

**MÉTRICAS PARA LA EVALUACIÓN DE CALIDAD DE LA INFORMACIÓN  
EN REDES SOCIALES VIRTUALES EN SALUD**



**Daniel Eduardo Cañón**

**Universidad del Cauca  
Facultad de Ingeniería Electrónica y Telecomunicaciones  
Departamento de Telemática  
Línea de Investigación en e-Salud  
Noviembre de 2013  
Popayán**

**MÉTRICAS PARA LA EVALUACIÓN DE CALIDAD DE LA INFORMACIÓN  
EN REDES SOCIALES VIRTUALES EN SALUD**



**Daniel Eduardo Cañón**

**Director: PhD. Ing. Diego Mauricio López**

**Universidad del Cauca  
Facultad de Ingeniería Electrónica y Telecomunicaciones  
Departamento de Telemática  
Línea de Investigación en e-Salud  
Noviembre 2013**

**A mi padre y mis hermanos**

## **AGRADECIMIENTO**

Quiero dedicar este trabajo de grado a mi padre Fabio Cañón Moya, a mis hermanos Fabio Alberto Cañón y Juan Manuel Cañón, ya que ellos son mi motor, mi motivación y cualquier palabra es insuficiente para expresar lo importantes que son para mí.

Al PhD. Ing. Diego Mauricio López, por su voto de confianza, su dedicación, y transmitir su conocimiento.

Al grupo G.I.G.H.A. de la universidad del Cauca, en especial a la profesora Rosa Álvarez, a Carolina Ortega y Alejandra Padilla, por aceptarme dentro de los proyecto y ayudarme a crecer no solo en lo académico, sino también como persona.

A mis amigos y en especial a: Ana María Duque, Cristina Pórtela, Alejandra González, Jessica Ñañez, Jesús Muñoz, Luis Eduardo Bastidas, Daniel Chavarro y Oscar Orozco, por hacer que estos años dentro y fuera de la universidad inolvidables.

## Contenido

<b>Introducción .....</b>	<b>11</b>
1.1 Introducción.....	11
1.2 Planteamiento del problema.....	11
1.3 Conceptos Básicos.....	13
1.4 Objetivos .....	15
1.4.1 Objetivo General .....	15
1.4.2 Objetivos Específicos .....	15
1.5 Organización del documento.....	15
<b>Síntesis de la base del conocimiento asociada a atributos y métricas de calidad en la web 2.0.....</b>	<b>17</b>
2.1. Introducción.....	17
2.2. Aproximación Metodológica .....	17
2.3. Análisis del estado del arte.....	18
2.3.1. Fase Preparatoria.....	18
2.3.1.1. Declaración del Alcance de la base del conocimiento:.....	18
2.3.1.2 Identificación de los Núcleos Temáticos de la base de conocimiento. ....	19
2.3.2. Fase Descriptiva.....	19
2.3.2.1. Conformación del Banco de Unidades de Análisis.....	19
2.3.2.2. Construcción de las Fichas Descriptoras .....	24
2.3.3. Fase de Interpretación por núcleos temáticos.....	27
2.3.3.1. Resumen de los parámetros de calidad obtenidos por medios de las fichas descriptoras.....	28
2.3.3.2. Aportes, vacíos y limitaciones del NT “Criterios y Métricas” correspondiente a la Unidad de Análisis Gestión de la calidad de la información en RSVS. ....	36
2.3.3.3. Tabla resumen de servicios, aplicaciones, aplicaciones Social Media(SM) implementaciones, y herramientas.....	41
2.3.3.4. Aportes, vacíos y limitaciones del NT “Servicios y Aplicaciones” correspondiente a la Unidad de Análisis Gestión de la calidad de la información en RSVS. ....	41
2.3.4. Fase de Construcción teórica global. ....	44



3.3.4.3.1 Requisitos de información.....	64
3.3.5 Especificación.....	65
3.3.5.1 Objetivos.....	65
3.3.5.2 Actividades y técnicas de representación:.....	65
3.4 Análisis y Diseño.....	66
3.4.1 Prototipo.....	66
3.4.2 MODULO HEALTHTRUST.....	67
3.5 Implementación.....	69
3.5.1 Interfaces web.....	69
3.5.2 PRUEBAS.....	76
3.5.2.1 Primer escenario de pruebas usando el prototipo de HealthTrust.....	76
3.5.2.2 Segundo escenario de pruebas usando el prototipo de HealthTrust.....	77
3.6 Conclusiones.....	80
<b>Evaluación del servicio.....</b>	<b>81</b>
4.1 Introducción.....	81
4.2 DESMET.....	81
4.3 Tipos de procesos de evaluación.....	81
4.4 Organización de la evaluación.....	81
4.5 Selección del método de evaluación.....	82
4.6 Análisis de Efectos Cualitativos-Opinión de expertos Intercalados.....	84
4.7 Escala de tiempo de la evaluación por parte de expertos.....	84
4.8 Desarrollo de la evaluación.....	84
4.8.1 Evaluación de la Relevancia.....	85
4.8.2 Evaluación de la Precisión.....	86
4.9 Resultados.....	86
4.9.1 Evaluación de Relevancia.....	86
4.9.2 Evaluación de la Precisión.....	94
<b>Conclusiones y trabajo futuro.....</b>	<b>96</b>
4.1 Conclusiones del Estado arte.....	96
4.2 Diseño e implementación de servicio.....	96

4.3 Conclusiones de la Evaluación .....	97
4.4 Trabajo futuro .....	97
<b>Referencias</b> .....	<b>97</b>

## **LISTA DE FIGURAS**

---

Figura 1. Modelo de organización jerárquico entre los diferentes parámetros de calidad .....	28
Figura 2. Criterios de calidad agrupados mediante dimensiones de calidad .....	38
Figura 3. Métricas de calidad para Wikipedia .....	47
Figura 4. Relaciones entre los contenidos y autores en una RSV .....	47
Figura 5. Hubs y Authorities en el contexto de blogs sobre diabetes.....	48
Figura 6. Implementación de HITS dentro de la librería JUNG .....	49
Figura 7. Ecuación de PageRank .....	50
Figura 8. Algoritmo del caminante aleatorio.....	50
Figura 9. Algoritmo del caminante aleatorio con probabilidad de salto .....	51
Figura 10. Modelado de HealthTrust para contenido .....	52
Figura 11. Modelado de HealthTrust para el autor.....	53
Figura 12. Algoritmo de la métrica HealthTrust.....	53
Figura 13. Aproximación esquema de AMIR-ST.....	57
Figura 14. Caso de uso del negocio, modelo inicial de publicación de videos videos .....	59
Figura 15. Caso de uso del negocio, modelo parcial para la publicación de videos .....	61
Figura 16. Modelo del dominio parcial para la publicación de videos .....	63
Figura 17. Caso de uso del negocio, modelo para la publicación de videos.....	64
Figura 18. Arquitectura lógica general para un sistema de moderación de videos en una RSVS .....	66
Figura 19. Arquitectura lógica módulo HealthTrust .....	67
Figura 20. Arquitectura de un servicio REST con JAX-RS.....	68
Figura 21. Patrón de diseño Façade.....	69
Figura 22. Interfaz web 1 servicio de búsqueda basado en métricas .....	70
Figura 23. Interfaz web 2 servicio de búsqueda basado en métricas .....	70



Figura 24. Interfaz web 3 servicio de búsqueda basado en métricas .....	71
Figura 25. Funcionalidad de la clase AbstractFacade .....	72
Figura 26. Funcionalidad de la clase VideoFacadeREST 1 .....	73
Figura 27. Funcionalidad de la clase VideoFacadeREST 2 .....	74
Figura 28. Modelo de la base de datos para la aplicación .....	75
Figura 29. Componente sintáctico para asistir a las búsquedas .....	75
Figura 30. Primera extracción de datos dentro del prototipo de HealthTrust .....	76
Figura 31. Grafo para la primera aproximación.....	77
Figura 32. Segunda extracción de datos dentro del prototipo de HealthTrust .....	78
Figura 33. Grafo para la segunda aproximación .....	79
Figura 34. Evaluación por medio de un grupo de discusión en la RSVS Unisalud.....	85
Figura 35. Base de datos referente a la evaluación en el programa SPSS .....	86
Figura 36. Calificación de relevancia para YouTube por parte del Evaluador 1.....	88
Figura 37. Calificación de relevancia para HealthTrust por parte del Evaluador 1.....	89
Figura 38. Relevancia absoluta del Evaluador 1, usando el servicio de búsqueda de YouTube .....	90
Figura 39. Relevancia absoluta del Evaluador 1, usando el servicio de búsqueda de HealthTrust .....	91
Figura 40. Calificación de Relevancia para YouTube por parte del Evaluador 2.....	92
Figura 41. Calificación de Relevancia para HealthTrust por parte del Evaluador 2.....	93
Figura 42: Relevancia absoluta del Evaluador 2, usando el servicio de búsqueda de YouTube .....	93
Figura 43: Relevancia absoluta del Evaluador 2, usando el servicio de búsqueda de HealthTrust .....	94

## **LISTA DE TABLAS**

---

Tabla 1. Discriminación de Núcleos Temáticos y fichas descriptoras.....	21
Tabla 2. Ficha representativa del Núcleo Temático Criterios y métricas.....	24
Tabla 3. Ficha representativa del Núcleo Temático Aplicaciones y servicios .....	26

Tabla 4. Dimensiones, criterios eEuro y principios HON, sub-criterios atributos y métricas .....	29
Tabla 5. Servicios, aplicaciones, implementaciones, y herramientas .....	41
Tabla 6. Estructura de las Fichas de construcción teórica global .....	44
Tabla 7. Atributo(s) y Métrica(s) preseleccionadas .....	44
Tabla 8. Atributo(s) y Métrica(s) seleccionadas .....	55
Tabla 9. Análisis correspondiente al caso de uso “Buscar video” .....	59
Tabla 10. Análisis correspondiente al caso de uso “Valorar video” .....	60
Tabla 11. Análisis correspondiente al caso de uso “Publicar video” .....	60
Tabla 12. Análisis correspondiente al caso de uso del servicio “Realizar búsqueda” .....	62
Tabla 13. Análisis correspondiente al caso de uso del servicio “Valorar videos” .....	62
Tabla 14. Análisis correspondiente al caso de uso del servicio “Publicar video” .....	62
Tabla 15. Funcionalidad de la clase VideoFacadeREST .....	72
Tabla 16. Tiempo de ejecución del escenario de pruebas 1 .....	77
Tabla 17. Tiempo de ejecución del escenario de pruebas 2 .....	80
Tabla 18. Evaluación de la selección del método DESMET .....	82
Tabla 19. Análisis descriptivo de los resultados.....	87
Tabla 20. Distribución de frecuencias para la variable relevancia del Evaluador 1 .....	88
Tabla 21. Distribución de frecuencias para la variable relevancia del Evaluador 1, con el valor de relevancia transformado de variable de escala, a variable en dos estados .....	89
Tabla 22. Distribución de frecuencias para la variable relevancia del Evaluador 2.....	91
Tabla 23. Distribución de frecuencias para la variable relevancia del Evaluador 2, con el valor de relevancia transformado de variable de escala, a variable en dos estados .....	92
Tabla 24. Evaluación de precisión de los servicios para al menos un revisor.....	94
Tabla 25. Evaluación de precisión de los servicios para ambos revisores.....	95

# Capítulo 1

## Introducción

### 1.1 Introducción

Las Redes Sociales Virtuales en Salud (RSVS) se han convertido en una nueva forma de promoción de la salud donde los pacientes pueden retroalimentar sus experiencias y compartir sus opiniones sobre la efectividad de medicamentos, tratamientos y brindar apoyo a personas que tienen su misma condición como pacientes. Pero, de igual forma, al traer nuevos beneficios también llevan consigo nuevos retos, como la moderación del contenido que ahora no solo es publicado y producido por expertos o profesionales en salud, sino también por los mismos usuarios o habituales consumidores de información.

La calidad de los contenidos de salud en internet, incluyendo los contenidos en las RSVS, es un tema de preocupación mundial, especialmente por el impacto negativo que puede causar en los consumidores finales de esta información. Ejemplo de ello son las múltiples campañas de empresas farmacéuticas por impulsar sus productos, campañas de curas milagrosas, sitios web pro-anorexia, y en general toda la desinformación que existe en la red.

### 1.2 Planteamiento del problema

En los últimos años, internet se ha convertido en una fuente importante para aquellas personas que buscan información en temas de salud, permitiendo que los ciudadanos se informen ampliando su conocimiento sobre enfermedades, sus cuidados, los medicamentos con sus respectivos tratamientos, estilos de vida y otra información de interés.

Estudios recientes sobre la eficiencia, la legibilidad y efectividad de la información en salud, han demostrado que tanto los medios de radiodifusión, los medios impresos y televisivos cuentan con niveles bajos de calidad de la información [1]. Internet y en especial las RSVS, que son el objeto de estudio de este trabajo de grado, no son la excepción. La gestión de la calidad de la información en las RSVS es compleja, dado que en general son los mismos usuarios, sin ser expertos en estos temas, los que generan y difunden contenido de interés en salud.

Toda esta avalancha de contenido y la naturaleza voluntaria y “viral” de las Redes Sociales Virtuales en Salud, hacen necesario que existan mecanismos de control de la calidad de la información que se publica. Las principales limitaciones tanto

de calidad y seguridad que se han identificado se pueden resumir como: a) la falta de concordancia del contenido informativo con las prácticas médicas b) la poca o ninguna relevancia que pueda llegar a tener esta información, que debe ser tomada en todos los casos como complementaria a la suministrada por el profesional de la salud, c) la posible falta de protección contra la desinformación y la publicidad engañosa, d) los problemas de accesibilidad que puedan tener los usuarios de redes sociales, e) la legibilidad dado que independientemente de la calidad de la información, esta debe ser comprensible por las personas que tienen acceso a ella, y f) falta de políticas de privacidad de la información.

En un trabajo recopilatorio realizado por A. Risk [2], se mencionan muchas iniciativas existentes que intentan generar una serie de mecanismos para garantizar la calidad de la información publicada en internet sobre salud. Las iniciativas mencionadas por Risk se agrupan en códigos de conducta, certificaciones por parte de terceros y evaluaciones basadas en herramientas. A excepción de “*Health on the Net*” HON [3], que ha hecho extensivo su código de conducta para Web 2.0, ninguna de estas iniciativas han sido desarrolladas especialmente para sitios Web 2.0 orientados a salud.

Además de las iniciativas anteriores, otros autores han propuesto métricas de calidad, algunas heredadas de la Web 1.0 y otras específicas para Web 2.0; que permiten apoyar las evaluaciones de calidad de los sitios Web. Muy pocas han sido definidas para sitios Web 2.0 orientados a salud. Por ejemplo, como se menciona en un estudio sobre Wikis en salud que aplica los criterios de calidad definidos por el “*Health Summit WorkingGroup*” HSWG[4], la mayoría de estos sitios aprobaron los criterios de calidad relacionados con el diseño, accesibilidad y fácil navegación; pero la mayoría obtuvo calificaciones bajas con respecto a la procedencia de la información y a su relevancia.

Es claro que en redes sociales donde existe un número grande de usuarios, es una tarea muy compleja para el administrador o moderador de la red social mantener el control de calidad de la información que se publica. Es por esto que se han desarrollado algunas iniciativas que pretenden automatizar el proceso de evaluación de la calidad. Estas Iniciativas como “Advogato”[5], Film Trust[6] y Trust Mail[7] han sido principalmente desarrolladas para dominios diferentes a la salud.

Considerando el problema anterior, la pregunta de investigación que se plantea en este proyecto es:

¿Cómo evaluar la calidad de la información en la Web 2.0, contenida en Redes Sociales Virtuales en Salud de manera automática o semiautomática?

La hipótesis propuesta es que es posible identificar métricas y desarrollar un servicio Web apoyado en dichas métricas, para evaluar de manera automática o semiautomática la calidad de los contenidos en redes sociales en salud.

### **1.3 Conceptos Básicos**

#### ***Calidad en salud***

Se entiende como la provisión de servicios de salud a los usuarios individuales y colectivos de manera accesible y equitativa, a través de un nivel profesional óptimo, teniendo en cuenta el balance entre beneficios, riesgos y costos, con el propósito de lograr la adhesión y satisfacción de dichos usuarios.

(Decreto presidencial 1011 de 2006)[8].

#### ***Calidad***

El glosario de la IEEE[9], cuenta con dos definiciones de calidad.(1) La calidad como el grado en que un componente de un sistema, o un proceso alcanza las expectativas o las necesidades de un cliente o un usuario.(2) La calidad como el grado en que un componente de un sistema, o un proceso alcanza o reúne unos requerimientos específicos.

#### ***Atributo de Calidad***

Un rasgo o característica que afecta la calidad de un elemento [9].

#### ***Métrica de Calidad***

(1) Es una medida cuantitativa del grado en que un ítem posee un atributo de calidad dado. (2) Una función cuyas entradas son datos de software y cuyas salidas son un valor numérico único que puede ser interpretado como el grado en que el software posee un atributo de calidad dado[9].

### **Redes Sociales en Salud**

Desde una perspectiva sociológica, las redes sociales describen una categoría diferente de los actores sociales (como individuos, como familias, como comunidades, organizaciones, instituciones gubernamentales, etc.) y sus relaciones sociales. Esto evidencia que aunque los individuos se encuentren socialmente conectados su condición de salud es interdependiente, y más allá de eso los servicios de salud provistos deben trascender la atención de los pacientes al grado de no solo impactar sobre la salud sino también generar rutinas de comportamiento saludable [10].

## **Redes Sociales Virtuales en Salud (RSVS)**

Las redes sociales de pacientes son espacios virtuales donde pacientes con condiciones similares se ponen en contacto y comparten tratamientos y problemas de salud. Estas redes ofrecen a las personas con problemas específicos de salud la oportunidad de compartir experiencias, buscar, recibir y proveer información, consejos e incluso apoyo emocional online [11].

### **Web 2.0**

La Web 2.0 es la red de Internet como plataforma de servicios, entregada a los dispositivos conectados. Las aplicaciones para la web 2.0 son las que aprovechan las ventajas intrínsecas de la plataforma, entregando software como un servicio continuamente actualizado, que mejora cada vez más cuando las personas lo usan, consumiendo y re-mezclando los datos de múltiples fuentes, incluyendo usuarios individuales; mientras se proveen sus propios datos, de una manera que permite mezclarlos con otros, creando un efecto de red por medio de una arquitectura de participación, mejorando la experiencia del usuario más allá de la naturaleza estática de la Web 1.0 [12].

### **Servicio Web**

Es una tecnología que utiliza un conjunto de protocolos y estándares que sirven para intercambiar datos entre aplicaciones. Distintas aplicaciones de software desarrolladas en lenguajes de programación diferentes, y ejecutadas sobre cualquier plataforma, pueden utilizar los servicios web para intercambiar datos en redes de ordenadores como Internet. La interoperabilidad se consigue mediante la adopción de estándares abiertos [13].

### **Comunidad Fuertemente Unida (Tightly knit Community, TKC) y su efecto.**

*TKC*: es un pequeño conjunto de sitios, pero altamente conectados. El *Efecto TKC* es un fenómeno que ocurre cuando esta pequeña comunidad obtiene valoraciones muy altas debido a la aplicación de algoritmos de análisis de enlaces, aún si los sitios dentro de la TKC no son considerados autoridades dentro del contexto o pertinentes dentro del mismo [14].

### **Moderación**

El contexto de este trabajo de grado se define moderación de contenido, como el conjunto de la decisión tomada por el usuario de la RSVS de publicar un contenido, y el apoyo que pueda brindar el servicio basado en métricas en la toma de esta decisión por medio de la valoración del contenido.

## ***Infodemiología (infodemiology) & Infovigilancia (infoveillance)***

La infodemiología puede ser definida como la ciencia que estudia la distribución y los determinantes de la información en un medio electrónico, especialmente en internet, o en una población; con un último objetivo de informar sobre políticas de salud pública. Los datos de la infodemiología pueden ser recolectados y analizados en tiempo real, mientras que el término infovigilancia se refiere a la vigilancia automatizada y continuo análisis de, información y textos libres sin estructura definida sobre internet. Esto incluye análisis de motores de búsqueda (desde el punto de vista de la persona que accede al servicio), blogs (desde el punto de vista del que provee el servicio) y otras aplicaciones de la Web 2.0 [15].

### **1.4 Objetivos**

#### **1.4.1 Objetivo General**

Proponer un mecanismo para la gestión semiautomática la calidad de los contenidos publicados en sitios de redes sociales en salud.

#### **1.4.2 Objetivos Específicos**

Los objetivos específicos planteados para responder a este objetivo general son los siguientes:

- Seleccionar un conjunto de atributos y métricas sobre calidad de la información en redes sociales en salud, que puedan ser utilizadas para evaluar automáticamente o semiautomáticamente la calidad de contenidos a ser publicados en sitios de redes sociales en salud.
- Desarrollar un servicio Web para la moderación de contenidos publicados en una Red Social en Salud, que se apoye en al menos uno de los atributos y/o métricas de calidad identificados en el objetivo anterior.
- Evaluar la precisión del servicio desarrollado usando métodos de evaluación en ingeniería de software.

### **1.5 Organización del documento**

A continuación se describen los capítulos que conforman el documento:

#### **Capítulo 2: Síntesis de base del conocimiento asociada a atributos y métricas de calidad en la web 2.0**

Este capítulo muestra el proceso llevado a cabo para la selección de atributos y métricas para la gestión de calidad de la información en RSVS. Además se realiza una descripción de la metodología utilizada y una revisión de trabajos anteriores sobre este tópico. El resultado es la construcción de un repositorio de atributos y métricas, además de las fichas descriptivas de los trabajos relacionados, lo que

permite realizar el proceso de selección de las métricas adecuadas y que darían cumplimiento al primer objetivo específico del trabajo de grado.

### **Capítulo 3: Desarrollo e implementación del servicio de moderación de contenidos en la RSVS de Unisalud.**

Este Capítulo muestra el proceso de construcción del servicio de moderación de contenidos en la RSVS de Unisalud. Primero se analizan los procesos de moderación en diferentes entornos de la Web 2.0, incluyendo la Red Unisalud que es una Red Social creada en un proyecto del Grupo de Ingeniería Telemática de la Universidad del Cauca. Posteriormente con base a una metodología para Ingeniería de requisitos, se procede a plantear cada una de las fases para el desarrollo del servicio, la retroalimentación con los posibles clientes y por último se generan los Modelos de Casos de uso y del dominio finales, y los requisitos finales para la posterior implementación del mismo.

### **Capítulo 4: Pruebas y análisis de resultados**

En este capítulo se selecciona una metodología de evaluación adecuada para la evaluación y el desarrollo de las pruebas de operación del servicio web dentro de la RSVS apoyados en la Metodología DESMET. Se presentan los resultados de la evaluación del servicio usando el método de Efectos Cualitativos-Opinión de expertos Intercalados de la metodología DESMET, analizando la precisión de las búsquedas obtenidas con el servicio propuesto.

### **Capítulo 5: Conclusiones y trabajo futuro**

En este capítulo se presentan las conclusiones del trabajo de grado, y las proyecciones del mismo después de la finalización de cada uno de los objetivos.



## Capítulo 2

# Síntesis de la base del conocimiento asociada a atributos y métricas de calidad en la web 2.0

### 2.1. Introducción

El presente capítulo de este trabajo de grado brinda un repositorio y análisis de atributos y métricas para la evaluación de la calidad de los contenidos publicados en redes sociales virtuales en salud.

En la sección 2.2 se presenta una aproximación metodológica que permite visualizar el proceso de selección de los atributos y métricas de calidad aplicables a las RSVS. En la sección 2.3 se presenta un análisis del estado del arte, en cual se desarrolla una revisión de métricas y atributos de calidad en salud orientados a la Web 2.0 y RSVS en la actualidad. Luego, en el apéndice A, se hace referencia a un conjunto de fichas descriptoras que tienen como propósito realizar una descripción de cada uno de los autores e iniciativas por definir atributos y métricas de calidad para la salud en la Web 2.0, y las cuales nos permiten hacer una clasificación detalladas en las secciones 2.4 y 2.5. Por último, en la sección 2.6, se presentan las conclusiones del capítulo.

