo, psicólogo u otra persona que puede ser beneficiada con el análisis y compresión del comportamiento de redes sociales. Debe estar sujeto a que estará extrayendo información de manera masiva. Actualmente se cuenta con un conjunto de sociólogos y psicólogos fuertemente interesados en el tema.

En un futuro se planea que la aplicación tenga una especialización para el ambiente político. Siendo sus principales consumidores abogados, periodistas, investigadores, y cualquier persona interesada en este tema.

Por igual se desarrollará una extensión para personas dentro del gremio académico e investigadores. Donde el mayor enfoque sea la representación de trabajos relacionados a los temas que domina el usuario o que desea expandir.

* 1. Hábitos de consumo actuales

El estado actual del mercado en su mayoría está dominado por herramientas de visualización de grafos en dos dimensiones. Tal es el caso de las herramientas Gephi [2], Graph-Tool [4] y Graphviz [1]. Muchas de estas herramientas piden que de manera manual que el usuario meta la información del cual el análisis se basará en.

Mientras tanto, podemos utilizar software similar para el análisis de una red social. Usando de ejemplo a Graph-Tool [4] y AllegroGraph [5]. Aunque siguen pidiendo que el usuario meta la información inicial para generar los grafos. Estas siguen estando limitadas en su mayoría en una visualización en dos dimensiones.

Finalmente, páginas de redes sociales tal como Facebook, Twitter y YouTube ya tienen soluciones que permiten el análisis de la red social. Su limitante es que no se despliegan como grafos al momento de escribir este documento. Sin mencionar que simplemente otorgar información de su propio sitio y no la relación entre plataformas.

* 1. Características funcionales

El sistema en su inicio va funcionar utilizando el dispositivo llamado HTC Vive. Este dispositivo permite interactuar y ver ambientes de realidad virtual. La solución se va dividir en 3 secciones principales.

Primero se encuentra la interfaz e interacción del usuario-máquina. En esta capa debemos ser capaces de interpretar las interacciones del usuario. Esto siendo cualquier entrada otorgada por él vía una interfaz gráfica (tal es el caso de los filtros deseados).

Después, en la segunda capa, se estará enfocando al despliegue del grafo mismo. Esto conlleva al cálculo de distancias “físicas” entre nodos. La forma en que vamos a colorear nodos. Al igual que aquí podemos crear las nubes de información relacionada.

Finalmente, nos vamos a encontrar con la capa del análisis y creación del grafo. Como el nombre indica, debe ser capaz de recibir los criterios obtenidos del usuario y encontrar la forma más eficiente de generar el grafo mismo. En ocasiones, deberá ser capaz de gestionar los sub-grafos y decir que nodos tienen mayor prioridad.

En esta investigación nos vamos a enfocar en la última sección. Se estará planteando distintos algoritmos y funcionalidades para la gestión de datos. Se busca hacer un comparativo entre diferentes soluciones propuestas con la nuestra. También haremos el esquemático de nuestro framework.

* 1. Novedad científica, tecnológica o aportación

Primero se estaría entregando una herramienta que su mayor enfoque es la visualización del grafo de manera virtual tridimensional. Donde el usuario pueda caminar entre y actualizar de acorde a sus necesidades.

Después tenemos la capacidad de generar el grafo en base a información dinámica que el usuario no mete de forma manual o que esté limitado a formatos. Esto ayuda a generar panoramas recientes y puntuales de cualquier tópico a analizar.

1. ESTADO DEL ARTE o de la TÉCNICA
   1. Virtual Reality Network Visualization

En este documento se detalla la problemática de navegar en un ambiente virtual utilizando el teclado y ratón. Se propone como usando un equipo especializado para estos ambientes mitiga mucho el conflicto. Haciendo más intuitivo la navegación. [6]

* 1. Un acercamiento inmersivo a la exploración visual geoespacial de base de datos de redes

La investigación gira alrededor de las ventajas de ver la representación de un grafo vía un espacio virtual en vez del tradicional espacio bidimensional. Se menciona la curvatura y proyección de los elementos a desplegar. Haciendo referencia las distancias claves e ideales entre el usuario y el objeto a explorar. También se toca el tema de cómo el ojo va a ir observando y escaneando el área virtual. [7]

1. MARCO TEÓRICO/CONCEPTUAL
   1. Grafo

Los grafos son una representación en forma de gráfica de problemas matemáticos. Estos tienden a representan objetos y lugares en la vida real, aunque también pueden representar conceptos abstractos. Se construye a partir de nodos y aristas.

Llamamos nodos a cualquier elemento dentro del sistema que estamos graficando. Pueden representar ciudades y personas hasta elementos como páginas web, rúters, entre otros. Los nodos se conectan a través de aristas. En algunos escenarios las conexiones pueden tener peso y/o dirección, de acuerdo sea el caso. [8]

* 1. Framework

Se considera un framework a cualquier solución de software que se puede re-usar y es considerado semi-completo. Se le denomina semi-completo a cualquier aplicación o implementación que todavía tiene métodos que se le necesita definir su funcionalidad.