### 2.2. Aproximación Metodológica

Como referencia metodológica para la identificación y selección de los atributos, así como de las métricas de calidad, fue utilizado el modelo para la investigación documental propuesta por Serrano [16]. Las fases de este modelo fueron adaptadas para su utilización, y se mencionan a continuación: Fase Preparatoria, Fase Descriptiva, Fase de Interpretación, Fase de Construcción, Fase de Extensión y Publicación. En la primera fase, después de llevar a cabo una revisión de la literatura sobre las diferentes iniciativas de gestión de calidad para la web en general y en específico para la Web 2.0 al igual que las RSVS, se identificaron los aspectos más importantes en los cuales se debía centrar la investigación. Posteriormente fue trazado el plan para identificar y seleccionar los atributos, y métricas de calidad. En la segunda fase (Descriptiva), partiendo de la literatura revisada, se procedió a identificar el conjunto global de atributos y métricas mencionadas, para posteriormente llevar a cabo la creación de unas fichas descriptoras que fueron consignadas en el apéndice A. Estas fichas dieron como

resultado una extensa tabla resumen de atributos y métricas de calidad. En la fase de Interpretación, fue realizado un estudio más profundo de los resultados encontrados en la fase anterior. El estudio se realizó revisando los atributos y métricas, lo que permitió seleccionar los atributos más críticos para la gestión de calidad del contenido en RSVS y aislar aquellas que tienen la posibilidad de ser automatizadas.

## **2.3. Análisis del estado del arte**

### **2.3.1. Fase Preparatoria**

En la fase preparatoria se describirá, con base en el modelo para la investigación documental, dos partes importantes: La definición de los Núcleos Temáticos (NT) y la definición de las fuentes que se utilizaron a la hora de elaborar las fichas descriptoras.

#### **2.3.1.1. Declaración del Alcance de la base del conocimiento:**

El propósito de generar esta base de conocimiento es tener una colección de atributos, métricas, y servicios que permitan asegurar que la calidad de la información en una RSVS sea óptima, centrándonos en la RSVS de la Unidad de Salud de la Universidad del Cauca – Unisalud <sup>1</sup>, de la cual son usuarios los Estudiantes, Administrativos y Profesores de la Universidad del Cauca.

Para llevar a cabo la generación de esta base de conocimientos es importante seguir las recomendaciones de Serrano, en el Cap2 y el anexo 2 del Modelo Integral para el profesional en ingeniería [16], donde se sugiere la necesidad de establecer unos límites que permitan enmarcar el estudio.

### **LÍMITES DEL ESTUDIO**

- (1) Relacionados con el tiempo.-** Fecha de inicio de la investigación: 1 diciembre 2012-fecha finalización 1 febrero de 2013. Se consideran documentos publicados en los últimos 10 años.
- (2) Relacionados con el material a estudiar.-** La metodología desarrollar es elegir las palabras adecuadas para realizar las búsquedas. Posteriormente, seleccionar las bibliotecas físicas o virtuales en las cuales se van a realizar cada una de las búsquedas, se intentara en su mayoría buscar artículos recientes.
- (3) Relacionados con el espacio:** A pesar que el interés del estudio es generar una colección de atributos y métricas para usar en el entorno Colombiano y en

---

<sup>1</sup> <http://esalud.unicauca.edu.co/redunisalud/>

especial el Caucaño, es importante señalar que, al ser Internet una red mundial de información, todos los hispano hablantes somos posibles receptores de la información publicada en una RSVS.

**(4) Relacionados con los integrantes del colectivo:** A pesar que el proyecto y las revisiones solo serán efectuadas por una persona, se pretende hacer una construcción lo más completa posible, en el tiempo establecido para ello.

### **2.3.1.2 Identificación de los Núcleos Temáticos de la base de conocimiento.**

Teniendo en cuenta que el objeto de estudio es la calidad de la información en RSVS, se han identificado dos NT que son de vital importancia y que permiten llegar al cumplimiento de los objetivos del trabajo de grado. De la misma forma se han seleccionado las fuentes relevantes para las búsquedas de trabajos relacionados.

#### **Definición Núcleos Temáticos:**

Los NT seleccionados son los siguientes:

- Criterios y métricas.
- Aplicaciones y servicios.

### **2.3.2. Fase Descriptiva**

En esta fase se procederá a hacer un estudio sobre los NT, teniendo en cuenta los referentes disciplinares y teóricos, y las delimitaciones espaciales, temporales, contextuales de los documentos y autores, ya mencionados en la declaración del alcance, teniendo en cuenta sus perspectivas metodológicas.

#### **2.3.2.1. Conformación del Banco de Unidades de Análisis**

Se efectuó una revisión general de la literatura correspondiente, donde se seleccionaron aquellas unidades de análisis necesarias para llevar a cabo la investigación. Para describir cada unidad de análisis, se utilizaron fichas descriptoras hechas según las necesidades de la investigación como puede observarse en el anexo A. A continuación, se describe la metodología utilizada para la selección de artículos y publicaciones.

a.) El propósito de la revisión general de la literatura es detectar la bibliografía y otros materiales que puedan ser útiles para conformar el banco de documentos más representativos con respecto a gestión de la calidad de la información en RSVS, por medio de atributos y métricas. Para esto se identificaron los documentos, los autores y los grupos de investigación más importantes, y se descartaron los artículos menos relevantes.

- b.) Los documentos objeto de esta investigación documental en un principio fueron las fuentes primarias o directas, ya que proporcionan información de primera mano. Ejemplos: Libros, antologías, artículos, monografías, tesis, disertaciones, documentos oficiales, reportes de asociaciones y ponencias.
- c.) Los esfuerzos se enfocaron en ubicar preferiblemente documentos resultantes de trabajos de investigación, es decir, de trabajos cuyo objetivo central haya sido la generación de conocimiento que socialmente no existía, como por ejemplo el trabajo realizado por Luque en HealthTrust[17].
- d.) También fue importante realizar búsquedas en fuentes secundarias de información relacionadas con el área temática respectiva ya que corresponden a compilaciones, resúmenes y listados de fuentes primarias publicadas en un área temática en particular. Ejemplos de estas fuentes secundarias son los artículos de recopilación desarrollados por Risk y colaboradores[2] , Luque y colaboradores [18].
- e.) Se procedió a realizar búsquedas en fuentes de información terciarias relacionadas con el área temática correspondiente. Este tipo de fuentes de información compendian: Nombres y títulos de revistas y otras publicaciones periódicas, así como nombres de boletines, conferencias y simposios; nombres de empresas, asociaciones industriales y de diversos servicios; catálogos de libros básicos que contienen referencias y datos bibliográficos; y nombres de instituciones nacionales e internacionales al servicio de la investigación.
- f.) Se tuvieron en cuentas las siguientes fuentes primarias de información, debido a que son las más utilizadas para la elaboración de marco teórico o de marcos conceptuales:
- IEEE.
  - ACM.
  - Pubmed.
  - ScienceDirect.
  - Springer.
  - Googlescholar.

En las cuales se realizaron búsquedas en inglés, haciendo combinaciones de las siguientes palabras claves:

*Recommendation, system, health, metrics, automatic, semiautomatic, attributes, quality, virtual social networks, e-health, information filtering, Web 2.0.*

Estas búsquedas dieron como resultado 258 artículos que a simple vista parecían tener información relevante, los cuales fueron filtrados teniendo en cuenta su

resumen (abstract), con la finalidad de encontrar artículos relacionados con los temas buscados “gestión de la calidad de la información en RSVS”, “métricas de calidad y atributos de calidad”, lo que dio como resultado una colección de 86 ítems entre artículos y referencias web. Teniendo en cuenta su relevancia, posteriormente se analizaron más a fondo y se eligieron 23 documentos para los cuales se realizó una ficha descriptora para cada uno.

Después de realizar la revisión se puede concluir que los métodos más populares escogidos por los autores para seleccionar métricas y atributos son: la utilización de grupos focales y la reunión de expertos.

**Grupos Focales:** Generalmente son formados con pacientes, profesionales de la salud y personas relacionados con salud pública o administración de entidades de salud. Estas reuniones son moderadas, para que no se pierda el tema central de la discusión. Usualmente este mecanismo se utiliza para hacer análisis de mercado y de productos, se puede realizar con grupos de población mixtos o heterogéneos.

**Reunión de expertos:** Son todas aquellas organizaciones que generalmente son no gubernamentales, las cuales son conformadas en su mayoría por expertos en salud pública, expertos en temas de prevención de salud, expertos en las tecnologías de la información y las comunicaciones aplicadas a la salud. Estos grupos usualmente se valen de sus conocimientos en diferentes áreas afines a la salud para publicar, como en el caso organizaciones como HON, parámetros o especificaciones para que las personas que acceden a información en salud, tengan certeza del tipo de contenido que están consumiendo.

A continuación se muestra una tabla resumen donde se puede observar los artículos seleccionados para hacer el estudio sobre los Núcleos Temáticos (NT).

**Tabla1.** Discriminación de Núcleos Temáticos y fichas descriptoras.

Artículo	Núcleo(s) Temático(s)	Autor	Tipo material
Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web	Criterios y métricas.	G. Eysenbach, John Powell, Oliver Kuss, Eun-RyoungSa,	artículo de revista JAMIA
¿Social but safe? Quality and safety of diabetes-related online social networks.	Criterios y métricas.	ElissaWeitzman, Emily Cole, LiljanaKaci	artículo de revista JAMIA

Review of Internet health Information Quality Initiatives.	Aplicaciones y servicios, Criterios y métricas.	Ahmad Risk, J. Dzenowagis	<i>Revista "Medical Internet Research"</i>
Certification for collaborative Websites / Websites with Web 2.0 elements.	Criterios y métricas.	fundaciónHON (Health On the Net)	<i>Página Web</i>
Assessment of the quality of medical wikis and Mighealthnet Wiki using Health Summit Working Group Quality Criteria	Criterios y métricas.	E. Dobrogowska, B. Niedźwiedzka	presentación slideshare
Advogato's Trust Metric	Aplicaciones y servicios, Criterios y métricas.	Comunidadde Advogato	<i>Página Web</i>
FilmTrust	Aplicaciones y servicios, Criterios y métricas.	FilmTrust	<i>Página Web</i>
Hi-Ethics: Tough Principles for Earning Consumer Trust	Criterios y métricas.	Donald Kemper	revista U R A C / Internet Healthcare Coalition
The Role of Social Media for Patients and Consumer Health.	Aplicaciones y servicios Criterios y métricas.	A. Y. S. Lau , K. A. Siek , L. Fernandez, H. Tange	revista IMIA Yearbook of Medical Informatics 2011
HealthTrust: a social network approach for retrieving online health videos	Aplicaciones y servicios, Criterios y métricas.	Luis Fernandez, Randi Karlsen, Genevieve B	<i>PublicaciónNorthern Research Institute</i>
Trustmail	Aplicaciones y servicios, Criterios y métricas.	Trustmail Solutions	<i>Página Web</i>

Mediline Plus	Aplicaciones y servicios	Biblioteca Nacional de Medicina de Estados Unidos	<i>Página Web</i>
Quality Guidelines: healthfinder.gov.	Criterios y métricas.	Departamento de Salud y servicios humanos de Estados Unidos	<i>Página Web</i>
Social Media Guidelines and Best Practices	Criterios y métricas.	Centro de prevención y control de enfermedades de Atlanta CDC	<i>Página Web</i>
DISCERN-The DISCERN Instrument	Criterios y métricas.	British Library and the University of Oxford	<i>Página Web</i>
A quality evaluation methodology for health-related websites based on a 2-tuple fuzzy linguistic approach	Criterios y métricas.	J. M. Moreno, J. M. Morales, C. Porcel, E. Herrera-Viedma	artículo de revista
Assessing information quality of a community-based encyclopedia	Criterios y métricas.	BesikiStvilia, Michael Twidale, Linda C. Smith, Les Gasser	artículo de revista
Evaluación de la calidad de blogs sobre educación ecológica	Criterios y métricas.	Fernando Ojeda, Francisco Perales, José Gutiérrez.	artículo de revista.
HealthFinland—A national semantic publishing network and portal for health information	Aplicaciones y servicios	Osma Suominena, Eero Hyvönena, Kim Viljanena, Eija Hukkab	Publicación
COMPUTING AND APPLYING TRUST IN WEB-BASED SOCIAL NETWORKS	Criterios y métricas./ Aplicaciones y servicios	Jennifer Ann Golbeck	Publicación

Videojuegos y Redes Sociales en Línea Para la Sensibilización Sobre la Tuberculosis Pulmonar	Aplicaciones y servicios	Miguel Ángel Niño Zambrano, Hendrys Fabián Tobar Muñoz, Jonatan Lemos Zuluaga	artículo de revista Ingenierías UIS Volumen 8, No. 2
Semi-Automatic Semantic Moderation of Web Annotations	Aplicaciones y servicios	Elaheh Momeni	artículo de revista
Quality Criteria for Health related Websites	Criterios y métricas	eEurope 2002:	Policy Paper Journal of Medical Internet Reserch 2002;4(3):e15

### 2.3.2.2. Construcción de las Fichas Descriptoras

Para cada unidad de análisis se elaboró una ficha descriptora que contiene el conocimiento científico relevante encontrado en dicha unidad, teniendo siempre como referente esencial la declaración del alcance de la base de conocimiento realizada anteriormente. La siguiente es la modificación estructurar realizada a la ficha descriptora, teniendo en cuenta sus parámetros generales:

1. Aspectos formales: Autor, Título de documento, Tipo de material.
2. Síntesis y palabras claves: Abstract, Palabras claves
3. Observaciones: Ventajas, Desventajas y Diferencias.

En el anexo A se encuentran especificadas detalladamente cada una de las fichas agrupadas según los núcleos temáticos. En las Tabla 2 y 3 se presentan dos ejemplos, correspondientes a las Fichas descriptoras para los núcleos temáticos “Criterios y Métricas” y “Servicios y Aplicaciones”.

**Tabla 2.** Ficha representativa del Núcleo Temático Criterios y métricas.

<b>1.- Información del documento</b>
<b>1.1- Autor:</b> Elissa R Weitzman, Emily Cole, LiljanaKaci
<b>1.2.- Título del documento:</b> ¿Social but safe? Quality and safety of diabetes-related online social networks
<b>1.3.- Tipo de material:</b> artículo de revista JAMIA
<b>2.- Comentarios</b>



## 2.1- Ventajas

- las evaluaciones de calidad del contenido se realizaron en el entorno de RSVS, lo que hace que este estudio sea muy relevante.
- Resalta la importancia de hacer una evaluación diferente a como se viene haciendo por parte de mecanismos como las certificaciones de terceros (HON).

## 2.2- Desventajas

- no define las métricas usadas, ni atributos de calidad usados, solo menciona los resultados.
- no tiene en cuenta información general en salud solamente es enfocada a diabetes.
- Se trata de evaluaciones realizadas de forma manual, no se menciona ningún mecanismo automático o semiautomático.

## 2.3- Diferencias

Este estudio contiene entre sus recomendaciones la necesidad de acompañar el mismo con un sistema de verificación de calidad que permita monitorear la publicación segura de contenidos de calidad dentro de la red social. Este sistema debe valorar la autoría individual y de la comunidad, agregando condiciones de innovación y protección al consumidor.

## 3.- Síntesis

### 3.1- Abstract

Poco es conocido de la calidad y seguridad de los sitios relacionados con Redes Sociales Virtuales de Salud. A pesar del reconocimiento que la internet en general y los medios sociales específicamente son vistos como fuentes cruciales por los consumidores a la hora de acceder a información en salud, los datos carecen del valor o impacto de estos recursos. Se han hecho valoraciones no formales para determinar la tendencia del contenido publicado en los sitios con la ciencia médica o/y las recomendaciones sobre las prácticas clínicas, efectividad del sitio, o la información que se publica sobre los miembros, la transparencia del contenido, de la publicidad, y la efectividad de las protecciones contra la desinformación y la publicidad engañosa.

**3.2- Palabras Clave** no se relacionan palabras claves en el artículo

La ficha correspondiente a la tabla 2, es una ficha representativa del núcleo temático 1) Criterios y métricas. Se escogió esta ficha en particular porque está orientada en un estudio de la calidad de información en redes sociales de diabetes en los estados unidos.

**Tabla 3.** Ficha representativa del Núcleo Temático Aplicaciones y servicios.

<b>1.- Información del documento</b>
<b>1.1- Autor:</b> LuisFernandez-Luque, Randi Karlsen, Genevieve B
<b>1.2.- Título del documento::</b> "HealthTrust: a social network approach for retrieving online health videos"
<b>1.3.- Tipo de material:</b> <i>PublicaciónNorthernm Research Institute, Tromse, Norway.</i>
<b>2.- Comentarios</b>
<b>2.1- Ventajas</b>
<ul style="list-style-type: none"> <li>• HealthTrust tiene un mejor desempeño para buscar contenido de calidad sobre la diabetes que el buscador de YouTube.</li> <li>• Se define el concepto de relevancia en búsquedas usando el algoritmo de HealthTrust, cuyo valor depende de si la búsqueda concuerda con el título o la descripción de la misma de la siguiente forma: concuerda con el título 100%, concuerda con la descripción 20%, con ninguna de las dos 0%.</li> <li>• HealthTrust se basa sobre la premisa que el contenido producido por un autor de confianza será confiable la mayor parte del tiempo. Por lo cual define que la autoría está relacionada en un 70% al autor y un 30% a las calificaciones del contenido.</li> <li>• En las pruebas realizadas a HealthTrust en la Red Social Virtual de videos YouTube, se comprobó que el algoritmo es eficiente filtrando videos con contenido publicitario y de información no adecuada.</li> </ul>
<b>2.2- Desventajas</b>
<ul style="list-style-type: none"> <li>• Algunos videos de fuentes referenciadas tuvieron calificaciones bajas.</li> <li>• Las calificaciones de HealthTrust pueden ser afectadas por bajas interacciones sociales en la RSVS (ej: amistad, suscripciones).</li> </ul>
<b>2.3- Diferencias</b>
HealthTrust está definida en general para cualquier comunidad virtual con contenidos de salud, pero en la actualidad solo se han realizado pruebas en la Red Social Virtual YouTube.
<b>3.- Síntesis</b>
<b>3.1- Abstract</b>
Los medios sociales se han vuelto algo de todos los días en el dominio de la salud. A pesar de los grandes volúmenes de información en salud acertada y de confianza disponible en las plataformas de medios sociales, actualmente encontrarla puede ser una tarea bastante dispendiosa. La información malintencionada puede ser muy popular (Ej: videos en contra la vacunación) y además puede obtener altas calificaciones en los motores de búsqueda. Los estudios indican que las comunidades online de consumidores de salud

pueden ser efectivas para encontrar información de salud confiable. Pensamos que la sabiduría de la comunidad acerca de la calidad de la información en salud puede ser tenida en cuenta para ayudar a crear herramientas para buscar y retornar contenido social en salud. Por medio del análisis de enlaces de una comunidad de salud, se extrajo información sobre la autoría de los proveedores de videos de diabetes. Implementamos una métrica para calcular la reputación de los proveedores de contenido dentro comunidad en línea. Nosotros evaluamos la métrica con la comunidad de YouTube de diabetes, incluyendo el uso de las métricas para la búsqueda de videos de diabetes.

### **3.2- Palabras Clave**

La ficha correspondiente a la tabla 3 es una ficha representativa del segundo núcleo temático “Aplicaciones y servicios”. Se escogió esta ficha en particular porque en este trabajo se propone realizar un filtrado de Calidad de información, basado en la métrica HealthTrust sobre una RSV, en este caso YouTube, y a su vez sobre canales específicos de temas de salud.

#### **2.3.3. Fase de Interpretación por núcleos temáticos.**

Para cada núcleo temático se construye una tabla de interpretación con base en las fichas descriptoras de las unidades de análisis correspondientes, que se encuentran en el anexo A.

En la figura 1 de la sección 2.3.3.1 se presenta un modelo de caracterización jerárquico propuesto. Esto para dar un mayor entendimiento a los términos que aparecen en cada uno de los trabajos enunciados en el anexo A. Luego en la tabla 4 se presenta, según el Modelo para Construcción de Soluciones, las relaciones, inferencias y asociaciones que se hicieron a la hora de realizar la investigación documental. Además, en las secciones 2.3.3.2 y 2.3.3.4 se presentan los aportes, vacíos y limitaciones de la correspondiente Unidad de Análisis en relación con el tema central respectivo y los núcleo temáticos. La sección 2.3.3.2 representa la Ficha de Interpretación para el Núcleo Temático “Criterios y Métricas”, y la sección 2.3.3.4 representa la Ficha de Interpretación para el Núcleo Temático “Servicios y Aplicaciones.

### 2.3.3.1. Resumen de los parámetros de calidad obtenidos por medios de las fichas descriptoras

**Figura 1.** Modelo de organización jerárquico entre los diferentes parámetros de calidad.



En la figura 1 se puede observar un modelo de organización jerárquica, el cual nace después de la comparación realizada entre los diferentes trabajos mencionados, y caracterizados por medio de las fichas descriptoras del anexo A.

**Tabla 4.** Dimensiones, criterios eEuro y principios HON, sub-criterios, atributos y métricas

<b>DIMENSIÓN</b>	<b>CRITERIOS eEURO/PRINCIPIOS HON</b>	<b>SUB-CRITERIOS</b>	<b>ATRIBUTOS</b>	<b>MÉTRICAS</b>
<b>CREDIBILIDAD</b>	AUTORIDAD ( LA INFORMACIÓN DE LA AUTORIA DEBE SER FIDEDIGNA)	Divulgación de la autoría	Autoría / Confianza	Advogato/ HealthTrust/ Autoridad/
		Divulgación de la fecha de creación y edición		
		Credenciales del autor divulgadas incluyendo la fecha de obtención de las mismas.		
		Referencias provistas		
		Debe ser posible identificar los roles de todos los participantes en la RSVS		
		Debe ser posible identificar por medio de sus credenciales quien es el dueño o responsable (moderador) del sitio web.		
	TRANSPARENCIA Y HONESTIDAD (PROPÓSITO DEL SITIO WEB, DIVULGACIÓN FINANCIERA, DETALLES DE CONTACTO DEL SITIO WEB)	El lugar de origen del sitio web debe estar especificado, junto con la dirección de contacto del dueño o responsable del sitio.	Autoría	
		Debe existir un formulario de contacto con los responsables del sitio web.		
		El propósito y objetivo del sitio web debe ser transparente.		
		Las metas especificadas deben ser satisfechas.		
		La audiencia debe ser especificada.		
		Los patrocinadores e inversionistas deben ser		

<b>CREDIBILIDAD</b>	TRANSPARENCIA Y HONESTIDAD (PROPÓSITO DEL SITIO WEB, DIVULGACIÓN FINANCIERA, DETALLES DE CONTACTO DEL SITIO WEB)	divulgados.		
		Las declaraciones de conflictos de intereses deben ser mostradas.		
		La publicidad y los contenidos deben ser diferenciados.		
		La publicidad no debería ser contradictoria con respecto a los contenidos del sitio web.		
	PRIVACIDAD Y PROTECCIÓN DE LOS DATOS(CONFIDENCIALIDAD)	Una declaración de la protección de datos personales debe ser mostrada.		
		Los post de todos los usuarios serán públicos por defecto.		
		Se debe establecer si los post son susceptibles a modificaciones o a ser borrados por su creador.		

<b>DIMENSIÓN</b>	<b>CRITERIOS eEURO/PRINCIPIOS HON</b>	<b>SUB-CRITERIOS</b>	<b>ATRIBUTOS</b>	<b>MÉTRICAS</b>
<b>CONTENIDO</b>	ACTUALIZACIÓN DE LA INFORMACIÓN(LA INFORMACIÓN DEBE SER DOCUMENTADA, FECHADA Y REFERENCIADA)	Evaluación personal del contenido del sitio web.		
		La fecha de publicación en la RSV debe aparecer en los documentos.	Actualización	Actualidad / Volatilidad
		El contenido debe estar completo y actualizado.	Precisión /completitud	Compleitud/ Consistencia/ Informatividad
		El lenguaje debe ser entendible	Legibilidad	Complejidad
	COMPROMISO DE VERACIDAD	El contenido debe ser acertado y veraz desde la experiencia y conocimiento del usuario.	Precisión	
		Bibliografía debe aparecer en los documentos.		
	POLÍTICAS DE PROPAGANDA	Se debe establecer si el usuario puede publicar publicidad a manera de contenido.		
<b>USABILIDAD</b>	ACCESIBILIDAD	EL diseño y la organización de un sitio web deben ser amigables para el usuario.	Apariencia	
		El sitio Web debería brindar opciones para moverse libremente sobre su estructura.	Presentación	

<b>USABILIDAD</b>	<b>ACCESIBILIDAD</b>	El sitio web no debería tener problemas de para realizar búsquedas.	Presentación	
		La velocidad de navegación sobre el sitio web debería ser alta.		
		El sitio web debería contener un motor de búsqueda versátil y amigable al usuario.		
		Las descargas deben ser amigables al usuario.		
		El sitio web no debería contener publicidad emergente.		
<b>ENLACES EXTERNOS</b>	<b>GESTIÓN RESPONSABLE</b>	Los enlaces externos deberían mostrar una descripción completa del Sitio al cual se enlaza.		
		Los enlaces externos deben ser distinguidos de los internos.		
<b>INTERACTIVIDAD</b>		El sitio web tiene que tener una sección de preguntas.		
		Si el sitio web tiene un foro, debería ser posible identificar al administrador.		
		Si el sitio web tiene un foro, el acceso debería ser con identificación.	Confidencialidad	
		Disponibilidad de herramientas de autodiagnóstico.		



Es importante destacar que la tabla 4 es el resultado de la aplicación del Modelo de descripción jerárquico descrito en la figura 1, y al ser un modelo es visto desde el punto de vista del investigador documental, lo que no lo hace correcto o incorrecto, pero constituye una herramienta didáctica y representativa que en el caso de este trabajo de grado se usa para correlacionar los diferentes trabajos enumerados en el Anexo A. Para elaborar este modelo se tuvo en cuenta las definiciones encontradas, buscando ir de lo general a lo particular y teniendo en cuenta que las políticas públicas (Quality Criteria eEurope) deben estar en un nivel más alto a las categorizaciones realizadas por entidades no gubernamentales, como es el caso de la fundación “Health On the Net” (HON) la cual se destaca como la organización más reconocida en internet, en el campo de la gestión y el aseguramiento de la calidad de la información en salud.

A continuación se realiza una descripción de cada una de las categorías mostradas en la tabla 4.

### **Dimensiones**

Las dimensiones seleccionadas teniendo en cuenta las fichas descriptoras del Anexo A, han sido obtenidas con base a los trabajos de Moreno y colaboradores [19] y del HSWG (Health Summit Working Group)[20], las cuales fueron consignadas en la tabla 4.

**D1 Credibilidad:** esta dimensión se refiere a todo aquello que permite que la información disponible dentro del sitio web sea creíble para los usuarios. De la misma forma en esta dimensión también se encuentran los criterios de calidad asociados a la identificación de los dueños del sitio web (La identificación de la institución u organización), patrocinio (divulgación de los patrocinadores y la naturaleza del soporte dado), objetivos (los objetivos deben estar especificados), las pautas (las pautas y los contenidos de la información deben ser diferenciados), etc...

**D2 Contenido:** esta dimensión está compuesta por los criterios relacionados a contenidos de información en salud provistos por sitios web. La precisión de la información es el atributo más obvio para la calidad del contenido, y los usuarios tienen el derecho de esperar que los sitios web provean información precisa.

**D3 Usabilidad:** esta dimensión se refiere a la funcionalidad para mejorar la facilidad de uso durante las visitas de los usuarios. El análisis a esta dimensión se realiza principalmente teniendo en cuenta principalmente criterios relacionados con el diseño del sitio web (organización lógica de los elementos de tal forma que los visitantes puedan entender fácilmente como utilizar el sitio web) y el mecanismo de navegación (Las herramientas de navegación provistas por el sitio web necesarias para acceder a información específica).

**D4 Enlaces externos:** esta dimensión está relacionada a las correlaciones desde un sitio Web hacía otros sitios externos, formando una estructura tipo web de información entre sitios web. Los proveedores de información del sitio web deben

poner especial atención a cada uno de los elementos que conforman esta dimensión, ya que los usuarios que llegan a su sitio web confían de la misma manera en los sitios relacionados.

**D5 Servicios Interactivos:** esta dimensión se refiere a los servicios interactivos provistos por el sitio web, así como por ejemplo, foros web o servicios en línea de preguntas (cuestionarios en línea).

## **CRITERIOS eEURO/PRINCIPIOS HON**

eEurope quality criteria[21] y los principios de HON[22], fueron agrupados por sus semejanzas, además que las dos iniciativas fueron desarrolladas en Europa. Los criterios de eEuropa son un estándar aceptado y HON es la entidad no gubernamental más reconocida en el aseguramiento de la calidad sobre la red, y sus principios para la Web 2.0 son muy recientes.

### **Autoridad**

La Autoridad se refiere a una característica de los participantes de un entorno, en este caso las RSVS, donde estos son reconocidos o destacados por sus niveles de conocimiento, invención, o influencia.

#### *La información de la autoría debe ser fidedigna*

- Todas las fuentes de información deben de ser provistas, la fecha de publicación y la fuente.
- El nombre y credenciales deben estar claramente establecidos, al igual que información sobre si la plataforma es moderada o no. Las cualificaciones/credenciales de los moderadores deben estar claramente establecidas. Los nombres verdaderos de los moderadores no son obligatorios. Los moderadores pueden usar pseudónimos. La frecuencia de las visitas de los moderadores deben estar publicadas.
- Debe ser establecido que todos los usuarios son considerados por defecto como “usuarios normales y no profesionales de la salud”. De lo contrario, debe ser especificado.
- Debe ser establecido que si un profesional de la salud brinda información como un usuario de la plataforma, su estatus como profesional médico debe ser identificado.
- Todas las personas y proveedores institucionales de información deben ser anunciados en el sitio, incluyendo fechas en las cuales las credenciales fueron recibidas.

### **Transparencia y honestidad**

El sitio propuesto debe tener un propósito totalmente transparente, el mismo debe ser enfocado en una audiencia objetivo, o los contenidos deben ser adaptados a esta población por medio de los perfiles de usuario. Es importante que sea visible para los usuarios si los moderadores son o no voluntarios y por último debe incluir

la siguiente información: nombre, la dirección física y la dirección electrónica de la persona u organización responsable de este sitio.

Detalles de contacto del sitio web:

Debe estar disponible en la RSVS y de una manera sencilla. Debe proveerse un formulario de contacto con los administradores o dueños del sitio web, o los datos de contacto de los mismos Ej: correo electrónico.

Propósito del sitio web:

Debe estar establecido claramente dentro de la RSVS que la información ahí encontrada no tiene como fin remplazar los dictámenes emitidos por el profesional de la salud, sino está diseñada para complementar los mismos. La edad mínima para el uso y la participación dentro de la RSVS debe estar establecida.

Divulgación financiera:

Es importante que se especifique si el personal involucrado en el sitio web es voluntario o no.

**Privacidad y protección de los datos:**

La política de privacidad y protección de los datos y el sistema para el procesamiento de los datos personales, incluyendo el procesamiento de usuarios sin registrar, debe ser establecido de conformidad con la legislación para protección de datos del lugar donde se encuentre la RSVS. En el caso de la comunidad Europea, donde se crearon los criterios de eEurope y principios HON, se debe tener en cuenta la legislación europea para protección de datos [23]. En Colombia debe considerarse la Ley Estatutaria No. 1581 “Disposiciones generales para la Protección de datos personales” [24].

Confidencialidad:

La plataforma debe tener una política de privacidad de los datos, y debe estar subrayado el hecho que todo el mundo tiene la posibilidad de leer un post y usar su contenido, de la misma forma debe estar establecido si el usuario tiene la posibilidad de modificar y borrar sus publicaciones.

**Actualización de la información:**

De debe proveer una clara y regular actualización del sitio, con fecha y actualización claramente mostrada en cada página o en cada ítem (si es relevante). El chequeo de la relevancia de la información debe realizarse de manera regular.

La información debe ser documentada, fechada y referenciada:

Una declaración pidiendo a los usuarios de la plataforma sobre su deber de dar referencias (enlaces por ejemplo), sobre la información de salud o médica que proveen cuando esta no sea basada en la experiencia personal.

Las experiencias personales se refieren a cualquier tipo de síntoma, test, tratamiento que la persona o un familiar cercano ha experimentado (sufrido). Todas las publicaciones se deben fechar automáticamente.

**Compromiso de veracidad:**

Debe estar estipulado que todos los usuarios de la plataforma deben comprometerse a diseminar solamente información verdadera a la luz de su propio conocimiento.

**Accesibilidad:**

Atención a las directrices sobre accesibilidad física, así como características generales de facilidad, búsqueda, lectura, utilización, etc.

**Gestión responsable:**

- Retroalimentación con los usuarios y una apropiada responsabilidad de supervisión (como un oficial de cumplimiento de la calidad designado para cada sitio).
- Asociación responsable: Deberían realizarse todos los esfuerzos para asegurar que las asociaciones o enlaces con otros sitios web, sean realizados solamente con individuos u organizaciones de confianza, y que estos a su vez cumplan con códigos relevantes de buenas prácticas.
- Política Editorial: Debe estar estipulado claramente cuáles son los procedimientos que fueron usados para la selección del contenido.

**Políticas de propaganda:**

Debe ser determinado si los usuarios de la plataforma tienen el derecho de publicar publicidad a manera de contenidos, banners o enlaces, etc...

**2.3.3.2. Aportes, vacíos y limitaciones del NT “Criterios y Métricas” correspondiente a la Unidad de Análisis Gestión de la calidad de la información en RSVS.**

En esta sección se analizan los aportes, vacíos y limitaciones del NT “Criterios y Métricas” correspondiente a la Unidad de Análisis Gestión de la calidad de la información en RSVS.

#### 2.3.3.2.1 HealthTrust: a social network approach for retrieving online health videos[17].

Se define una métrica denominada HealthTrust, que evalúa la confianza en los autores de los contenidos, valiéndose del análisis de enlaces que en este caso puntual, son los canales de video de Youtube y sus suscriptores. HealthTrust, aunque sólo ha sido probada inicialmente en la red social YouTube, tiene todos los elementos necesarios para ser aplicada en una red social en salud, debido a que fue concebida en el campo de la salud (canales de diabetes) y es específica de la Web 2.0.

#### 2.3.3.2.2. Guide to Healthy Web Surfing: MedlinePlus[25].

Este documento muestra como la biblioteca Nacional de Medicina de Estados Unidos, por medio de su iniciativa Medline plus, propone una serie de recomendaciones que deben tener en cuenta todas las personas que buscan información en la red relacionada con salud. A pesar de no definir los criterios de manera detallada, se mencionan en su orden criterios como la autoría de la fuente "Authorship", pertinencia, actualidad de la información y también la privacidad de la información sobre los perfiles de usuario. Los criterios no están explícitos en las recomendaciones y se quedan simplemente en recomendaciones. A pesar de estar enfocado en el campo de la salud, no es específico de la Web 2.0.

#### 2.3.3.2.3. Quality Guidelines: healthfinder.gov - Your Source for Reliable Health Information[26].

En este documento se puede apreciar como healthfinder.gov describe una guía de calidad para la información que se comparte en este sitio Web. En este formulario se hace una especie de revisión inicial de los sitios Web a los cuales se les permitirá hacer parte de este buscador. Este sitio Web es un híbrido de tecnologías Web 1.0 y Web 2.0 porque no permite que el usuario cree sus propios contenidos, o recomiende contenidos fuera de los que se encuentran en la misma Web. Sin embargo se vale de los medios de las redes sociales como Facebook, Twitter y dos Widgets creados por ellos mismos, que permiten ser añadidos a cualquier web para compartir la información del sitio. A pesar que es una iniciativa interesante en el campo de la búsqueda sobre Internet de contenidos de calidad en salud, no hace referencia a atributos ni métricas de calidad concretos.

#### 2.3.3.2.4. Social Media Guidelines and Best Practices[27].

El centro de prevención y control de enfermedades de Atlanta (CDC, Centers for Disease Control and Prevention), brinda una serie de características y buenas prácticas que deben ser cumplidas por sus empleados para la publicación de videos concernientes a información de prevención en salud, en su canal de YouTube. Dentro de las buenas prácticas se puede destacar la recomendación en la cual los videos y el contenido deben estar acordes para la población a la cual está destinada, teniendo en cuenta la forma de promoción y marketing que para este caso específico es el carácter viral de las RSVS. También se resalta que toda la información e imágenes utilizadas, hagan parte de los contenidos aprobados por

el mismo centro de prevención de enfermedades (autoría). A pesar de que la CDC ofrece una serie de lineamientos válidos para el contenido en salud, los criterios no están explícitos, y a pesar que pueden ser válidos en un entorno de redes sociales de salud, no están específicamente diseñados para la Web 2.0.

2.3.3.2.5. A quality evaluation methodology for health-related websites based on a 2-tuple fuzzy linguistic approach[19].

En este estudio se hace uso de los grupos focales para definir unos criterios de calidad de información teniendo como base a los pacientes de un Hospital en la zona rural de Murcia en España. Hacen parte de este grupo pacientes crónicos, pacientes interesados en información de salud sobre Internet y profesionales de la salud. Como resultado del mismo se obtienen 5 Dimensiones que a su vez se dividen en 30 criterios de calidad más específicos, como se muestra en la figura 2:

**Figura 2.** Criterios de calidad agrupados mediante dimensiones de calidad[19].

Quality dimensions	Quality criteria
Credibility	C <sub>1</sub> , C <sub>2</sub> , C <sub>3</sub> , C <sub>4</sub> , C <sub>5</sub> , C <sub>6</sub> , C <sub>7</sub> , C <sub>8</sub> , C <sub>9</sub> , C <sub>10</sub> , C <sub>11</sub> , C <sub>12</sub>
Content	C <sub>13</sub> , C <sub>14</sub> , C <sub>15</sub> , C <sub>16</sub> , C <sub>17</sub>
Usability	C <sub>18</sub> , C <sub>19</sub> , C <sub>20</sub> , C <sub>21</sub> , C <sub>22</sub> , C <sub>23</sub> , C <sub>24</sub>
External links	C <sub>25</sub> , C <sub>26</sub>
Interactive services	C <sub>27</sub> , C <sub>28</sub> , C <sub>29</sub> , C <sub>30</sub>

Este trabajo es muy completo, pero lamentablemente al contar con tantos criterios de calidad, se hace muy difícil asegurar que la evaluación de los usuarios pueda ser totalmente acertada. Como lo afirma Muñoz[28], un cuestionario largo produce fatiga y rechazo en el sujeto que lo completa, con el riesgo añadido que se quede incompleto o se conteste sin la debida reflexión. Por otro lado cuando se realiza la evaluación de la información solo toma en cuenta la opinión un grupo de usuarios, lo que lleva a pensar que está puede ser sesgada ya que la imparcialidad de los usuarios se puede ver nublada por su misma condición de pacientes. Es importante resaltar que a pesar de que las dimensiones y los criterios de calidad fueron elaboradas específicamente para temas concernientes al campo de la salud, no son específicas para la Web 2.0.

2.3.3.2.6. Assessing information quality of a community-based encyclopedia[29].

Besiki Stvilia et al, proponen una forma automática de evaluar los contenidos de Wikis como Wikipedia. Los criterios y métricas de calidad propuestos tienen como prioridad la escalabilidad, ya que en el artículo se referencian muchas iniciativas que no son viables por el tamaño de los cálculos que se deben realizar. Los criterios mencionados son en su orden: Autoría/Reputación, que se refiere a la calidad del autor; Completitud, que refleja si la publicación en Wikipedia abarca el tema en su totalidad; Complejidad, esta métrica se refiere al nivel de conocimiento

que se debe tener para poder entender el artículo; Informatividad, este atributo tiene que ver con que tanto apoyo se tiene por parte de fuentes externas, así como qué tanto se soporta de imágenes para complementar el artículo; Consistencia, se refiere al tiempo que se ha mantenido el artículo en Wikipedia y al número de ediciones por parte del administrador; Actualidad, se refiere a qué tan actualizado está el artículo; Volatilidad, se refiere a qué tantas veces se ha devuelto un artículo a un estado anterior en el proceso de edición por parte del administrador. Estos criterios y métricas no se aplican directamente a información en salud, pero constituyen una referencia importante puesto que se realizaron en el marco de la Web 2.0.

#### 2.3.3.2.7. Quality evaluation of websites and blogs for environmental education[30].

En un proyecto sobre el filtrado de la información disponible sobre temas ambientales, se definen los siguientes criterios de calidad para evaluar sitios Web, blogs y Sindicación Realmente Simple (RSS, Really Simple Syndication) utilizados en educación ambiental: Dinamismo, Relevancia, Validez/Credibilidad, Fiabilidad/Consistencia, Visibilidad. Estos criterios no se aplican directamente a información en salud, pero constituyen una referencia importante ya que se realizaron en el marco de la Web 2.0. Tampoco se definen métricas de calidad.

#### 2.3.3.2.8. eEurope 2002: Quality Criteria for Health related Websites[21].

Propone un conjunto de criterios para asegurar la calidad de la información, en sitios web relacionados con temas de salud, además formula que se usen los criterios relevantes para cada población objetivo, lo que permitiría maximizar las ventajas del filtrado de información. También se hace claridad en que no solo se debe asegurar que los criterios de calidad se cumplan de manera efectiva, sino que esto también tiene que ir de la mano con una educación del usuario y de los administradores de los sitios, para lo cual se hace referencia a una serie de campañas llevadas a cabo por la Comunidad Europea. Por último se relacionan los criterios de calidad de la información para sitios relativos a salud: Transparencia y Honestidad, Autoridad, Privacidad y protección de los datos, Actualización de la información, Responsabilidad (ética), Accesibilidad.

Aunque los criterios de calidad están definidos en el marco de sitios relacionados con contenido de salud, no son específicas de la Web 2.0.

#### 2.3.3.2.9. Review Of Internet Health Information Quality Initiatives[2].

En esta se realiza una revisión de mecanismos para garantizar la calidad de la información publicada en Internet sobre salud. Las iniciativas mencionadas por Risk, se pueden agrupar en:

- Los códigos de conducta y éticos consisten en una serie de principios básicos y criterios de calidad que se deben respetar, más no determina un proceso formal de evaluación del cumplimiento de los mismos. Estos principios son

establecidos por organizaciones como “e-healthcoalition”<sup>2</sup>,”hi-ethics”[10] y “HealthOnthe Net” (HON).

- Las certificaciones por parte de terceros son procesos de certificación continuos donde se valida formalmente el cumplimiento de ciertos estándares de calidad, como por ejemplo el cumplimiento de códigos de conducta. Estas también hacen uso de meta-tags y descriptores XML para asegurar una mejor eficacia en la búsqueda de información por parte de los usuarios. Recientemente HON ha establecido, además de su código de conducta, procesos de certificación. Un ejemplo de organizaciones certificadoras es la organización MedPics, con su certificación MedCertain.
- Evaluaciones basadas en herramientas: Consiste en una serie de cuestionarios evaluados por el usuario, que permiten darle un peso a la información según la calidad definida por el mismo.

#### 2.3.3.2.10. Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web[1].

Eysenbach y colaboradores proponen un framework para evaluar la calidad de sitios web, basado en como se evalúa la calidad de la información en salud en la práctica. Además el uso de los siguientes atributos: Precisión, Completitud, Criterios técnicos, Legibilidad, Criterios de diseño (aesthetics). De manera complementaria en los Criterios técnicos se relacionan atributos de calidad adicionales de los cuales se seleccionó, teniendo como parámetro los que recogen más estudios y los cuales se pueden aplicar a las RSVS: a) Divulgación de la autoría b) Divulgación del patrocinio c) Divulgación de la fecha de creación o edición d) Credenciales del autor divulgadas e) Referencias provistas f) Dirección de correo electrónico provista. Se puede hacer un estudio de los atributos propuestos en esta revisión para ver cuáles son pertinentes a la hora de evaluar contenidos específicos del área de salud sobre Redes Sociales Virtuales en Salud (RSVS).

#### 2.3.3.2.11. Social but safe? Quality and safety of diabetes-related online social networks[32].

Las evaluaciones de calidad del contenido se realizaron en el entorno de RSVS, lo que hace que este estudio sea muy relevante. Además resalta la importancia de hacer una evaluación diferente a como se viene haciendo por parte de mecanismos como las certificaciones de terceros (HON).

Este estudio contiene entre sus recomendaciones la necesidad de acompañar el mismo con un sistema de verificación de calidad que permita monitorear la publicación segura de contenidos de calidad dentro de la red social. Este sistema debe valorar la autoría individual y de la comunidad, agregando condiciones de innovación y protección al consumidor.

---

<sup>2</sup><http://www.ihealthcoalition.org/>



#### 2.3.3.2.12. Advogato - Trust Metric[5].

Se define una métrica basada en la confianza “Advogato”, donde se utilizan los lazos de amistad en redes sociales para establecer una relación adecuada de cercanía entre pares de usuarios que interactúan en la red, teniendo como base, la verificación de credenciales por parte de las personas que realizan publicaciones en los sitios de desarrollo de software libre y la necesidad de la protección de la información personal. Este último es un problema latente entre la mayoría de los usuarios de las redes sociales y que está obligando a que las mismas usen las restricciones máximas de la información de sus perfiles, lo que hace que los servicios de redes sociales pierdan sus beneficios iniciales. La implementación de esta métrica automática de confianza, a pesar de no ser específica de salud, supone un caso de estudio importante debido a que está enfocada en la Web 2.0, y se llevó a cabo por medio de un módulo de Apache llamado “modvirgule”. Con “Advogato” se puede evidenciar la importancia de la métrica de confianza en los entornos de Web 2.0.

#### 2.3.3.3. Tabla resumen de servicios, aplicaciones, aplicaciones Social Media(SM) implementaciones, y herramientas.

**Tabla 5.** Servicios, aplicaciones, implementaciones, y herramientas.

<b>NOMBRE</b>	<b>TIPO</b>
FilmTrust	Aplicación SM
Trustmail	Aplicación SM
HealthTrust	Implementación métrica
Healthfinder.org	Aplicación
DISCERN	Herramienta
2-tuple fuzzy linguistic approach	Implementación
HealthFinland	Servicio
PlanetTB	Aplicación SM

#### 2.3.3.4. Aportes, vacíos y limitaciones del NT “Servicios y Aplicaciones” correspondiente a la Unidad de Análisis Gestión de la calidad de la información en RSVS.

#### 2.3.3.4.1. HealthTrust: a social network approach for retrieving online health videos[17].

En la métrica HealthTrust definida por Luque y colaboradores, se implementa un servicio basado en la Interfaz de programación de aplicación (API, Application Programming Interface) de YouTube, que permite retornar videos realizando una búsqueda de los contenidos de los canales de diabetes contenidos en una Lista Blanca (“White list”), conformada en su mayoría por canales de video de entidades gubernamentales encargadas de la promoción en salud. A pesar de utilizar la métrica de calidad “HealthTrust”, este no es un servicio para la gestión de la calidad, es un buscador que presenta unas mejoras en las búsquedas orientadas a diabetes con respecto al buscador por defecto de la red social Youtube.