Es esperado que un framework tenga una colección de métodos orientados a solucionar un solo problema. Aunque, por la definición de ser semi-completo, nos da un pequeño margen de cómo se debe solucionar el problema. [9]

1. Plan del proyecto TOG
   1. Alcance
      1. Estructura del desglose del trabajo (WBS)

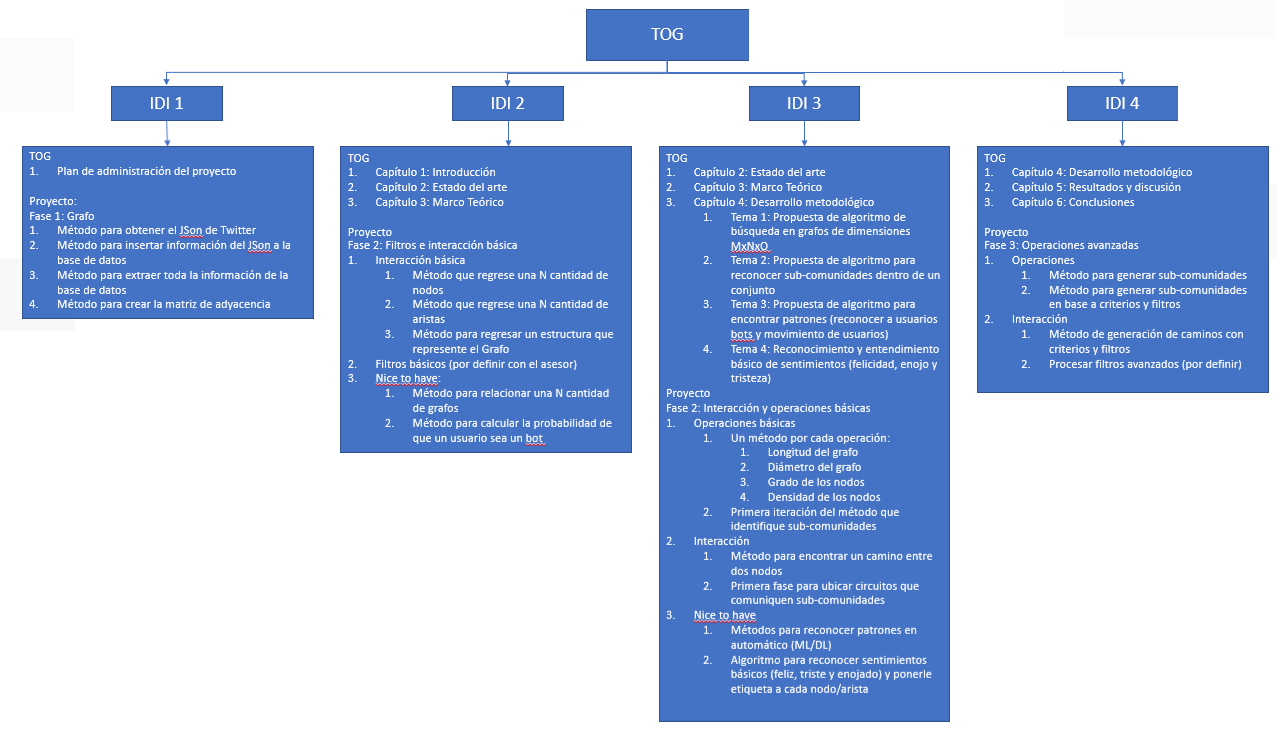


Figura 4.1 Estructura del desglose del trabajo (WBS)