#### 2.3.3.4.2. HealthFinland—A national semantic publishing network[33][34].

HealthFinland es un sitio web de Finlandia promovido por el grupo de Investigación en Computación Semántica (SeCo) de la Universidad de tecnología de Helsinki(TKK), El Departamento de ciencias de la computación de la Universidad de Helsinki y el Instituto Nacional para la Salud y el Bienestar de Finlandia (THL), que tiene como función principal garantizar que los usuarios que acceden a la información en salud, puedan encontrar información personalizada de acuerdo con sus necesidades, por medio de un motor de búsqueda especializado. El proyecto se fundamenta en tres pilares fundamentales: 1) minimizar el número de trabajos redundantes o duplicados mediante un repositorio global (mashup Web 2.0 services), 2) servicios inteligentes que permitan al usuario final encontrar información que se adapte a sus necesidades; estos son búsqueda semántica, búsquedas multi-vista centradas en el usuario, búsqueda de cadenas de caracteres basadas en ontologías, 3) bajo costo de mantenimiento, lo que implica que el servicio por si solo se encargue de enlazar el contenido. HealthFinland[22] es una iniciativa muy completa, pero centrada especialmente en búsquedas semánticas. La calidad de la información es asegurada mediante políticas de calidad para la publicación, pero no mediante servicios que gestionen de manera automática la calidad de la información.

#### 2.3.3.4.3. DISCERN - The DISCERN Instrument[35].

DISCERN define una herramienta que consiste en un cuestionario, el cual permite a los administradores y usuarios de sitios Web en salud, evaluar si la información que se ha publicado cumple con todas las características de calidad necesarias para esta clase de sitios. Entre estas se encuentran: 1) ¿la información tiene una meta bien definida?, 2) ¿cumple con la(s) meta(s) propuestas?, 3) ¿es relevante para los consumidores?, 4) ¿son visibles las fuentes de donde proviene la información?, 5) ¿es visible la fecha de creación de la información?, 6) ¿la información es equilibrada e imparcial?, 7) ¿provee una lista de fuentes de información adicional?, 8)¿intenta llenar los vacíos de conocimiento o las diferencias que se puedan generar por opiniones por parte de los expertos en la materia?, 9) ¿describe cómo funcionan los tratamientos clínicos?, 10) ¿describe los beneficios del tratamiento?, 11)¿describe los riesgos del tratamiento?, 12)

¿describe que sucedería sin el tratamiento?, 13) ¿describe los efectos del tratamiento y sus alternativas, así como su impacto en la calidad de vida del paciente?, 14) ¿hace claridad sobre la existencia de más de una opción de tratamiento?, 15) ¿provee soporte en la toma correcta de decisiones, sobre los posibles tratamientos?. El proyecto DISCERN está basado en el trabajo de la División de Salud Pública y Cuidado Primario de la Universidad de Oxford. Cabe resaltar que estas características no son específicas para Web 2.0, y están centradas específicamente en la información de tratamientos médicos, en lugar de información general sobre temas de salud.

#### 2.3.3.4.4. FilmTrust y TrustMail Solutions

FilmTrust[6] y TrustMail[7] son dos aplicaciones de social media que se enfocan en la confianza. La primera consiste en una Web con características de red social y a su vez permite la calificación de películas, en el cual dependiendo de la relación que tenga el usuario con cada uno de las personas que integran sus conexiones sociales y sus evaluaciones de contenido correspondientes, se le realizaran recomendaciones al mismo. Las ventajas de este sitio, se puede apreciar mejor en el caso que las opiniones o evaluaciones sobre las películas por parte de los usuarios sean divergentes. TrustMail es un cliente de correo electrónico que usa los índices de confianza de cada usuario (remitente), como puntaje de calificación para el mensaje. Esto permite que los mensajes de alta importancia queden en la parte más alta del buzón. Este método de filtrado es una mejora al sistema conocido como “White List”, en el cual el usuario ingresa manualmente todos los correos que desea filtrar, mientras que con TrustMail el filtrado se realiza de manera automática basado en las relaciones obtenidas por medio de las redes sociales. La meta de este sistema es proveer altas valoraciones a las personas confiables, y bajas valoraciones al spam. Esta herramienta no está vista como una solución definitiva al spam, sino como un punto de partida para otras herramientas. Aunque FilmTrust y TrustMail no son específicos para la gestión de la calidad de la información en salud, consisten en un punto de partida importante debido a que reafirman la importancia actual de las métricas de confianza o “trust” en la Web 2.0.

#### 2.3.3.4.5. Apache Lucene y Apache Solr

Apache Lucene y Apache Solr, son dos plataformas de código abierto “open source” que permiten realizar búsquedas personalizadas y minería de datos. Entre sus capacidades de búsqueda se encuentran: búsquedas de texto-completo, resaltar los resultados exitosos, “clustering” dinámico, integración de bases de datos, búsqueda dentro de documentos (ej: Word, PDF), y búsqueda geo-espacial. Además, las plataformas son completamente configurables por medio de XML (en campos específicos). En Apache Lucene y Solr, las configuraciones son aplicables en tiempo de indexación, durante la búsqueda o en ambos. Estas plataformas permiten pensar que es posible realizar una configuración de acuerdo a las necesidades de calidad de la información definidas por el usuario, para hacer búsquedas de recursos en la Web que puedan ser publicados en la red social. Sin

embargo, no proveen funcionalidades para búsquedas basadas en criterios de calidad.

### 2.3.4. Fase de Construcción teórica global.

#### 2.3.4.1. Construcción de la Ficha de Construcción Teórica Global.

Con base en el conjunto de fichas de interpretación de núcleo temático, se debe construir una ficha que describa el estado alcanzado de conocimiento, teniendo como referencia esencial la declaración del alcance de la base de conocimiento realizada. Esta ficha representa el nivel más alto de abstracción del soporte teórico construido. En la tabla 6 se observa la estructura general de referencia para la ficha de construcción teórica global.

**Tabla 6.** Estructura de las Fichas de construcción teórica global[16].

<b>F5.- Ficha de construcción teórica global</b>
1. Identificación del tema y de los núcleos temáticos analizados
2. Síntesis del estado actual del conocimiento (Diagnóstico)
3. Estado pretendido del conocimiento (Pronóstico)

#### 2.3.4.2. Síntesis del estado actual del conocimiento (Diagnóstico)

Como el objetivo del análisis del estado del arte es: seleccionar un conjunto de atributos y métricas sobre calidad de la información en RSVS. En la Tabla 7, se presentan los diferentes atributos y métricas categorizados en dimensiones, las cuales fueron resultado del estudio realizado a la tabla 4, según el modelo de categorización descrito en la figura 1.

**Tabla 7.** Atributo(s) y Métrica(s) preseleccionadas.

<b>Atributo(s)</b>	<b>Métrica(s)</b>
Autoría	Advogado, HealthTrust
Confianza	
Precisión	Compleitud/Consistencia/Informatividad
Compleitud	
Legibilidad	Complejidad
Actualización.	Actualidad/Volatilidad
Apariencia	
Presentación	
Confidencialidad	

#### 2.3.4.2.1. Atributos seleccionados

En esta unidad se analizarán uno a uno para poder llegar a un conjunto de atributos acordes a las necesidades de un RSVS. Entre los atributos preseleccionados se encuentran: *Autoría, Confianza, Actualización, Precisión, Compleitud, Legibilidad, Apariencia, Presentación y Confidencialidad.*

“Presentación” y “Apariencia” a pesar de ser atributos importantes no inciden directamente en la calidad de los contenidos publicados en entornos Web 2.0 como lo afirma[4] en su estudio, ya que los blogs analizados en el mismo cumplen con estos dos atributos a muy alto nivel y al mismo tiempo no aseguran la calidad.

“Autoría” y “Confianza” son dos atributos que van de la mano, y que han demostrado ser muy valiosos para evaluar la calidad en otros entornos diferentes a la salud, como Advogato, FilmTrust, TrustMail, Wikipedia. Y han sido evaluados de manera inicial en HealthTrust.

“Actualización”, este atributo resalta la importancia la idea de que la información en salud este actualizada, ya que se perderían beneficios adicionales para el paciente, que podrían repercutir en su calidad de vida.

“Precisión”, “Compleitud” y “Legibilidad” Estos tres atributos se refieren al contenido y serán seleccionados teniendo en cuenta estudios realizados, para la RSVS Unisalud y los criterios ya implementados en su sistema de recomendación.[10]

“Confidencialidad” este atributo se refiere más a la seguridad de los datos suministrados por el paciente, a pesar de ser vital y de suma importancia, no está directamente relacionado con los contenidos en salud.

Por lo mencionado anteriormente podemos decir que los atributos seleccionados son:

- Autoría
- Confianza
- Precisión
- Compleitud
- Legibilidad

#### **2.3.4.2.2. Métricas seleccionadas**

En esta unidad se analizaran una a una para poder llegar a un conjunto de métricas acordes a las necesidades de un RSVS. Entre las métricas pres seleccionadas se encuentran: *Advogato, HealthTrust, Autoridad, Actualidad, Informatividad, Consistencia, Volatilidad, Complejidad y Compleitud.*

“**Advogato**”: es una métrica desarrollada para evaluar la relevancia de las publicaciones en foros de desarrolladores de software libre, lo cual no tiene compatibilidad con una RSVS, pero resalta la importancia de las métricas de confianza “Trust”.

**“HealthTrust”**: es una métrica específica para evaluar contenidos publicados en RSVS, y tiene en cuenta la confianza en el autor “autoría” y en el contenido “confianza”.

**“Autoridad”**: es una métrica basada en los perfiles públicos, que usa Wikipedia para relacionar a sus autores, y no se tendrá en cuenta debido a que realiza un trabajo similar a HealthTrust, pero “Autoridad” no es específica de RSVS.

**“Actualidad”**: es una medida de acuerdo a una marca de tiempo, de la última modificación del documento, en otras palabras que tan actualizado es el documento.

**“Informatividad”**: la informatividad implica que un contenido tiene información nueva y es comprensible para la persona que lo accede. Esta métrica esta modelada, teniendo en cuenta la proporción de información útil contenida dentro de un contenido dado (Info Noise), el número total de personas editando el contenido(diversidad) y el número de imágenes que contiene el documento.

**“Consistencia”**: Se refiere a varios aspectos del contenido. En particular a la a que tan homogénea es la información y la evolución de la misma en el tiempo. Esta métrica tiene en cuenta cuántas ediciones ha tenido el documento por parte del administrador del sitio, además de cuántos años lleva de haber sido publicado.

**“Volatilidad”**: Se refiere al espacio de tiempo en que un contenido puede o no ser válido. En el caso de Wikipedia se tiene como base los valores de tiempo de retorno de los artículos destacados de 199/9 minutos contra los artículos cualquiera 712/20 minutos[29]

**“Complejidad”**: Esta métrica está centrada en que tan comprensible es el contenido publicado en términos de legibilidad. La métrica esta basadas en las calificaciones de legibilidad de Flesch, que son diseñadas para indicar la dificultad de comprensión cuando se lee un pasaje de Inglés contemporáneo [30]

**“Compleitud”**: Esta métrica se refiere a que tan completo es el contenido, y se expresa en función de los enlaces caídos, el número de enlaces internos y la longitud del artículo.

En la figura 3, se puede observar la descripción de las métricas Autoridad, Compleitud, Complejidad, Informatividad, Consistencia y Volatilidad, relacionadas a Wikipedia y que fueron descritas anteriormente.

**Figura 3.** Métricas de calidad para Wikipedia[29].

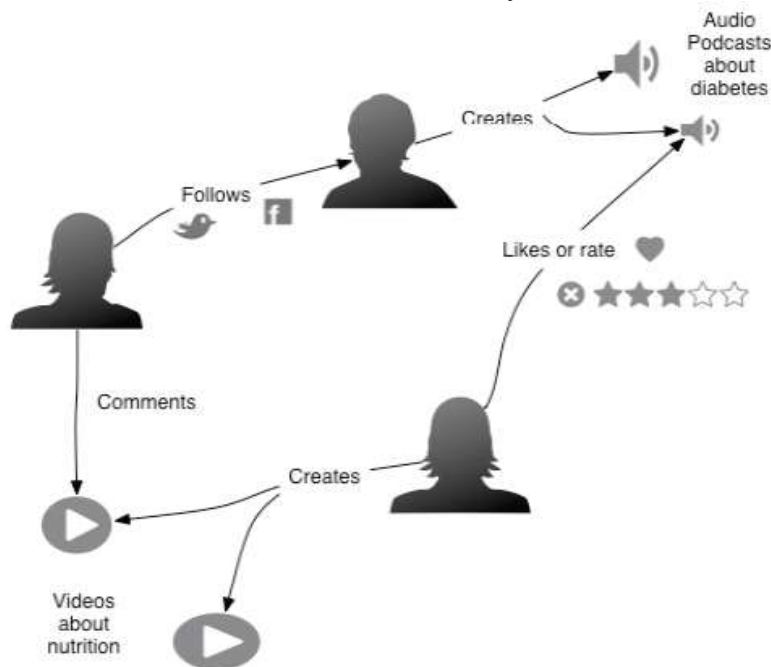
1. **Authority/Reputation** =  $0.2 * \text{Num. Unique. Editors} + 0.2 * \text{Total Num. Edits} + 0.1 * \text{Connectivity} + 0.3 * \text{Num. of Reverts} + 0.2 * \text{Num. External Links} + 0.1 * \text{Num. Registered User Edits} + 0.2 * \text{Num. Anonymous User Edits}$ .
2. **Completeness** =  $0.4 * \text{Num. Internal Broken Links} + 0.4 * \text{Num. Internal Links} + 0.2 * \text{Article Length}$
3. **Complexity** = *Flesch Readability Score*
4. **Informativeness** =  $0.6 * \text{InfoNoise} - 0.6 * \text{Diversity} + 0.3 * \text{Num. Images}$
5. **Consistency** =  $0.6 * \text{Admin. Edit Share} + 0.5 * \text{Age}$
6. **Currency** = *Currency*
7. **Volatility** = *Median Revert Time*

### 2.3.4.2.2.1. HealthTrust

De las métricas analizadas, la única construida específicamente para el dominio de la salud, es HealthTrust por lo tanto, se profundizara a fondo sobre su base teórica, y sobre su adaptación a una comunidad de contenido multimedia online de diabetes, ya que como se mostrará a continuación, esta métrica fue modelada para cualquier tipo de contenido y temática relacionada con el campo de la salud.

**Definición:** En [17] Luque y colaboradores, definen confianza dentro de su métrica HealthTrust, como “Confianza asegurada en la calidad de los usuarios y contenido dentro de una red social”. Por lo cual esta se puede modelar desde dos puntos de vista: qué tanto se confía, en un productor de contenido, o en el contenido como tal. En definitiva el valor de HealthTrust para una pieza de contenido dentro de una RSVS, o un RSV modelada como se muestra en la Figura 4, es la suma de la confianza en el autor, y el contenido.

**Figura 4.** Relaciones entre los contenidos y autores en una RSV[17].

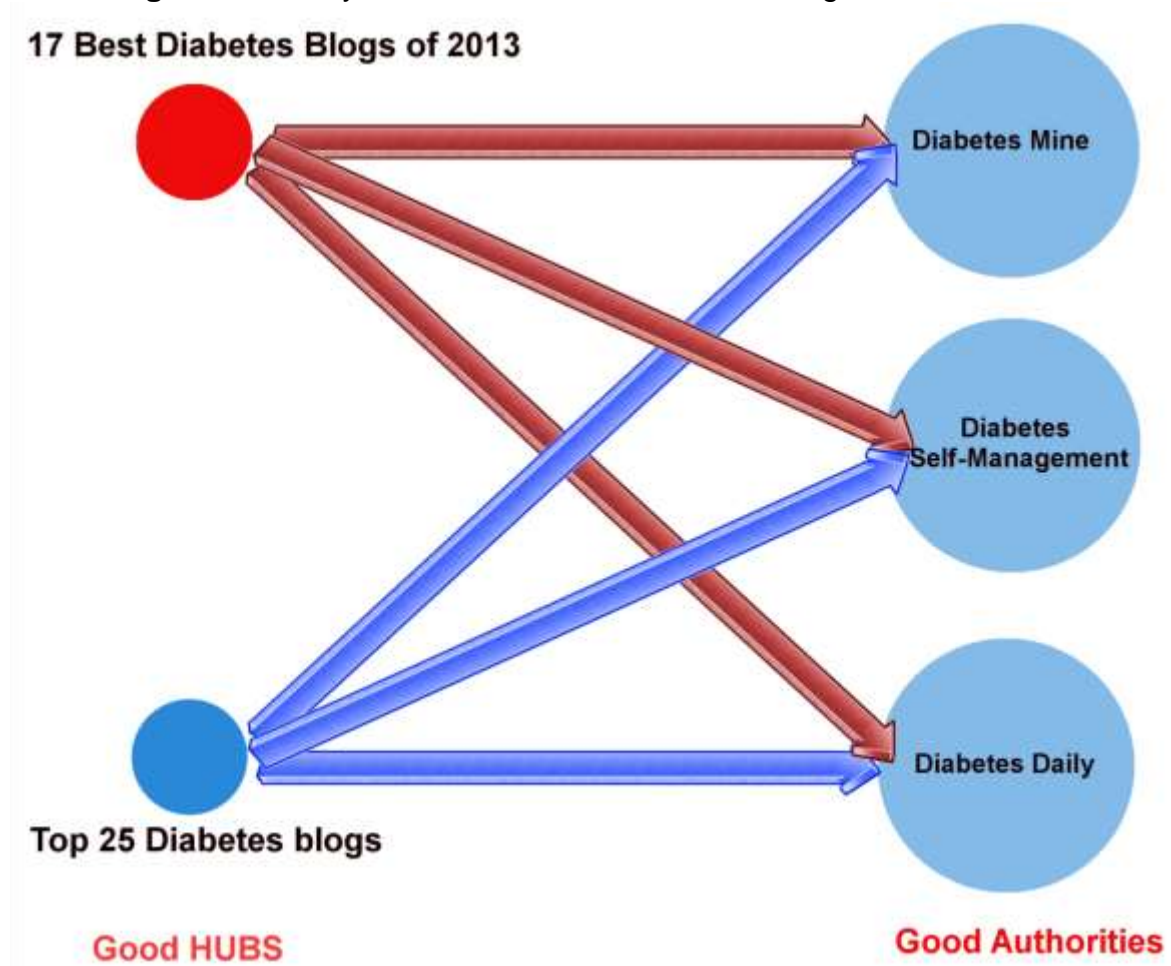


## 2.3.4.2.2. Algoritmo de la métrica de HealthTrust

### 2.3.4.2.2.1. Conceptos Generales

“Autoridades” y “Hubs”: Los “Hubs” son nodos dentro de un grafo social, los cuales no constituyen una fuente relevante de información, pero apuntan a otros nodos. Mientras que los nodos conocidos como “Autoridades”, contienen información relevante y son referenciados por muchos otros nodos, en especial por “Hubs”. Por otro lado como se observa en la Figura 5, los “buenos Hubs” son nodos que referencian a muchas “Autoridades”, mientras que las “Buenas Autoridades” son aquellas que son referenciadas por muchos “Hubs”[37].

**Figura 5.** Hubs y Authorities en el contexto de blogs sobre diabetes.



### 2.3.4.2.2.2. HITS

HITS, del Inglés “Hyperlink-Induced Topic Search” también conocido como “Hubs and authorities”, es un algoritmo de Análisis de enlaces desarrollado por Jon Kleinberg [37], el cual usa una aproximación simple para verificar la calidad de un Nodo dentro de una red, inicialmente asume que si un nodo A contiene un hipervínculo hacia un Nodo B, A piensa que B contiene información de calidad. Si



A apunta hacia muchos documentos de buena calidad entonces la opinión de A también se tiene en cuenta como muy importante[38].

En HITS se le asignan dos valores a cada nodo, uno como “Hub” y otro como “Autoridad”. Los valores no son iguales para todos los nodos de la red; si el enlace viene desde una Autoridad, este tendrá un valor más alto que el valor dado por un nodo con un nivel de prestigio menor. Para el caso particular de HealthTrust se escogen los valores de “Autoridad”, ya que se desea saber el impacto que tiene la propagación de la autoría desde los nodos definidos como autoridades por la red, y los nodos a los que estos enlazan.

### **Implementación de HITS dentro de HealthTrust**

El algoritmo HITS dentro de HealthTrust, se implementará por medio de la librería de Java (JUNG)[39]. La implementación de este algoritmo dentro de la librería sigue el proceso mostrado en la Figura 6, donde se puede observar que se hace la implementación tradicional de HITS, se asignan 2 valores a cada nodo “Hub” y “Autoridad”, luego se tiene en cuenta la suma de nodos sucesores y predecesores. Esto se refiere a los enlaces que salen y entran de cada nodo, para por ultimo normalizan los valores para que la suma de los cuadrados de cada uno sea igual a uno. Se realizan iteraciones sucesivas de este proceso hasta que los valores converjan.

**Figura 6.** Implementación de HITS dentro de la librería JUNG[40].

```
assign equal initial hub and authority values to each vertex
repeat
  for each vertex w:
    w.hub = sum over successors x of x.authority
    w.authority = sum over predecessors v of v.hub
  normalize hub and authority scores so that the sum of the squares of each = 1
until scores converge
```

#### **2.3.4.2.2.2.3. PageRank**

El algoritmo PageRank desarrollado originalmente en la Universidad de Stanford por Larry Page y Sergey Brin. Es un método basado en análisis de enlaces usado por Google para evaluar la “Relevancia” de los Sitios Web. PageRank permite asignar una posición a cada nodo dentro de un grafo social, teniendo en cuenta cuando un Sitio Web es enlazado por otro. Este término se conoce como “backlink”. Si un “backlink” proviene desde un Sitio Web importante, este da un mayor puntaje en el ranking generado por PageRank, con respecto a un Sitio menos importante[37] [41] [42].

El valor del PageRank de un nodo dado  $i$ , denotado como  $PR(i)$ , se encuentra definido recursivamente según la ecuación de la figura 7. Donde se tienen en cuenta todo los nodos  $j$  pertenecientes a la red, y que tienen conexión con  $i$ ;  $N(j)$  es el número total de enlaces originados desde el nodo  $j$ ;  $d$  es conocido como el

factor de amortiguación, es un número entre 0 y 1 y es usado para controlar el exceso de influencia de los nodos, y D es un distribución de probabilidad sobre todas los Sitios Web(Ej: uniforme).

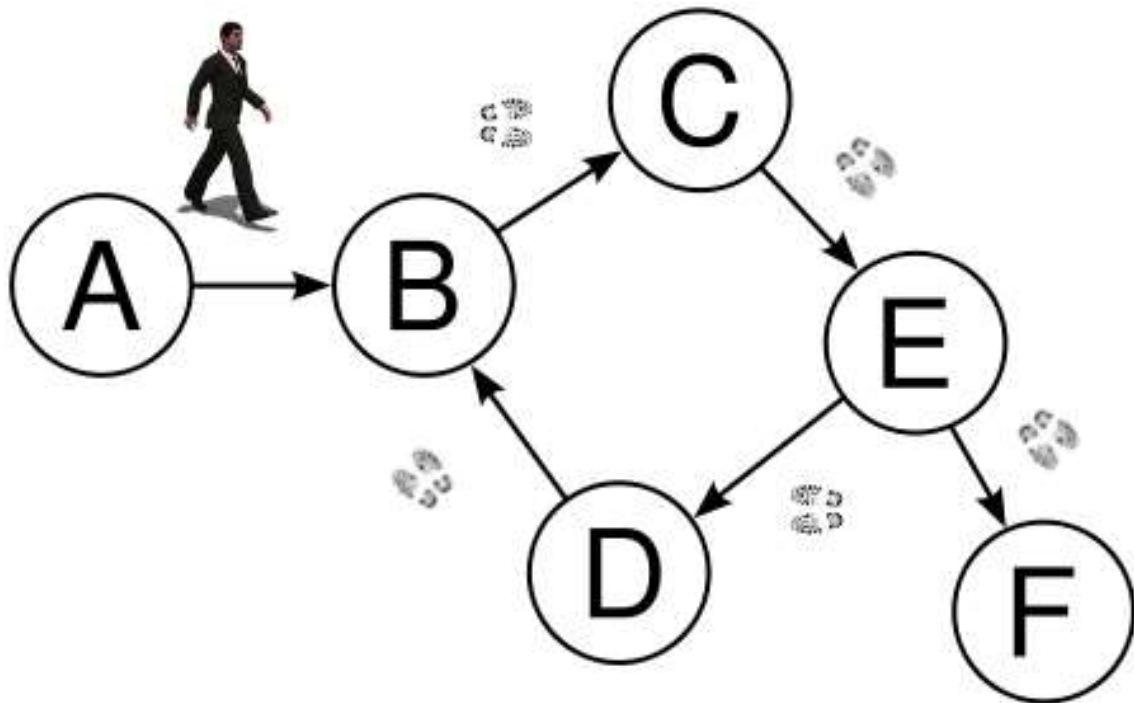
**Figura 7.** Ecuación de PageRank[42].

$$PR(i) = dD(i) + (1 - d) \sum_{j \rightarrow i} [PR(j) / N(j)]$$

### Implementación de PageRank dentro de HealthTrust

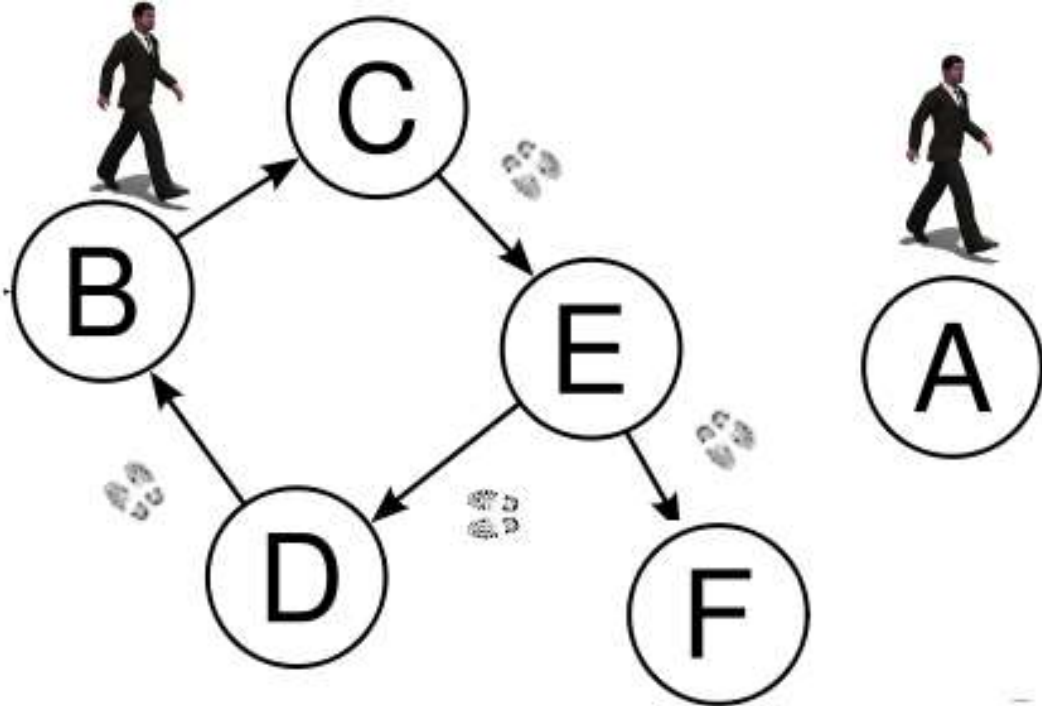
El valor de PageRank dado a cada nodo puede ser visto, como la fracción del tiempo que se tarda en visitar este nodo con respecto al tiempo total de recorrido del grafo social por parte de un caminante aleatorio como se puede observar en la Figura 8.

**Figura 8.** Algoritmo del caminante aleatorio



Como se puede observar en la figura 9, en algunas ocasiones los nodos no tienen conexiones entre sí, por lo cual se propone el modelo del caminante aleatorio con posibilidad de salto, este añade la probabilidad de saltar a un nodo cualquiera de la red cada vez que se recorre un conexión hacia otro nodo. Los valores típicos de la probabilidad de salto se encuentra entre 0.1 y 0.2 y si un nodo no tiene conexiones de salida como el nodo F, la probabilidad de salto es de 1.

**Figura 9.** Algoritmo del caminante aleatorio con probabilidad de salto.



#### **2.3.4.2.2.4. Contexto del algoritmo de la métrica de HealthTrust**

HealthTrust basa sus calificaciones en una adaptación de algoritmos de búsqueda como HITS[38] y PageRank[42], fundamentándose en la premisa de que las comunidades en línea de salud son efectivas filtrando contenido impreciso y desinformante, ya que un usuario que genere información de este tipo dentro de una comunidad de salud tiene pocas posibilidades de ganar prestigio dentro de esta comunidad.

Como se sugiere en [14] el efecto TKO genera un impacto negativo a la hora de aplicar algoritmos basados en análisis de enlaces. Luque y colaboradores[17] demuestran por medio del algoritmo de la métrica HealthTrust que este efecto es opuesto en el dominio de la salud, ya que el contenido relevante dentro de una RSV de videos no especifica en salud como YouTube, puede ser desplazado por videos altamente populares, más no necesariamente adecuados para temas de salud.

#### **2.3.4.2.2.3. Elección del campo de acción del algoritmo de la métrica de HealthTrust**

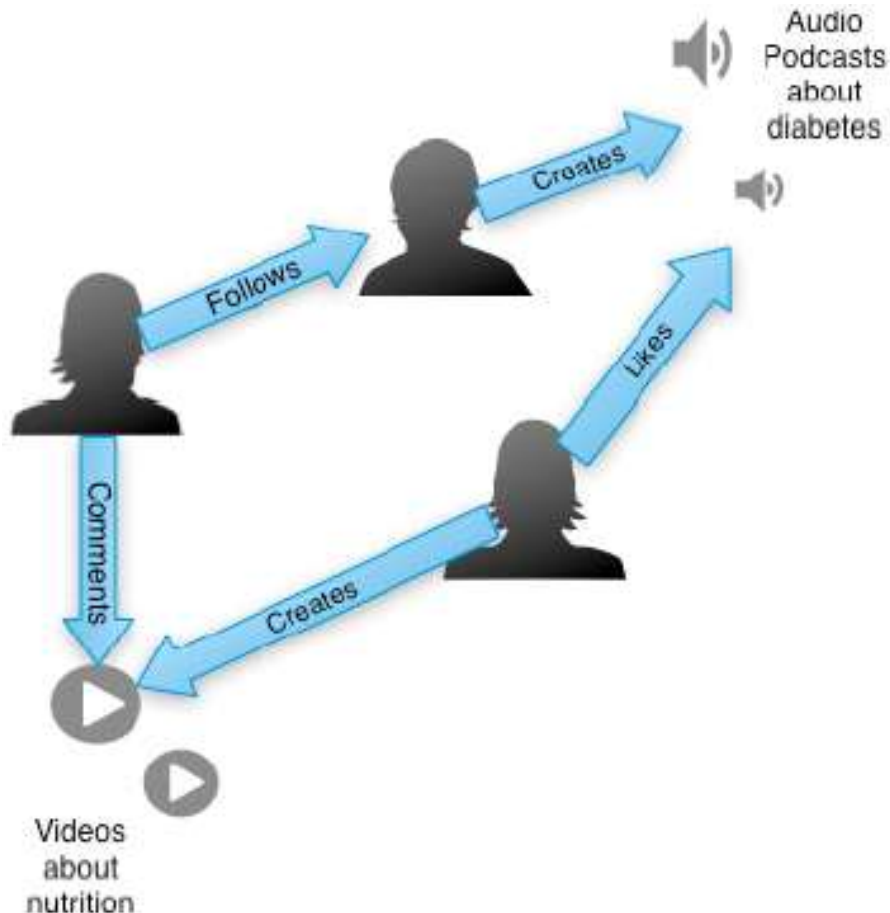
La elección de aplicar HealthTrust a una comunidad online de videos de diabetes tiene como fundamento el alto número de pacientes en el departamento del Cauca con esta enfermedad, y también la vinculación de la línea de investigación en e-salud de la Facultad de Electrónica y Telecomunicaciones de la Universidad del Cauca con proyectos sobre el tema, pertenecientes a la facultad de Ciencias de la

Salud. Por último se seleccionó YouTube, porque es la RSV de videos más popular actualmente.

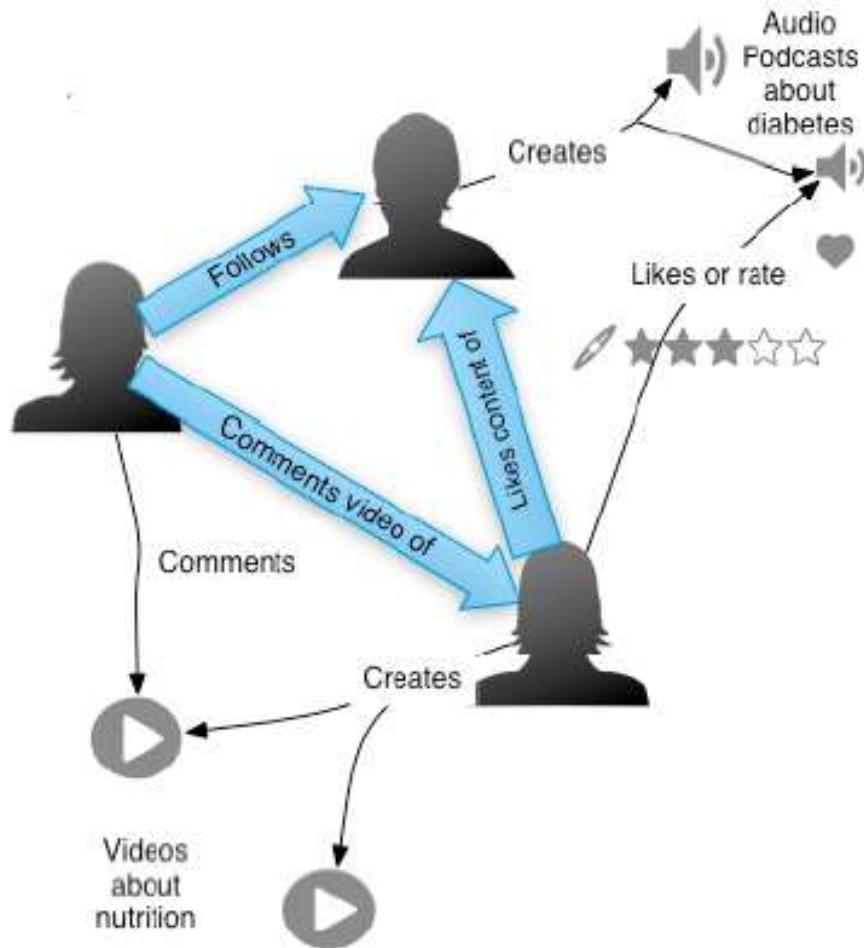
**Pasos requeridos para la elaboración del algoritmo de HealthTrust:**

- 1) El primer paso es definir la RSVS donde se va a aplicar el algoritmo. Para nuestro caso la comunidad de diabetes dentro de la RSV YouTube.
- 2) Conformar un set de datos, con información de la comunidad, por medio de la API de datos de YouTube, lo que incluye, todas las relaciones sociales de los canales y videos (Ej: suscripciones, favoritos, comentarios, relaciones de amistad).
- 3) Calcular los valores de HITS “authoritative” o PageRank para todos los miembros de la comunidad definida (contenido figura 10 y autor figura 11). Como se dijo anteriormente, estos valores representan la confianza en la fuente dentro de la comunidad.

**Figura 10.** Modelado de HealthTrust para contenido [17].



**Figura 11.** Modelado de HealthTrust para el autor [17].



4) Por último, se conforma la valoración para cada pieza de contenido, como se observa en la figura 12, donde el factor de herencia (inheritance factor) es definida por Luque y colaboradores[17] como 0.7, teniendo como criterio que el 70% de la valoración de la métrica es debida a la confianza en el autor y el 30% restante al valor del contenido .

**Figura 12.** Algoritmo de la métrica HealthTrust[17].

*HealthTrust (content, community)*

$$= \text{Authoritativeness} (\text{content}, \text{community}) \times (1 - \text{InheritanceFactor})$$

$$+ \text{Authoritativeness} (\text{author of content}, \text{community}) \times \text{InheritanceFactor}$$

#### 2.3.4.2.2.4. Complejidad del algoritmo de HealthTrust

Para poder calcular la complejidad del algoritmo de HealthTrust, primero tenemos que analizar la complejidad de los algoritmos de las métricas sociales que la

componen, para lo cual analizaremos el mejor y el peor de los escenarios de aplicación de los mismos.

### Peor Escenario

como lo afirma[43], la complejidad de los algoritmos de las métricas sociales HITS y PageRank, está dada por :

*Complejidad de HITS =  $O(n * m)^k$  Complejidad de PageRank =  $O(n * m)$*   
 Donde n, m respectivamente se refieren al número de nodos y al número de ejes del grafo social. En el caso de la complejidad de HITS, se tiene en cuenta el número de iteraciones del algoritmo para llegar a un valor convergente denotado con la letra k.

A continuación se realiza una generalización para el algoritmo HealthTrust basado en HITS:

$$Complejidad\ algoritmo\ de\ HealthTrust(HITS) = O((q * m)^k (p * m)^k)$$

En el cual se tiene en cuenta los modelamientos de Luque para los grafos de Contenido y Autor explicados en la sección anterior. Donde q se puede expresar como el número de nodos Autores y nodos Contenido  $q = na + nc$ , mientras que p puede ser expresado como el número de nodos Autores  $p = na$ .

$$Complejidad\ algoritmo\ de\ HealthTrust(HITS) = O(((na + nc) * m)^k ((na) * m)^k)$$

$$Complejidad\ algoritmo\ de\ HealthTrust(HITS) = O((nc * m)^k (na * m)^{2k})$$

Siguiendo un proceso análogo encontramos que la complejidad para el algoritmo de HealthTrust basado en PageRank es:

$$Complejidad\ algoritmo\ de\ HealthTrust(PageRank) = O((nc * m)(na * m)^2)$$

### Mejor Escenario

El mejor escenario de la aplicación de los algoritmos ocurre cuando ya se tiene la valoración de la mayoría de los nodos de la red, y simplemente se calculan los valores de nuevos sets de nodos sin recalculan las valoraciones de toda la red. Esta complejidad estaría dada por:

*Complejidad de HITS =  $O(n + m)^k$  Complejidad de PageRank =  $O(n + m)$*   
 Luego de realizar un análisis similar al descrito en el punto anterior tenemos

$$Complejidad\ algoritmo\ de\ HealthTrust(HITS) = O((na + nc + m)^k (na + m)^k)$$

$$Complejidad\ algoritmo\ de\ HealthTrust(PageRank) = O(mc(na * m) + (na + m)^2)$$

De lo anterior se puede concluir que el tiempo empleado por el algoritmo de HealthTrust basado en HITS es mayor para los dos escenarios planteados

### 2.3.4.3. Estado pretendido del conocimiento (Pronóstico)

Del análisis y selección realizada con anterioridad podemos concluir que los criterios, atributos y métricas más adecuados para ser aplicados en un RSVS son los que se encuentran referenciados en la tabla resumen siguiente:

**Tabla 8.** Atributo(s) y Métrica(s) seleccionadas.

Atributos	Métricas
Autoría	HealthTrust
Confianza	
Precisión	
Complejidad	
Legibilidad	

### 2.3.5. Fase de extensión y publicación:

La fase de extensión y publicación cuenta con las siguientes actividades para llevarse a cabo.

- Como ya se había mencionado en el anteproyecto, la escritura del capítulo dos de esta monografía titulado “Síntesis de la base del conocimiento asociada a atributos y métricas de calidad en la Web 2.0”.
- Se procederá a escribir un artículo de revisión.
- Se realizó una participación en la modalidad poster y exposición en el congreso internacional de informática médica y telesalud organizado por la fundación Cardio Infantil y ACIESA en la ciudad de Bogotá los días 28-30 de mayo de 2013.

## 2.4. Conclusiones

En este capítulo se describió el proceso llevado a cabo para la selección de atributos y métricas para la gestión de calidad de la información en RSVS. Del análisis y selección realizada se puede concluir que los criterios más adecuados para ser aplicados en un RSVS son: Autoría, Confianza, Precisión, Complejidad y Legibilidad. Y los únicos que han sido implementados por medio de una herramienta software son los de Autoría y Confianza, a través de la métrica HealthTrust. En el siguiente capítulo se describe el proceso de desarrollo de un servicio web que apoya a la moderación de contenidos para una RSVS, soportado en la métrica HealthTrust.

## CAPITULO 3

# Desarrollo e implementación del servicio de moderación de contenidos en una RSVS

En este capítulo se desarrolla el servicio web que apoya a la moderación de contenidos para una RSVS. Es importante aclarar que este servicio no consiste en un buscador de contenido, sino una herramienta que permite valorar los resultados de las búsquedas, ya sea por autor o contenido por medio de la consulta de un servicio web. Este capítulo se concentra en su mayoría en la construcción del servicio utilizando la Ingeniería de requisitos para servicios telemáticos. Las fases restantes del desarrollo del servicio se presentan en el Anexo B, usando el Proceso Unificado de Rational (Rational Unified Process, RUP).

### 3.1 Sistemas de moderación de contenidos en RSVS

A continuación se caracterizan los diferentes tipos de mecanismos de moderación en RSVS.

Red Social Virtual de salud con moderadores: En este tipo de RSVS, los moderadores tienen la tarea de prevenir la avalancha de información negativa (estas consisten en terapias no adecuadas, e información médica imprecisa), como también prevenir comportamientos inadecuados que vayan en contravía de los comportamientos definidos como saludables. Un ejemplo común es que un miembro muy popular de una RSVS de diabetes que motive de alguna manera la ingesta de refrescos con alto contenido de azúcar, o que promueva conductas irrespetuosas dentro de los canales de comunicación[44].

El tipo de RSVS descrita anteriormente, cuenta con la dificultad que la moderación debe realizarse en tiempo real, y la mayoría de las RSVS como las anteriormente descritas hacen parte de organizaciones sin ánimo de lucro, y los moderadores son voluntarios.

Red Social Virtual de salud con participación activa de profesionales de la salud: generalmente consisten en foros o plataformas de discusión y los médicos participantes cuentan con un horario determinado para responder las preguntas. La moderación de la información en este caso en particular se basa en la disponibilidad horaria de los profesionales de la salud, que generalmente hacen parte de un proyecto y cuenta con recursos limitados.

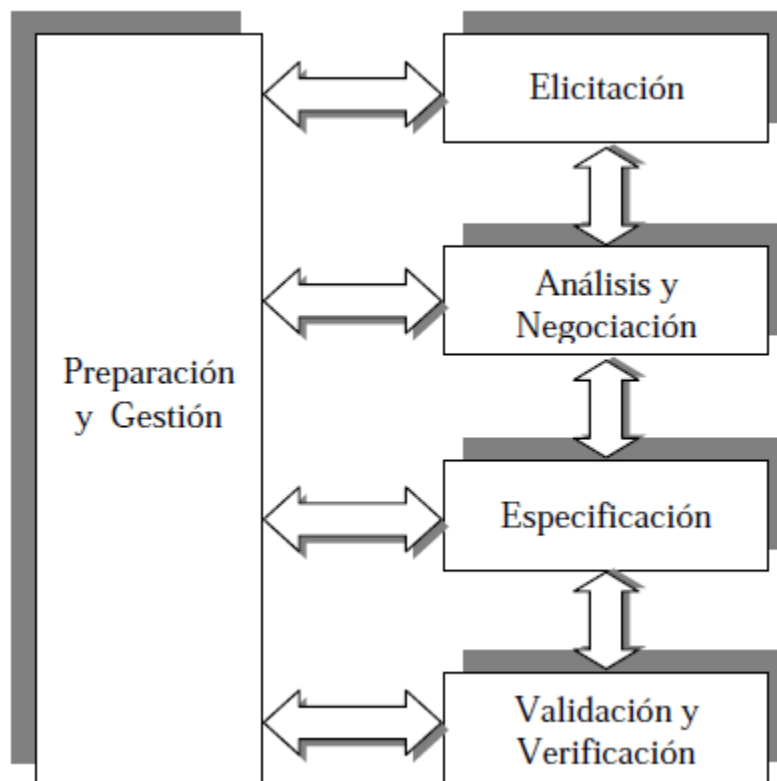


Red Social Virtual con moderación por parte de los usuarios: En este tipo de RSVS, son los mismos usuarios que generalmente cuentan con un grado alto de experticia en el tema específico, que moderan los contenidos de acuerdo a las prestaciones de la plataforma para ese fin. Por ejemplo: sistema de calificaciones, sistema de reporte de SPAM, etc.

### 3.2 Aproximación Metodológica

Como referencia metodológica, tal como se mencionó anteriormente, se utilizara una versión simplificada de RUP, la cual es una metodología de Ingeniería de Software que busca asegurar productos software de calidad. Dos de las disciplinas del RUP son **a) el Modelado de la organización** y **b) Captura de requisitos**. Para el desarrollo de estos dos componentes se utilizara como metodología base la Ingeniería de Requisitos la cual permite hacer una construcción formal y adecuada del servicio de moderación a utilizar en una RSVS. Esta metodología de Ingeniería de Requisitos se describe en la Figura 13 [45].

**Figura 13.** Aproximación esquema de AMIR-ST<sup>3</sup>[45].



A continuación se describen las disciplinas restantes, abordadas durante el desarrollo del servicio:

<sup>3</sup>Aproximación Metodológica para la Ingeniería de Requisitos de Sistemas Telemáticos

**Disciplinas de Análisis y Diseño:** En este componente se presenta la arquitectura lógica, los diagramas de clases, de paquetes, lo cuales se encuentran descritos a fondo en el anexo B. El propósito concreto de esta etapa es una descripción de la funcionalidad del sistema, además de las tecnologías, software y hardware que se utilizarán para hacer realidad el servicio.

**Implementación:** en esta disciplina se muestran las interfaces del servicio, la implementación del servicio web RESTful y el diagrama de base de datos.

**Pruebas:** El plan de pruebas se realizara teniendo en cuenta la metodología de evaluación más adecuada, que resultara de un análisis entre varias metodologías de evaluación de software propuestas por DESMET.

**Puesta en servicio:** esta etapa consiste en la puesta en funcionamiento del servicio web que permita obtener una valoración de un video o autor, con respecto a la comunidad de diabetes de YouTube.

A continuación se describen en detalle los diferentes modelos obtenidos durante el desarrollo del servicio.

### **3. 3 Modelado de la organización y Captura de requisitos**

#### **3.3.1 Preparación y Gestión**

Esta etapa fue realizada casi en su totalidad en la fase del anteproyecto de trabajo de grado, y tienen que ver con el planeamiento inicial para el desarrollo de cada una de las actividades propuestas, así como la gestión del tiempo por medio de cronogramas y flujos de trabajo.

##### *3.3.1.1 Objetivo principal:*

Elaborar el Plan de Trabajo para la realización de la fase de captura de Requisitos para un servicio de moderación de contenidos en una RSVS.

##### *3.3.1.2 Actividades y Técnicas de representación:*

Las etapas y entregables correspondientes: *Conformación del equipo de trabajo, elaboración del plan de trabajo, establecimiento de los criterios de evaluación.* Los dos primeros entregables ya fueron descritos paso a paso en el anteproyecto de este mismo trabajo de grado, y en el siguiente capítulo se analizará a fondo la selección y ejecución de la evaluación del servicio.

#### **3.3.2 Elicitación**

En esta etapa se realiza un estudio del dominio del problema, desarrollando una interacción con el cliente y los usuarios que permita relacionar las necesidades de los mismos. Se realizara un estudio de los modelos de negocio existentes, el análisis de los modelos de negocios propuestos, teniendo en cuenta a los clientes y usuarios implicados para la correspondiente re alimentación para el mejoramiento del sistema.

Es importante destacar que, aunque el fin de este trabajo de grado no es elaborar un servicio de moderación y publicación de contenidos, es importante para el mismo modelar este servicio para poder establecer cómo puede hacer parte el servicio web de evaluación de contenido, dentro de este contexto.