* + 1. Requerimientos

|  |  |  |
| --- | --- | --- |
| ID | Requerimientos | Entregable |
| REQ01 | El diseño UML por lo menos debe incluir el diagrama de caso de usos. Diagrama de actividad para el proceso de convertir   el archivo Json a un grafo y el proceso de filtrar la información. Debe haber el diagrama de clases de la estructura general | Fase 1: Grafo Fase 2: Filtros e interaccion básica Fase 3: Interacción operaciones básicas Fase 4: Operaciones Avanzadas |
| REQ02 | Cualquier propuesta hecha debe incluir el algoritmo o la solución que actualmente existe. Una comparativa entre los 2 y enseñar el porqué es más eficiente el algoritmo propuesto | Top of Form  \* Método para relacionar una N cantidad de grafos​ \* Método para calcular la probabilidad de que un usuario sea un bot \* Métodos para reconocer patrones en automático (ML/DL)​ \* Algoritmo para reconocer sentimientos básicos (feliz, triste y enojado) y ponerle etiqueta a cada nodo/arista \* Método de generación de caminos con criterios y filtrosBottom of Form |
| REQ03 | Cualquier implementación de código debe seguir el patrón de diseño MVC. Toda función pública debe tener un valor de retorno, por default debe ser un booleano | \* Método para obtener el JSon de Twitter​ \* Método para insertar información del JSon a la base de datos​ \* Método para extraer toda la información de la base de datos​ \* Método para crear la matriz de adyacencia \* Método que regrese una N cantidad de nodos​ \* Método que regrese una N cantidad de aristas​ \* Método para regresar un estructura que represente el Grafo \* Métodos que realizan operaciones matemáticas \* Método para generar sub-comunidades​ \* Método para generar sub-comunidades en base a criterios y filtros \* Método para relacionar una N cantidad de grafos​ \* Método para calcular la probabilidad de que un usuario sea un bot \* Métodos para reconocer patrones en automático (ML/DL)​ \* Algoritmo para reconocer sentimientos básicos (feliz, triste y enojado) y ponerle etiqueta a cada nodo/arista \* Método de generación de caminos con criterios y filtros |
| REQ04 | Todo método debe tener por lo menos 4 casos de pruebas. De esos dos deben ser para escenarios positivos y otros dos para el caso negativo. | \* Método para obtener el JSon de Twitter​ \* Método para insertar información del JSon a la base de datos​ \* Método para extraer toda la información de la base de datos​ \* Método para crear la matriz de adyacencia \* Método que regrese una N cantidad de nodos​ \* Método que regrese una N cantidad de aristas​ \* Método para regresar un estructura que represente el Grafo \* Métodos que realizan operaciones matemáticas \* Método para generar sub-comunidades​ \* Método para generar sub-comunidades en base a criterios y filtros \* Método para relacionar una N cantidad de grafos​c \* Método para calcular la probabilidad de que un usuario sea un bot \* Métodos para reconocer patrones en automático (ML/DL)​ \* Algoritmo para reconocer sentimientos básicos (feliz, triste y enojado) y ponerle etiqueta a cada nodo/arista \* Método de generación de caminos con criterios y filtros |
| REQ05 | Todo tema expuesto en el desarrollo metodológico, en los casos de explicación de algoritmo, debe haber una explicación sobre sus duraciones y un análisis de su complejidad espacial y temporal | \* Tema 1: Propuesta de algoritmo de búsqueda en grafos de dimensiones MxNxO​ \* Tema 2: Propuesta de algoritmo para reconocer sub-comunidades dentro de un conjunto​ \* Tema 3: Propuesta de algoritmo para encontrar patrones (reconocer a usuarios bots y movimiento de usuarios)​ \* Tema 4: Reconocimiento y entendimiento básico de sentimientos (felicidad, enojo y tristeza) |
| REQ06 | Los métodos relacionados al procesamiento de filtros deben regresar los valores en formas de arreglos o listas | \* Método de generación de caminos con criterios y filtros \* Método para generar sub-comunidades en base a criterios y filtros |

Tabla 4.1 Tabla de Requerimientos

* 1. Cronograma



Figura 4.2 Cronograma

* 1. Riesgos

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Identificación | | Análisis cualitativo | | | Mitigación | |
| ID | Riesgo | Probabilidad | Impacto | Descripción del Impacto | Estrategia de respuesta | |
| R1 | Dejar la maestría | Media | Alto | Dejar de darle continuidad a la maestría | Documentar todo el código y poner cualquier referencia utilizada durante lo que estuve. | |
| R2 | Cancelación del proyecto | Baja | Alto | Nuestro cliente y asesor cancele el proyecto antes de terminar la maestría | Asegurar que los involucrados firmen un contrato donde todos están de acuerdo que el proyecto por lo menos dure 2 años (la duración de la maestría) | |
| R3 | El API utilizado para comunicar nuestro sistema con la base de datos de Neo4j deje de ser soportada | Baja | Baja | Ya no sea posible hacer conexiones ni interacciones con la base de datos | Conseguir la última versión estable del API y mantenernos con ella hasta la finalización del proyecto | |
| R4 | Twitter modifique su API y regrese valores distintos a los establecidos bajo su documentación | Baja | Baja | La forma en que se obtiene datos se vuelve obsoleta | Asegurar que los métodos que consigan los valores sean lo más genéricos posibles para que su modificaciones sean las menos posibles | |
| R5 | Compañeros de equipos se salgan del proyecto antes de su finalización | Baja | Alto | Los entregables no se lograran a tiempo. Generará una sobre-carga a los miembros sobrantes | Tener una sesión de transferencia antes de su partida para que sus conocimientos e ideas que tenía no se pierdan. De ahí, actualizar el cronograma junto con WBS re-considerando que cosas son posibles lograr dentro del tiempo | |
| R6 | Me retrase con la implementación de mis funciones públicas para el resto del equipo | Baja | Media | Funcionalidades pre-establecidas no se podrán probar ni continuar con su desarrollo si dependen de las funciones públicas | | Asegurar tener por lo menos valores de prueba que se pueden regresar mientras la funcionalidad se completa. De esta forma, todos pueden continuar sin problema. |