### 3.3.2.1 Objetivo Principal:

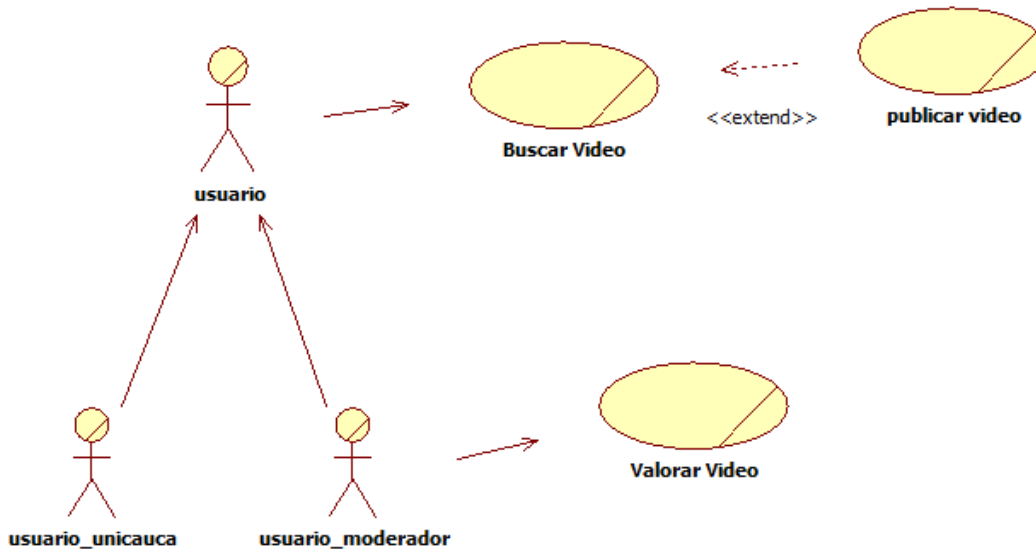
Descubrir por medio de la interacción inicial con el cliente, los procesos actuales de la organización para identificar las necesidades futuras de los mismos en una RSVS, tanto en su “parte frontal” como en la aplicación para gestión del negocio.

### 3.3.2.2 Actividades y técnicas de la representación:

#### 3.3.2.2.1 Obtención del Modelo de negocio

Por medio del uso de los estereotipos de organización de UML, se planteó el Modelo del negocio de la forma en la cual se realizan las publicaciones en la RSVS Unisalud a la fecha de inicio de este proyecto. Los casos de uso del negocio que se observan en la figura 14 y están explicados a detalle en la tabla 9.

**Figura 14.** Caso de uso del negocio, modelo inicial de publicación de videos.



**Tabla 9.** Análisis correspondiente al caso de uso “Buscar video”

<b>Caso De Uso</b>	Buscar video
<b>Actor</b>	Usuario
<b>Autor</b>	Daniel Eduardo Cañón
<b>Fecha</b>	20/04/2013
<b>Descripción Breve</b>	El usuario busca por sus propios medios un video a ser publicado dentro de la RSVS
<b>Condiciones Anteriores</b>	

<b>Condiciones Posteriores</b>	El usuario copia la URL en la que se encuentra el video
<b>Flujo de Eventos</b>	<ol style="list-style-type: none"> <li>1.El usuario ingresa a una RSV de videos como "YOUTUBE" y realiza una busqueda de un video.</li> <li>2. El usuario selecciona un video del ranking desplegado en la busqueda.</li> </ol>

**Tabla 10.** Análisis correspondiente al caso de uso "Valorar video"

<b>Caso De Uso</b>	Valorar video
<b>Actor</b>	Usuario_moderador
<b>Autor</b>	Daniel Eduardo Cañón
<b>Fecha</b>	20/04/2013
<b>Descripción Breve</b>	El usuario valora la calidad de un video manualmente por medio de una herramienta de excel que cuanta con criterios de evaluación anteriormente definidos.
<b>Condiciones Anteriores</b>	Seleccionar uno o varios videos para evaluar.
<b>Flujo de Eventos</b>	<ol style="list-style-type: none"> <li>1. El usuario accede a la herramienta de excel</li> <li>2. El usuario asigna valores a los criterios de evaluación</li> <li>3. El usuario toma una decisión basado en los resultados de la valoración</li> </ol>

**Tabla 11.** Análisis correspondiente al caso de uso "Publicar video"

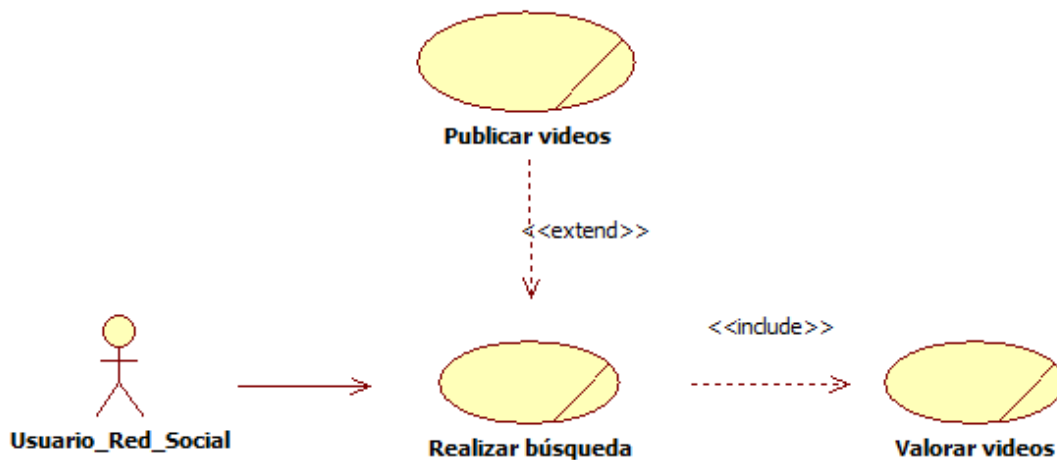
<b>Caso De Uso</b>	Publicar Video
<b>Actor</b>	Usuario
<b>Tipo</b>	Primario
<b>Autor</b>	Daniel Eduardo Cañón
<b>Fecha</b>	20/04/2013
<b>Descripción Breve</b>	El Usuario publica de un video dentro de la RSVS de unisalud, llenando manualmente todos los campos de

	información requeridos.
<b>Condiciones Anteriores</b>	Haber realizado el casos de uso "Buscar video"
<b>Flujo de Eventos</b>	<ol style="list-style-type: none"> <li>1.El usuario introduce la dirección del video encontrado en el caso de uso "Buscar video"</li> <li>2. El usuario llena manualmente todos los campos de nformación requeridos.</li> <li>3. El usuario presiona el boton publicar en la interface respectiva.</li> </ol>

### 3.3.2.3 Identificación de las necesidades de clientes y usuarios

Se describe de manera preliminar los requisitos que la RSVS Unisalud debe implementar para soportar los procesos de negocios de la organización de manera automática o semiautomática, según la validación a realizar con clientes y usuarios de la misma.

**Figura 15:** Caso de uso del negocio, modelo parcial para la publicación de videos



**Tabla 12.** Análisis correspondiente al caso de uso del servicio “Realizar Búsqueda”

<b>Caso De Uso</b>	Realizar búsqueda
<b>Actor</b>	Usuario
<b>Tipo</b>	Primario
<b>Autor</b>	Daniel Eduardo Cañón
<b>Fecha</b>	20/04/2013
<b>Descripción Breve</b>	Por medio de este caso de uso es posible ingresar todos los datos necesarios para buscar un video dentro de la RSV de videos.

**Tabla 13.** Análisis correspondiente al caso de uso del servicio “Valorar video”.

<b>Caso De Uso</b>	Valorar video
<b>Actor</b>	Usuario
<b>Tipo</b>	Primario
<b>Autor</b>	Daniel Eduardo Cañón
<b>Fecha</b>	20/04/2013
<b>Descripción Breve</b>	Por medio de este caso de uso es posible evaluar si un video dentro de una RSV es adecuado para la publicación en la RSVS de unisalud. Esto se realiza por medio de un servicio web y la valoración de HealthTrust.

**Tabla 14.** Análisis correspondiente al caso de uso del servicio “Publicar video”

<b>Caso De Uso</b>	Publicar Video
<b>Actor</b>	Usuario
<b>Tipo</b>	Primario
<b>Autor</b>	Daniel Eduardo Cañón
<b>Fecha</b>	20/04/2013
<b>Descripción Breve</b>	El usuario según su criterio y la ayuda de las valoraciones de HealthTrust decide cual video publicar en la RSVS

### 3.3.3 Análisis y negociación

En esta etapa se estudiará la información obtenida en la fase inmediatamente anterior. Además se llevará a cabo una negociación con los posibles clientes del servicio de moderación teniendo en cuenta la primera aproximación para llevar a cabo un refinamiento de los procesos de negocio.

#### 3.3.3.1 Objetivo Principal

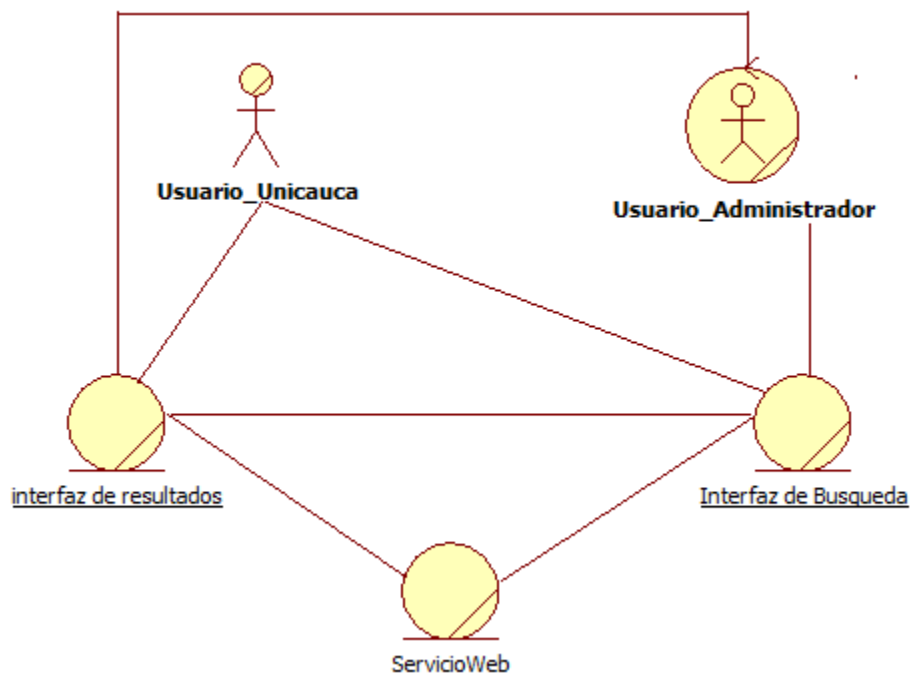
Profundizar en el conocimiento de la organización para asegurar un acuerdo con los posibles clientes sobre los procesos dentro de la organización que el servicio de moderación va a dar soporte.

### 3.3.4 Actividades y técnicas de representación

#### 3.3.4.1 Obtención del modelo del dominio

Consiste en moldear el ambiente en el cual trabajará el servicio de moderación teniendo en cuenta lo definido en el numeral 3.3.2 y así obtener el Modelo de Dominio para la organización, el cual a su vez consta de conceptos (objetos) y sus respectivas relaciones. Las fuentes en las cuales se basará su construcción son los objetos de información que fluyen entre las actividades en cada uno de los procesos de negocio.

Figura 16. Modelo del dominio parcial para la publicación de videos

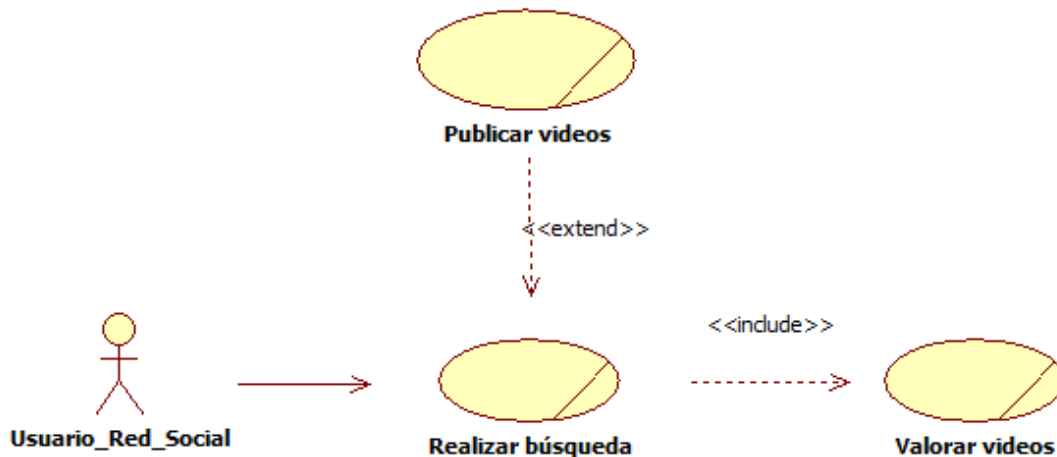


#### 3.3.4.2 Obtención del modelo de caso de uso del Negocio propuesto

Este proceso de refinamiento del caso de uso del Negocio es opcional, dependiendo de las necesidades de cada proyecto. Se procederá a discutir con

los posibles clientes modificaciones a los procesos identificados en la etapa anterior y que son candidatos a ser soportados por el servicio de moderación. Para ello se utilizan los mismos elementos de notación empleados en la descripción del Modelo de Negocios elaborado en la etapa anterior.

**Figura 17.** Caso de uso del negocio, modelo para la publicación de videos



### 3.3.4.3 Análisis de Requisitos de Información, Funcionales y no funcionales

Estos requisitos son centrados solamente en el caso de uso valorar videos.

#### 3.3.4.3.1 Requisitos de información

- El sistema debe almacenar una base inicial de los canales/autores de contenido de la RVS YouTube, para los cuales se asume que los contenidos son de calidad ej: canal de videos de la CDC, asociaciones de Diabetes, etc.
- El sistema debe almacenar los videos relacionados a cada uno de los canales, incluyendo sus relaciones: reproducciones, me gusta (likes), no me gusta (dislikes), favoritos, suscripciones, y comentarios.

#### 3.3.4.3.2 Requisitos funcionales

- El servicio web debe permitir búsquedas de valoraciones correspondientes a autores o videos de salud, relacionados en la red social YouTube. Estas búsquedas deben ser basadas en una o varias métricas definidas con anterioridad, las cuales tienen como finalidad brindarle un alto de grado de confiabilidad a la información publicada por el usuario.

#### 3.3.4.3.3 Requisitos no funcionales

- Se desarrollara un servicio web REST, que permitirá ser consumido desde cualquier tipo de plataforma.



- El servicio Web, se desarrollara en java, usando la API de YouTube para realizar un análisis del grafo social de los canales y sus diferentes relaciones.

#### **3.3.4.3.4 Negociación de requisitos**

Para llevar a cabo esta etapa se realizó una retroalimentación con los posibles clientes del servicio, divididos en dos actividades lo cual permite modelar los casos de uso del negocio desde tres puntos de vista usuario, moderador y experto:

- Usando la técnica de grupos focales como se resume en el Anexo C, se procedió a realizar una actividad con dos (2) grupos de estudiantes de comunicación social. Los cuales cuentan con un perfil como creador y moderador de contenidos dentro de una RSV.
- Se realizó una entrevista a un experto la enfermera Martha Ortega Alegría, quien se encargó de realizar la evaluación manual de los contenidos publicados con anterioridad en la RSVS Unisalud. Esto permitió la retroalimentación del servicio, teniendo en cuenta un punto de vista por parte de un experto.

### **3.3.5 Especificación**

#### **3.3.5.1 Objetivos**

Documentar cada uno de los requisitos elicitados, usando todos los métodos necesarios y pertinentes, para que las personas directamente beneficiadas tengan una mejor comprensión de los mismos.

#### **3.3.5.2 Actividades y técnicas de representación:**

##### **3.3.5.2.1 Priorización de los requisitos del sistema**

A continuación se encuentran listados uno a uno los requisitos del sistema en orden de importancia y dificultad.

- Servicio Web “ranking de canales de videos en salud de la red social YouTube, basados en la métrica HealthTrust”.
- El sistema debe brindar una búsqueda o sistema de organización avanzada.
- Sugerir videos publicados con anterioridad en la red social, ya sea por medio de la métrica HealthTrust o la publicación manual en la red social.
- Permitir la evaluación manual de los contenidos por parte del usuario.

##### **3.3.5.2.2 Descripción de los requisitos para el sistema**

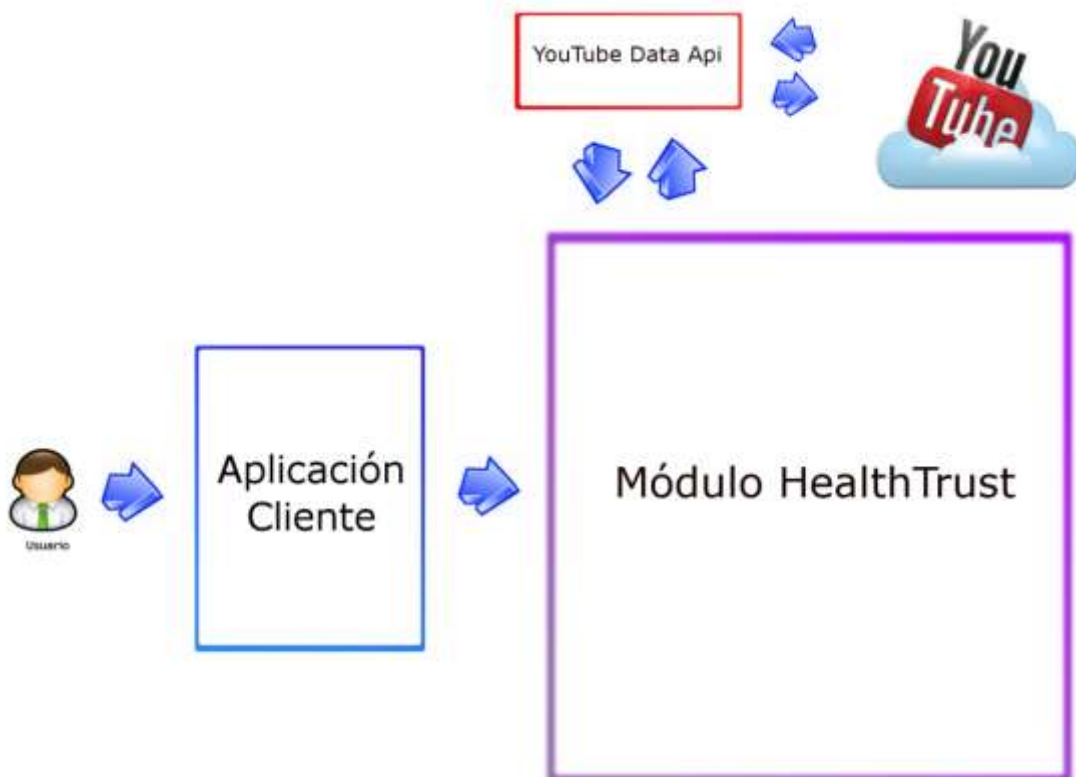
- **Servicio Web:** Este servicio Web desarrollado en java, retorna la valoración de un video o autor basado en la métrica descrita en el capítulo anterior HealthTrust. Para llevar a cabo esta tarea se utiliza el API de YouTube y se construye un grafo social a partir de la información recolectada.

### 3.4 Análisis y Diseño

#### 3.4.1 Prototipo

En los numerales anteriores de este capítulo se modeló y describió como debería ser un sistema completo de gestión de contenidos para una RSVS. En la figura 18 se presenta la arquitectura lógica más general, donde el usuario interactúa con la RSVS por medio de su interfaz gráfica, la cual puede estar construida con tecnologías del lado del cliente o Front-END como (HTML, HTML 5, PHP, JSF, Java, Etc.), o ser parte de una plataforma para construcción de RSV, como ELGG. Dentro de la aplicación cliente también se encuentra toda la lógica necesaria para consumir el servicio web RESTful. Esta se hace por medio del protocolo HTTP y tiene las funcionalidades (PUT, POST, GET, DELETE, LIST). Las respuestas del mismo se encuentran en formato XML, y JSON, como se puede observar en la figura 20. Por otro lado el módulo HealthTrust es el que se encarga de recuperar todos los datos sociales de los autores y contenido, para posteriormente generar el grafo y su respectiva valoración. La razón principal por la cual el API de YouTube se encuentra fuera de este módulo, es resaltar que este servicio se puede adaptar para cualquier tipo de contenido dentro de una RSV, no solamente videos dentro de la RSV YouTube.

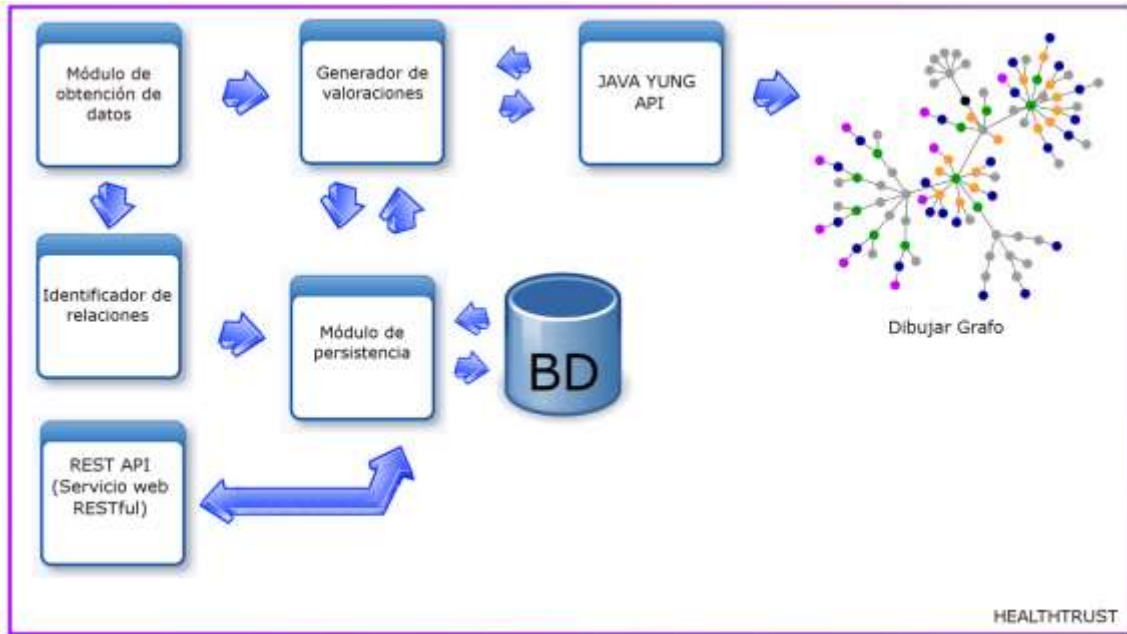
**Figura 18.** Arquitectura lógica general para un sistema de moderación de videos en una RSVS



### 3.4.2 MODULO HEALTHTRUST

En la figura 19, se puede observar una visión más detallada de los componentes de este módulo, los cuales se describen a continuación:

**Figura 19.** Arquitectura lógica módulo HealthTrust



**Módulo de obtención de datos:** Este módulo se encarga de hacer todas las peticiones de los atributos sociales de cada contenido y autor relacionado (autores de los comentarios, suscripciones, favoritos...etc.) a la API de datos de YouTube. Posteriormente esta información se pasa al identificador de relaciones, e informa al generador de valoraciones que ya está lista la información para generar el grafo social.

**Generador de valoraciones:** El generador de valoraciones se encarga de armar el grafo social, por medio de la información que se encuentra almacenada en la base de datos y que es recuperada usando el módulo de persistencia. Posteriormente esta información es entregada a la Java Jung API. Por último el generador de valoraciones se encarga de armar las calificaciones de cada contenido/autor, para posteriormente persistirlo.

**Java Jung API:** Esta API recibe la información de los nodos y las relaciones del grafo social por parte del generador de valoraciones, genera el grafo social y calcula los valores de HITS y PageRank para cada nodo (en el anexo B, se hace una descripción detallada de la API), luego posteriormente grafica este grafo social, y sus relaciones.

**Identificador de Relaciones:** Este módulo se encarga de identificar las relaciones entre los nodos existentes dentro de la red social. Por ejemplo, a que otros nodos está suscrito, si ha comentado videos de otros autores, si tiene videos de salud

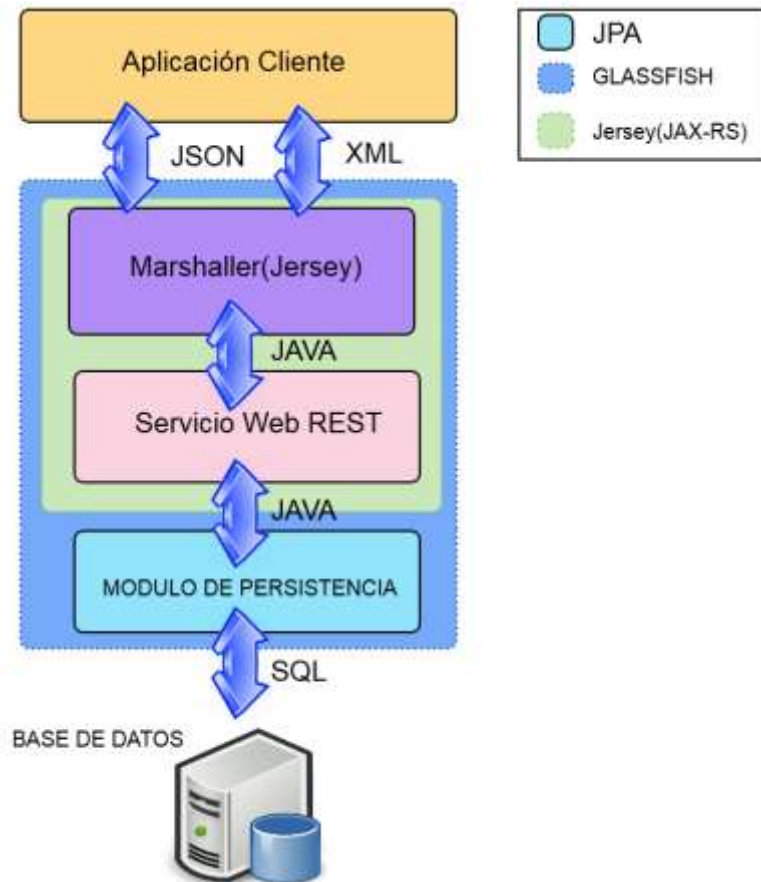
entre sus favoritos. Luego de verificar estas relaciones, envía esta información al siguiente modulo para ser persistida.

**Módulo de persistencia:** Este módulo está compuesto por las entidades de persistencia: Video, Canal, y la unidad de persistencia por medio de la utilización del API de persistencia de JAVA (JPA), se encarga de hacer el mapeo objeto-relacional de la información que se quiere persistir, y también de la que ya se encuentra en la base de datos.

**REST API (servicio web RESTful):** Este servicio tiene como fin permitir que la información de las valoraciones del algoritmo HealthTrust, pueda ser consumida fácilmente.

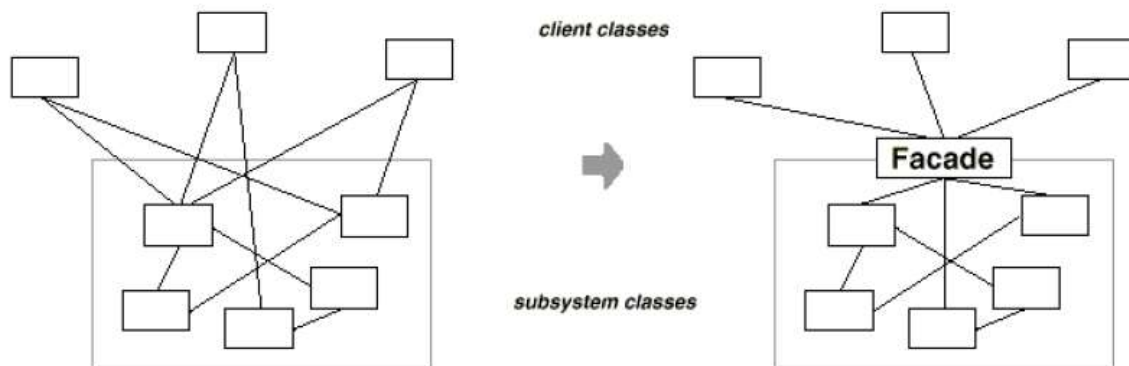
Como se especifica en el anexo B, los servicios web Restful, son una opción más liviana a su contraparte basada en SOAP, ya que fueron pensados para enviar datos directamente usando el protocolo HTTP. En la figura 20, se puede observar la arquitectura de un servicio RESTfull con JAX-RS y Jersey. La funcionalidad de crear este tipo de servicios desde entidades JPA (Java Persistence API), o bases de datos existentes se encuentra incluida por defecto en los Wizards del entorno de desarrollo NetBeans.

**Figura 20. Arquitectura de un servicio REST con JAX-RS**



El servicio web REST está diseñado usando el patrón de diseño Façade [46], el cual consiste en brindar una clase padre o genérica a un sistema más complejo (figura 21). Para el caso del servicio web, se crea una clase AbstractFaçade, que presenta una funcionalidad estándar que después es reutilizada en las clases CanalFaçadeREST y VideoFaçadeREST. En esencia, estas dos clases funcionan como un envoltorio para cada una de las clases de entidad (Video y Canal), lo que permite acceder a la base de datos desde cualquier tipo de aplicación del lado del cliente por medio de peticiones HTTP.

**Figura 21.** Patrón de diseño Façade [46].



**BD (Base de Datos):** Esta es la base de datos que contiene toda la información correspondiente a las valoraciones de HealthTrust, así como toda la información de las relaciones sociales de cada autor/contenido recolectado por medio del módulo HealthTrust.

### 3.5 Implementación

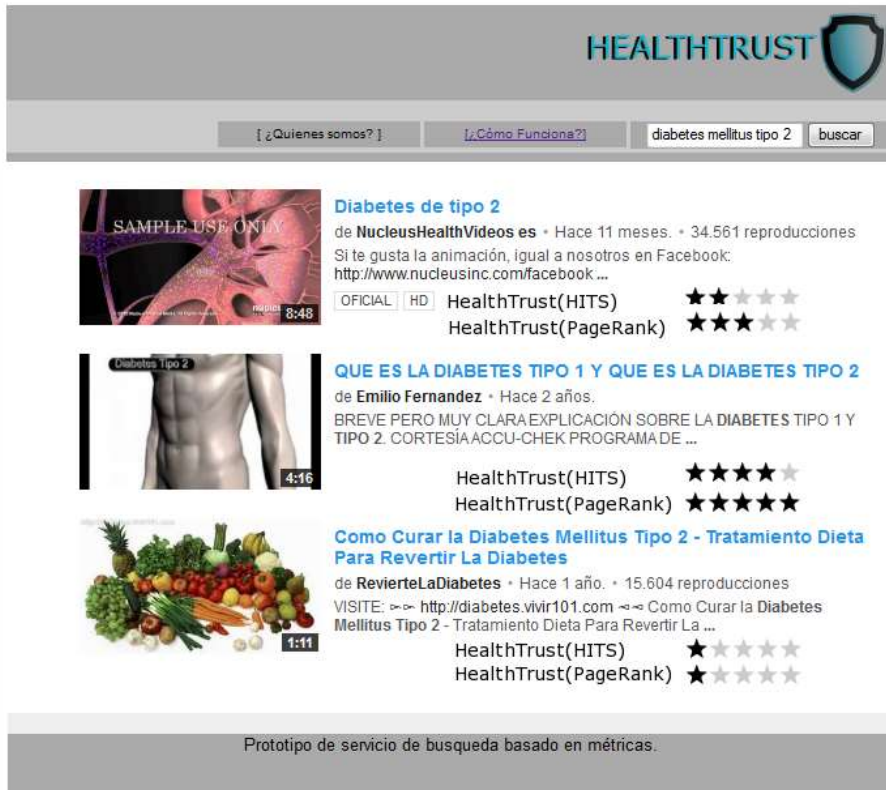
#### 3.5.1 Interfaces web

Las figuras 22, 23, 24 corresponden a las interfaces del servicio de búsqueda basado en métricas y fueron desarrolladas utilizando tecnologías del lado del cliente. Cada una de estas vistas corresponde a los diseños (Mockups) presentados en el Anexo C.

Figura 22. Interfaz web 1 servicio de búsqueda basado en métricas.

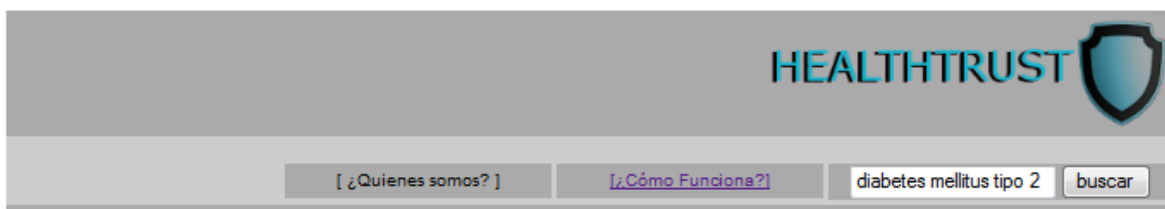


Figura 23. Interfaz web 2 servicio de búsqueda basado en métricas.

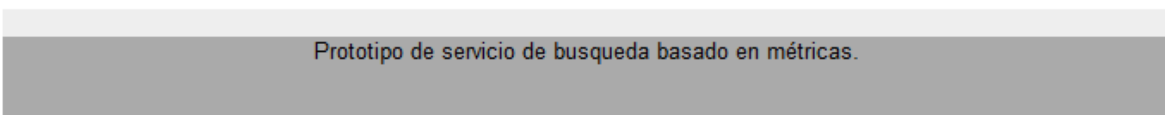


El servicio muestra los videos de las búsquedas por defecto de YouTube, adicionándoles las calificaciones correspondientes al algoritmo HealthTrust, basados en HITS y PageRank respectivamente. Mientras en la figura 24, se observa la interfaz correspondiente a la sección “¿Quiénes somos?”, donde se da una pequeña descripción del servicio y de las personas involucradas en su construcción.

**Figura 24.** Interfaz web 3 servicio de búsqueda basado en métricas.



HealthTrust es un algoritmo basado en confianza, el cual permite valorar la confianza en un autor y su contenido, teniendo en cuenta la opinión de la comunidad. Para este caso específico las opiniones son obtenidas dentro de la comunidad de Diabetes de YouTube.



Las interfaces mostradas en las Figuras 22-24, sirven para ilustrar como se podría implementar el servicio, pero por facilidad se realizara la evaluación del mismo en la RSVS Unisalud.

## **API REST**

La clase AbstractFacade cuenta con la siguiente funcionalidad genérica: crear, editar, remover, encontrar, encontrar por rango, y contar el número de registros. La implementación está escrita en código JPA Estándar.

**Figura 25.** Funcionalidad de la clase AbstractFacade.

```

public abstract class AbstractFacade<T> {
    private Class<T> entityClass;

    public AbstractFacade(Class<T> entityClass) {
        this.entityClass = entityClass;
    }

    protected abstract EntityManager getEntityManager();

    public void create(T entity) {
        getEntityManager().persist(entity);
    }

    public void edit(T entity) {
        getEntityManager().merge(entity);
    }

    public void remove(T entity) {
        getEntityManager().remove(getEntityManager().merge(entity));
    }

    public T find(Object id) {
        return getEntityManager().find(entityClass, id);
    }
}

```

En la figura 26 se puede observar la implementación de la clase VideoFacadeREST, que como se había señalado anteriormente reutiliza funcionalidad de la clase AbstractFacade (figura 25). A continuación en la tabla 15, se hace una relación de la funcionalidad que se presenta en esta figura por medio de las palabras listadas:

**Tabla 15.** Funcionalidad de la clase VideoFacadeREST.

Comando	Descripción
@stateless	Indica que la clase es un Bean de sesión sin estado.
@path("service.video")	La notación path es usada para identificar la URI (Uniform Resource Identifier), con la que nuestra clase va responder a la solicitud
@Path("{from}/{to}")	La notación Path en este caso indica que los métodos listados a continuación necesitan un parámetro para ser ejecutados ej: id
@Path("{id}")	
@Path("count")	
@POST, @PUT,@DELETE,@GET:	Esto indica que los métodos asociados a estas notaciones, responderán automáticamente, cuando se lleven a cabo las invocaciones http correspondientes.
@Produces({"application/xml", "application/json"})	Esta notación indica que la respuesta a la petición http se hará en un formato específico ej: XML



**Figura 26.** Funcionalidad de la clase VideoFacadeREST.

```
@Stateless
@Path("service.video")
public class VideoFacadeREST extends AbstractFacade<Video> {
    @PersistenceContext(unitName = "RestHealthTrustPU")
    private EntityManager em;

    public VideoFacadeREST() {
        super(Video.class);
    }

    @POST
    @Override
    @Consumes({"application/xml", "application/json"})
    public void create(Video entity) {
        super.create(entity);
    }

    @PUT
    @Override
    @Consumes({"application/xml", "application/json"})
    public void edit(Video entity) {
        super.edit(entity);
    }
}
```

La funcionalidad de la clase CanalFacadeREST es equivalente a la de la Clase VideoFacadeRest mostrada en la figura 27, ya que ambos son una especialización de la clase AbstractFacade, por lo cual no se describirá dentro de esta sección. De igual modo se pone a disposición su versión digital por medio de una carpeta de DropBox (ver anexo B).

**Figura 27.** Funcionalidad de la clase VideoFacadeRest.

```
@DELETE
@Path("/{id}")
public void remove(@PathParam("id") String id) {
    super.remove(super.find(id));
}

@GET
@Path("/{id}")
@Produces({"application/xml", "application/json"})
public Video find(@PathParam("id") String id) {
    return super.find(id);
}

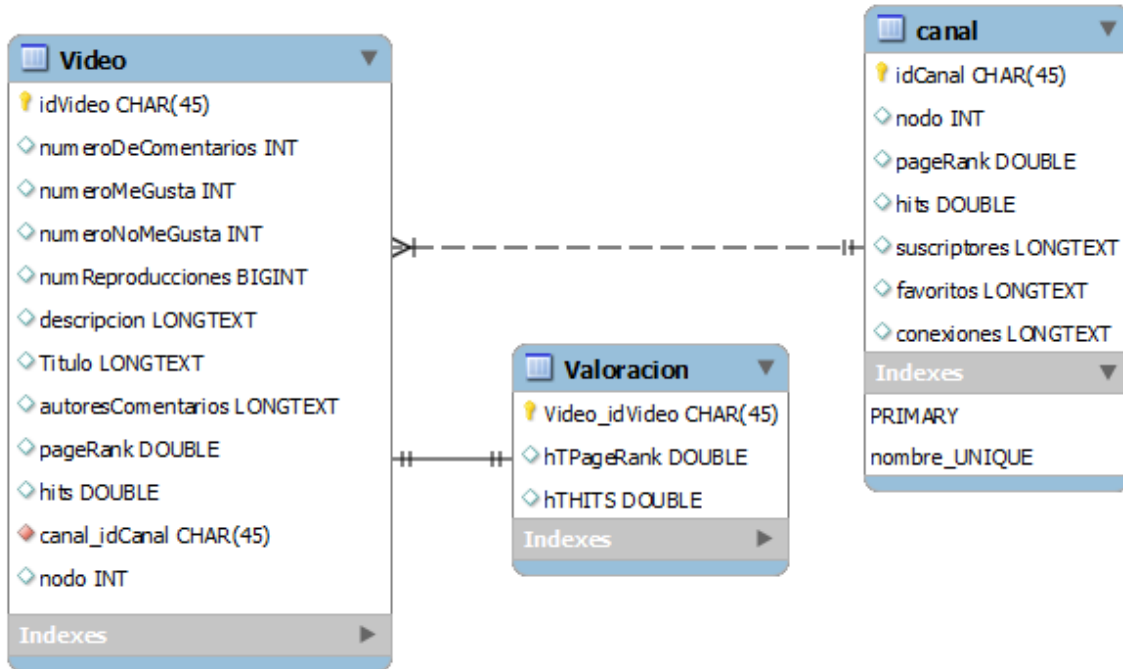
@GET
@Override
@Produces({"application/xml", "application/json"})
public List<Video> findAll() {
    return super.findAll();
}

@GET
@Path("/{from}/{to}")
@Produces({"application/xml", "application/json"})
public List<Video> findRange(@PathParam("from") Integer from, @PathParam("to") Integer to)
    return super.findRange(new int[]{from, to});
}
```

## MODELO DE LA BASE DE DATOS

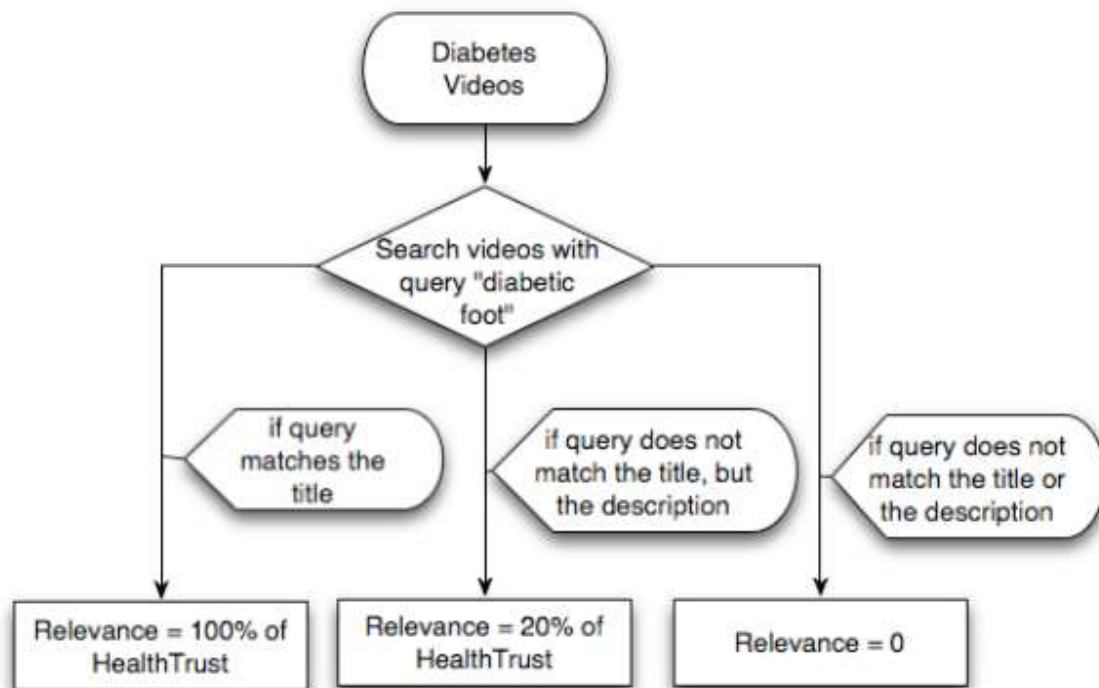
En la figura 28 se puede observar el modelo de la base de datos, donde los canales representan a los autores de contenido de la RSV YouTube, y los videos el contenido generado por esos autores. Cada uno de los atributos representan las relaciones sociales y las valoraciones relacionadas a cada uno de los nodos (autores y contenido). Por otro lado, la tabla valoración reúne las calificaciones de cada pieza de contenido, siguiendo el algoritmo HealthTrust planteado por Luque, y descrito en la sección anterior. También es importante destacar que las relaciones presentes en el modelo implican que un Canal puede contener muchos videos, pero cada video solo puede tener una valoración disponible por parte del algoritmo.

**Figura 28.** Modelo de la base de datos para la aplicación.



### Componente Sintáctico para asistir a las búsquedas

**Figura 29.** Componente sintáctico para asistir a las búsquedas[17].



El componente mostrado en la figura 29 tiene como funcionalidad comparar los resultados de la búsqueda con las palabras claves relacionadas a ella, y según su

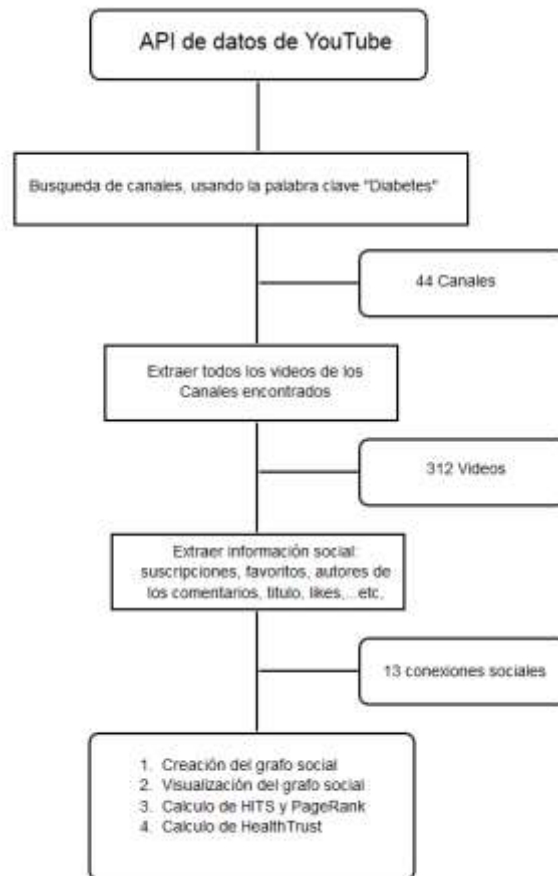
semejanza mantener su valor de calificación teniendo en cuenta su relevancia: a) si las palabras clave están contenidas en el título del video, la calificación de HealthTrust se mantiene. En otras palabras se le asigna el 100% del valor al video. B) si las palabras clave no están contenidas en el título pero si en la descripción del mismo, se les asigna el 20% de dicho valor c) si no coincide ni el título, ni la descripción con las palabras clave correspondientes a la búsqueda, se les asigna el 0%.

### 3.5.2 PRUEBAS

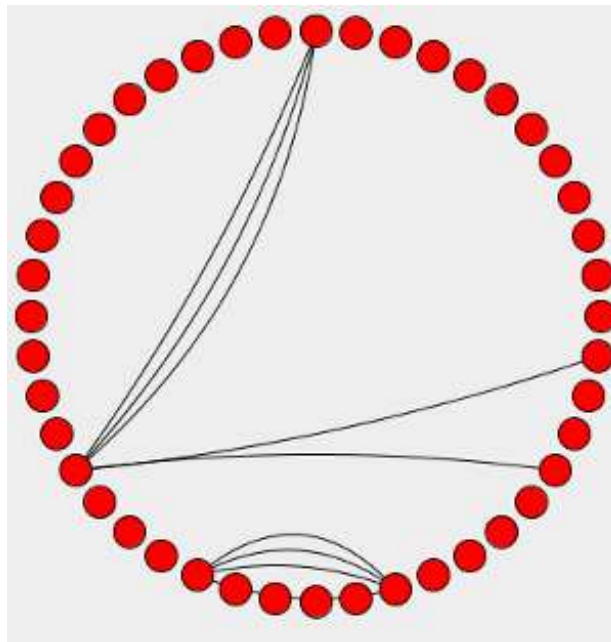
#### 3.5.2.1 Primer escenario de pruebas usando el prototipo de HealthTrust

En la figura 30 se observa el primer intento de extracción de datos, teniendo como la principal limitante que la API de YouTube no permite la búsqueda de Canales en un idioma específico. Además el resultado de la búsqueda está limitado a 500 canales o videos. Por lo anterior, al retornar los 500 elementos (canales o videos) permitidos por la misma (ver Anexo B), se encontró que solo 44/500 canales estaban en Español. Y estos canales tan solo contenían 312 videos, y tenían solo 13 conexiones sociales entre los nodos como se puede apreciar en la figura 31. En este mismo anexo se puede observar la tabla con los valores resultantes de HITS y PageRank.