Tabla 4.2 Tabla de riesgos

* + 1. Supuestos

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Supuesto | Validado por: | Fecha de validación | Impacto de no ser valido |
| S1 | El proyecto simplemente involucra el HTC Vive | Contacto del Signa-lab y asesor | 11 de Septiembre de 2017 | Se debe considerar cómo hacer la visualización entre distintas plataformas. |
| S2 | Las actividades están planeadas para ser realizadas dentro de un año y medio | Asesor | 11 de Septiembre de 2017 | Se tendrá que decidir que funcionalidades se harán en nuestra estadía y cuáles se pasarán a otro grupo de estudiantes |

Tabla 4. Tabla de supuestos

* + 1. Elementos fuera de alcance

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ID | Descripción | Validado por: | Fecha de validación | Razón de estar fuera de alcance |
| OS1 | No se va desarrollar para funcionar en el Oculus Rift. | Asesor | 11 de Septiembre de 2017 | Solo tenemos a la mano el HTC Vive. |

Tabla 4. Tabla de elementos fuera de alcance

* + 1. Criterios de éxito

|  |  |  |
| --- | --- | --- |
| ID | Descripción | ¿Cómo se validará? |
| CE1 | El tiempo de ejecución para la obtención y generación del grafo debe ser que los tiempos que maneja el cliente actualmente (por lo menos disminuir los tiempo en 10%) | Se hará una corrida en paralelo de los dos sistemas. Haciendo búsquedas aleatorias con diferente volumen de datos |
| CE2 | El sistema debe manejar un volumen de descarga de tweets de 200mil búsquedas | El tiempo de duración debe ser menor a 5 horas y 10% más eficiente que el del usuario |

Tabla 4. Tabla de criterios de éxito

* 1. Secciones adicionales (Costos, calidad, recursos humanos, riesgos, comunicaciones).

1. CONCLUSIONES
   1. Conclusiones

[Las conclusiones deben responde a los objetivos establecidos]

* 1. Trabajo Futuro

[Se refiere a recomendaciones o descripciones sobre líneas de investigación que abre este trabajo, aplicaciones inmediatas que se derivan, o desarrollo de componentes o extensiones del desarrollo.]

BIBLIOGRAFÍA

|  |  |
| --- | --- |
| [1] | J. Ellson, E. Gansner, E. Koutsofios, S. North y G. Woodhull, «https://www2.graphviz.org,» 2003. [En línea]. Available: https://graphviz.gitlab.io/\_pages/Documentation/EGKNW03.pdf. [Último acceso: 13 Noviembre 2017]. |
| [2] | D. Khokhar, «Gephi Cookbook,» de *Gephi Cookbook*, Birmingham, Packt Publishing Ltd, 2015, p. 296. |
| [3] | P. Gil, «Lifewire,» 8 Noviembre 2017. [En línea]. Available: https://www.lifewire.com/what-exactly-is-twitter-2483331. [Último acceso: 2017 Noviembre 2017]. |
| [4] | T. d. P. Peixoto, «Graph-Tool,» 9 Noviembre 2017. [En línea]. Available: https://graph-tool.skewed.de/static/doc/demos/index.html. [Último acceso: 2017 Noviembre 2017]. |
| [5] | Franz Incorporated, «Franz,» 3 Noviembre 2017. [En línea]. Available: https://franz.com/agraph/support/documentation/current/agraph-introduction.html#ai-overview. [Último acceso: 13 Noviembre 2017]. |
| [6] | L. Beisel, P. Lamouric, J. C. Mayer y H. Dinh, «LUCA BEISEL,» Diciembre 2016. [En línea]. Available: http://lucabeisel.de/vr-network-visualization/. [Último acceso: 30 Octubre 2017]. |
| [7] | M.-J. Zhang y K. Zhang, «An Immersive Approach to the Visual Exploration of GeospatialNetwork Datasets,» de *CCS Concepts: Human-centered computing ~ Virtual reality;Human-centered computing ~ Geographic visualization*, Zhuhai, China, 2016. |
| [8] | D. P. Kumar, S. Archana y B. T. Kumar, «Graph Theory in an Object Oriented Approach,» *Journal of Computer Sciences and Applications,* vol. 3, nº 6, pp. 123-126, 2015. |
| [9] | D. C. Schmidt, «Applying Patterns and Frameworks to Develop,» de *The Handbook of Programming Languages*, St. Louis, Macmillan Technical Pub, 1997, p. 250. |

**APÉNDICE A. Título**