**Figura 30.** Primera extracción de datos dentro del prototipo de HealthTrust.



**Figura 31.** Grafo social para el primer escenario de pruebas



### Tiempo empleado en el primer escenario de pruebas

Las pruebas realizadas se llevaron a cabo en un servidor con las siguientes características técnicas:

Intel(R) Core(TM) i5-2430M CPU 2.4 GHZ Memoria RAM 8,00 GHZ

Posteriormente se llevaron a cabo las siguientes búsquedas descritas en la Tabla 16, con la finalidad de estimar el tiempo promedio de la ejecución del primer escenario de pruebas.

**Tabla 16.** Tiempo de ejecución del escenario de pruebas 1.

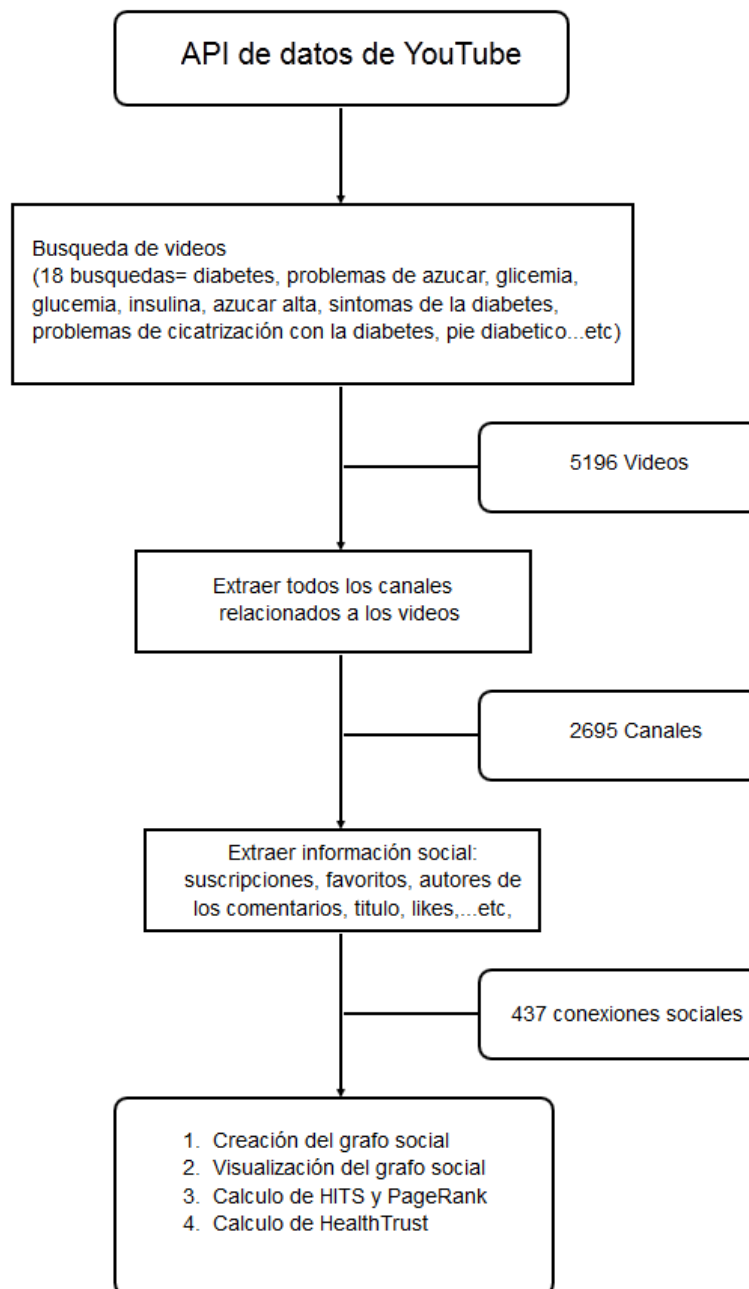
Pruebas	Tiempo
Búsqueda 1 con palabra clave diabetes	5:15 minutos
Búsqueda 2 con palabra clave diabetes	4:95 minutos
Búsqueda 3 con palabra clave diabetes	5:10 minutos

### 3.5.2.2 Segundo escenario de pruebas usando el prototipo de HealthTrust

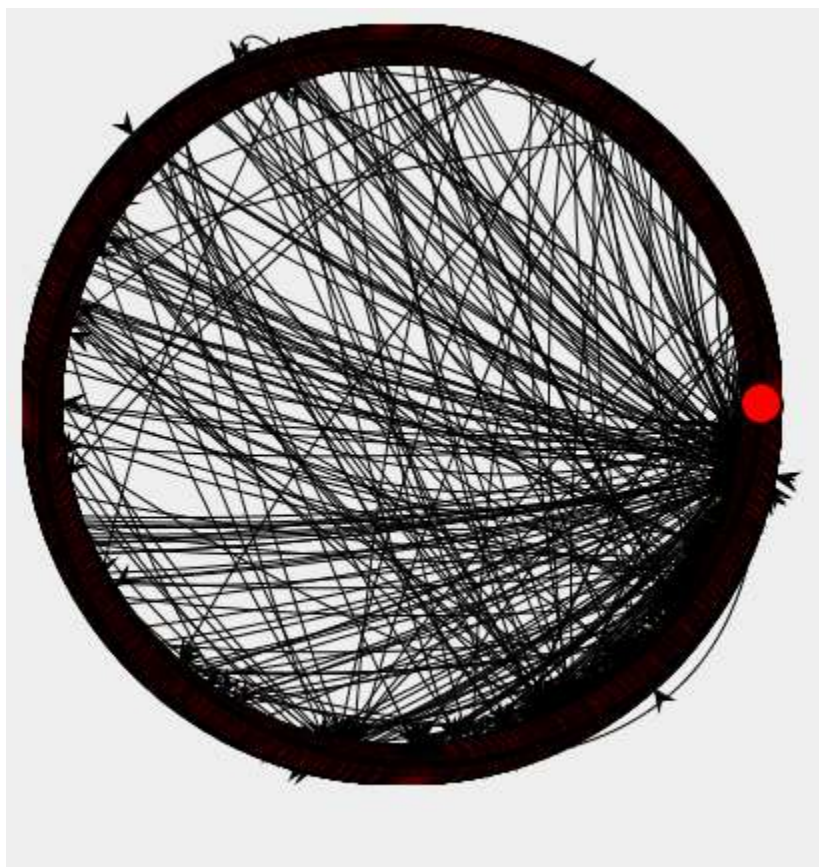
Para esta aproximación se buscó otro enfoque. Por medio de una entrevista a dos personas capacitadas en el área de Diabetes, se obtuvieron 18 búsquedas posibles sobre la diabetes. Los resultados fueron: *diabetes, problemas de azúcar, glicemia, glucemia, insulina, azúcar alta, síntomas de la diabetes, problemas de cicatrización con la diabetes, pie diabético, complicaciones de la diabetes, dietas para pacientes con diabetes, control para pacientes diabéticos, antecedentes para*

*pacientes diabéticos, examen de tolerancia a la glucosa, examen pre y post con carga, examen hbA1C, diabetes gestacional, diabetes mellitus tipo 2.* Como se puede observar en la figura 32, al realizar búsquedas sucesivas de videos, se pueden obtener más canales dentro de la comunidad de diabetes de YouTube, eso se relaciona a lo expresado en el anexo B, ya que la API de YouTube en su versión 2 tiene la capacidad de buscar Videos en Español más no canales. También es importante destacar que se descartaron los videos repetidos entre búsquedas.

**Figura 32:** Segunda extracción de datos dentro del prototipo de HealthTrust.



**Figura 33.** Grafo para la segunda aproximación.



En la figura 33 se puede observar el grafo correspondiente a la segunda aproximación de HealthTrust para el contenido. Debido a que el análisis se centrará directamente en los resultados numéricos mostrados por las métricas HITS y PageRank, solo se muestra el grafo a manera ilustrativa, ya que con su disposición actual es imposible realizar cualquier tipo de análisis.

### **Tiempo empleado en el segundo escenario de pruebas**

Las pruebas realizadas se llevaron a cabo en un servidor con las siguientes características técnicas:

Intel(R) Core(TM) i5-2430M CPU 2.4 GHZ Memoria RAM 8,00 GHZ

Posteriormente se llevaron a cabo las siguientes búsquedas descritas en la Tabla 17, con la finalidad de estimar el tiempo promedio de la ejecución del segundo escenario de pruebas, el cual se puede estimar como dos horas.

**Tabla 17.** Tiempo de ejecución del escenario de pruebas 2.

Pruebas	Tiempo
Búsqueda 1 con las 18 palabras claves mencionadas en el escenario 2	1 hora 55 minutos 15 segundos
Búsqueda 2 con las 18 palabras claves mencionadas en el escenario 2	1 hora 49 minutos 5 segundos
Búsqueda 3 con las 18 palabras claves mencionadas en el escenario 2	2 horas 10 minutos 12 segundos

### **3.6 Conclusiones**

En este capítulo se muestra la captura de requerimientos, análisis diseño, e implementación y pruebas del servicio. Un diseño más riguroso del servicio desde el punto de vista de la metodología UML, se referencia en el anexo B.

De las pruebas realizadas con HealthTrust se puede apreciar como es posible usar el servicio para generar el grafo social, el cálculo de las métricas HITS y PageRank, para finalmente calcular el valor de HealthTrust para el autor. En el próximo capítulo se presenta una evaluación más formal usando el método de Análisis de Efectos Cualitativos derivado de la metodología DESMET.



# Capítulo 4

## Evaluación del servicio

### 4.1 Introducción

El presente capítulo tiene como objetivo seleccionar y aplicar una metodología de evaluación, para demostrar la relevancia y precisión del servicio. Para seleccionar la mejor metodología de evaluación, de acuerdo a las condiciones particulares del servicio y del contexto en el cual se evaluará, se usará la tabla resumen proporcionada por el método DESMET[47], el cual contiene nueve metodologías diferentes de evaluación de métodos o herramientas de software. Después de haber seleccionado el método se realizará la evaluación formal de la relevancia y precisión del servicio.

### 4.2 DESMET

El método DESMET tiene como objetivo principal ayudar al evaluador miembro de una organización en particular, a planear y ejecutar una evaluación adecuada de un método o herramienta de software. Además DESMET cuenta con un mecanismo para evaluar la madurez de una organización, directrices para afrontar problemas de factores humanos, un manual de medición y una herramienta de recopilación de datos. Para propósitos de este trabajo de grado, la evaluación se limitará a verificar la relevancia de los resultados y precisión del servicio web.

### 4.3 Tipos de procesos de evaluación

**Evaluaciones cuantitativas:** destinadas al establecimiento de efectos medibles de la utilización de un método/herramienta software. Los efectos medibles se basan generalmente en la reducción del tiempo de producción, reconstrucción, mantenimiento o costos.

**Evaluaciones cualitativas:** destinadas al establecimiento de la idoneidad del método o herramienta software. Los evaluadores miden el grado en que el método/herramienta proporciona las características necesarias de manera útil y eficaz con base, generalmente, en la opinión personal.

**Evaluación Híbridas:** Finalmente, los métodos híbridos son aquellos que involucran tanto elementos objetivos como subjetivos.

### 4.4 Organización de la evaluación

DESMET ha identificado tres formas para organizar un ejercicio de evaluación:

**Experimento formal** donde a muchos sujetos (por ejemplo, ingenieros de software) se les pide realizar una o varias tareas utilizando los diferentes métodos que se están investigando. Los sujetos son asignados a cada método de forma

que los resultados son imparciales y pueden ser analizados empleando técnicas estadísticas estándar.

**Estudio de caso** donde cada método que se está investigando se prueba en un proyecto real utilizando procedimientos para el desarrollo de un proyecto estándar de la organización que está realizando la evaluación.

**Encuesta** donde el personal/organizaciones que han empleado métodos en proyectos pasados son interrogados para que proporcionen información sobre dicho método ó herramienta software. La información de los usuarios puede ser analizada utilizando técnicas estadísticas estándar.

#### 4.5 Selección del método de evaluación

DESMET describe una tabla que se lista a continuación, la cual permite evaluar el escenario del servicio de gestión de contenido desarrollado. Esta tabla que se puede encontrar en la referencia [47], ha sido modificada para saber si las condiciones se encuentran presentes dentro del servicio y en qué porcentaje.

**Tabla 18.** Evaluación de la selección del método DESMET[47].

Método de Evaluación	Condiciones Favorables del método	Condición Presente		Porcentaje (%)
		Sí	No	
Experimento Cuantitativo	Beneficios claramente cuantificables.		X	50%
	Disponibilidad del personal para participar en el experimento (por ejemplo, el desempeño de trabajo no productivo).	X		
	Método de adquisición relacionado con una sola tarea/actividad.	X		
	Beneficios directamente medibles de la salida de una tarea.		X	
	Tiempo de aprendizaje relativamente corto.	X		
	Deseo de realizar evaluaciones del método de adquisición independientes del contexto.		X	
Estudio de Caso Cuantitativo	Beneficios cuantificables en un solo proyecto.	X		60%
	Beneficios cuantificables antes del lanzamiento del producto.	X		
	Procedimientos de desarrollo estables.	X		

	Personal con experiencia en mediciones.		X	
	Plazos de evaluación proporcionales con el tiempo de desarrollo de los proyectos de tamaño normal.		X	
Encuestas Cuantitativas	Beneficios no cuantificables en un solo proyecto.	X		66.6%
	Existencia de una base de datos con logros del proyecto incluyendo: productividad, calidad, datos del método/herramienta.	X		
	Proyectos con experiencia en el uso del método.		X	
Análisis de Características por Proyección.	Amplio número de métodos a evaluar.		X	50%
	Cortos periodos de tiempo para realizar la evaluación.	X		
Análisis de Características por Estudio de Caso	Beneficios difíciles de cuantificar.		X	20%
	Beneficios observables en un solo proyecto.		X	
	Procedimientos de desarrollo estable.	X		
	Población de usuarios del método/herramienta limitado.		X	
	Plazos de evaluación proporcionales con el tiempo de desarrollo de los proyectos de tamaño normal.		X	
Análisis de Características por Experimento	Beneficios difíciles de cuantificar.		X	50%
	Beneficios directamente observables de la salida de una tarea.	X		
	Tiempo de aprendizaje relativamente corto.	X		
	Población de usuarios del método muy variados.		X	
Análisis de Características por Encuesta	Beneficios difíciles de cuantificar.		X	0%
	Población de usuarios del método muy variado.		X	
	Beneficios no observables en un solo proyecto.		X	
	Proyectos con experiencia en el uso del método, o proyectos preparados para aprender sobre el método.		X	

Análisis de Efectos Cualitativos - Opinión de expertos Intercalados	Disponibilidad de opiniones de expertos en evaluaciones de métodos similares.	X		75%
	Ausencia de procedimientos de desarrollo estables.	X		
	Requerimientos para mezclar y comparar métodos.		X	
	Interés en la evaluación de métodos genéricos.	X		
Benchmarking	Método enfocado en máquina, no en humanos.		X	50%
	Salidas del método capaces de ser clasificadas en términos de algún "buen" criterio.	X		

Como se puede observar el mejor porcentaje de adecuación al contexto del proyecto de grado lo obtuvo el método de Análisis de Efectos Cualitativos (75%), que a su vez es una metodología de evaluación mixta, la cual brinda más versatilidad a la evaluación.

#### **4.6 Análisis de Efectos Cualitativos-Opinión de expertos Intercalados**

Una forma común de realizar evaluaciones es utilizar la opinión de expertos. Para poder llevar a cabo la evaluación del servicio de gestión de calidad HealthTrust, se tendrán en cuenta dos expertos. Estas dos personas son asistentes investigadores de los proyectos “perfil epidemiológico de la diabetes mellitus dos, en el departamento del Cauca” y “Eficacia de la Hemoglobina Glicosilada como prueba diagnóstica de la diabetes mellitus tipo 2, en una población del departamento del Cauca” de la facultad de salud de la Universidad del Cauca, financiados por la Vicerrectoría de investigaciones y relacionados al área de diabetes. Estos expertos Además de tener experiencia en el dominio debido a que interactúan con pacientes con diabetes, también cuentan con educación y preparación en todos los temas relacionados a esta enfermedad. También es importante destacar que se buscó personas con perfiles similares a moderadores de una RSVS, que usualmente no son profesionales de la salud, pero generalmente son personas con formación o conocimientos de nivel medio o alto, con respecto a los temas que en ella se tratan.

#### **4.7 Escala de tiempo de la evaluación por parte de expertos**

El tiempo de evaluación por medio de la consulta a expertos es muy corto (unos pocos días), pero esto se encuentra supeditado a su disponibilidad de tiempo. Por otro lado, extraer los datos de un método/herramienta puede llevar bastante tiempo.

#### **4.8 Desarrollo de la evaluación**

Para el desarrollo de la evaluación se buscó analizar los Servicios de búsqueda YouTube y el servicio basado en el algoritmo de HealthTrust usando PageRank por medio de un Top de los 20 videos más relevantes. Para esto se utilizó la

búsqueda referenciada en la figura 32 con las palabras claves: *diabetes, problemas de azúcar, glicemia, glucemia, insulina, azúcar alta, síntomas de la diabetes, problemas de cicatrización con la diabetes, pie diabético, complicaciones de la diabetes, dietas para pacientes con diabetes, control para pacientes diabéticos, antecedentes para pacientes diabéticos, examen de tolerancia a la glucosa, examen pre y post con carga, examen hbA1C, diabetes gestacional, diabetes mellitus tipo 2*. Estas dos búsquedas dieron como resultado una lista conjunta de 40 videos, los cuales fueron organizados en orden alfabético, para posteriormente ser entregados a los evaluadores por medio de un grupo de discusión dentro de la RSVS Unisalud de la Universidad del Cauca como se muestra en la figura 34.

**Figura 34.** Evaluación por medio de un grupo de discusión en la RSVS Unisalud.



The image shows a screenshot of the Unisalud website interface. At the top, there is a header with the Universidad del Cauca logo and the text 'Unidad de Salud'. Below the header is a navigation menu with items: 'Actividad', 'Grupos Unisalud', 'Relaciones', 'Blogs', 'Multimedia', 'Ayuda', and 'más'. The main content area displays search results for the category 'diabetes mellitus'. Three results are visible:

- Carol tiene Diabetes**: Por Moderador Unisalud ayer. Categorías: diabetes mellitus. "Carol tiene diabetes". Cortometraje de animación, de 7 minutos de duración, creado por la Fundación para la Diabetes. La protagonista del corto, Carol, es una niña de nueve años que tiene diabetes tipo 1,...
- Cascada de la insulina. Bioquímica**: Por Moderador Unisalud ayer. Categorías: diabetes mellitus. <http://www.youtube.com/watch?v=Fax18EXt1Mo>
- CASCADA INSULINA**: Por Moderador Unisalud ayer. Categorías: diabetes mellitus.

#### 4.8.1 Evaluación de la Relevancia

El criterio de evaluación que llevarán a cabo los expertos consta de cuatro atributos (Confianza, Relevancia, Presentación y Completitud) con una calificación de uno (1) a cinco (5), teniendo en cuenta que uno (1) corresponde a una valoración "muy baja"; dos (2) corresponde a una valoración "baja"; tres (3) corresponde a una valoración "media"; cuatro (4) corresponde a una valoración

“alta”, y cinco (5) corresponde a una valoración “muy alta”. Esta evaluación de calidad fue definida en un proyecto anterior del grupo de ingeniería Telemática[10].

De los atributos evaluados se tendrá en cuenta principalmente el atributo de relevancia ya que se recomendó a los expertos que calificasen este atributo como “la relevancia del video para decidir si recomendaría o no el video a un paciente con diabetes”, pero para llegar a ese punto el revisor tiene que haber visto los títulos, las descripciones y la mayor parte del contenido de los videos. Para llevar a cabo esta evaluación se procedió a crear una base de datos en el programa de análisis estadístico SPSS en su versión 19 como se muestra en la figura 35. La versión completa de la base de datos se encuentra disponible en los anexos.

**Figura 35.** Base de datos referente a la evaluación en el programa SPSS

Ti	CCev1	Rev1	PeV1	CLev1	CCev2	Rev2	PeV2	CLev2	Id	posY	posPR	grupo	Rev1Consolid.	Rev2Consolid.
¿Cuál tiene diabetes?	Alta	Alta	Alta	Alta	Muy Alta	Alta	Alta	Alta	hOazdQaryl	16	1142	YouTube	Relevante	Relevante
Cocada de insulina Bioquímica	Alta	Media	Media	Media	Muy Alta	Media	Media	Media	Fax18EX1fMs	18	0	YouTube	Relevante	Relevante
Cocada de insulina	Alta	Media	Media	Media	Muy Alta	Media	Alta	Media	FVSD3jB2AQ	17	1904	YouTube	Relevante	Relevante
Como Aplicar a Insulina	Baja	Baja	Baja	Baja	Media	Baja	Media	Media	ZurprgmYY	14	0	YouTube	No Relevante	No Relevante
Cómo realizar una medición de glucemia capilar #1.	Alta	Alta	Alta	Alta	Muy Alta	Muy Alta	Muy Alta	Muy Alta	pkfaFFs0WY	12	2160	YouTube	Relevante	Relevante
Cuales son los sintomas de la Diabetes	Alta	Alta	Media	Baja	Media	Media	Media	Media	YG_Dty587s	9	284	YouTube	Relevante	Relevante
Diabetes - Tipos de insulina y sus usos - Diabetes	Alta	Alta	Media	Alta	Alta	Alta	Alta	Alta	slwQ1e9SH6	2	3518	YouTube	Relevante	Relevante
Diabetes - Valores bajos y elevados de azúcar en s.	Alta	Alta	Alta	Alta	Alta	Alta	Alta	Alta	gKEgJULBM4	13	102	YouTube	Relevante	Relevante
DIABETES CURACION TOTAL	Baja	Baja	Baja	Baja	Baja	Baja	Alta	Baja	BYM5GcDjp0	8	327	YouTube	No Relevante	No Relevante
Diabetes Cure - Treatment To Reverse Diabetes Th.	Baja	Baja	Baja	Baja	Baja	Baja	Baja	Baja	h0DaCwBtS6c	10	0	YouTube	No Relevante	No Relevante
DIABETES ¿PODEMOS CURARLA?	Alta	Alta	Alta	Alta	Alta	Alta	Alta	Alta	E0aK0eRTH4w	19	403	YouTube	Relevante	Relevante
Glucosa, insulina y diabetes	Media	Alta	Baja	Baja	Alta	Alta	Alta	Alta	fYutbwWLUhY	6	0	YouTube	Relevante	Relevante
La Insulina, la glucosa y ti	Baja	Baja	Baja	Baja	Baja	Baja	Baja	Baja	Mtw9jcnPbQ	4	2804	YouTube	No Relevante	No Relevante
primeros sintomas de la diabetes - sintomas del dia	Media	Alta	Media	Media	Muy Baja	Muy Baja	Muy Baja	Muy Baja	ucCWGRQGR8E	20	3718	YouTube	Relevante	No Relevante
Que es la diabetes	Alta	Alta	Alta	Alta	Muy Alta	Alta	Media	Alta	10s4X_Lgt98	1	208	YouTube	Relevante	Relevante
Que es la diabetes mellitus? - Diferencias entre la d	Media	Media	Media	Baja	Alta	Alta	Media	Media	sGm0bzD1Vgw	11	322	YouTube	Relevante	Relevante
QUE ES LA DIABETES TIPO 1 Y QUE ES LA DIA.	Muy Alta	Alta	Alta	Alta	Media	Media	Media	Media	ZTem5SDGz5k	6	4264	YouTube	Relevante	Relevante
regulación de la glucemia	Alta	Alta	Media	Media	Media	Media	Muy Baja	Baja	stFjatrUis_M	3	411	YouTube	Relevante	Relevante
REMEDIOS CASEROS PARA LA DIABETES	Media	Baja	Baja	Baja	Media	Baja	Alta	Media	bys8tklprs	7	76	YouTube	No Relevante	No Relevante
Simplemente Cuido Revertiendo la diabetes en 30 d	Media	Baja	Media	Baja	Baja	Baja	Baja	Baja	X124gnGPWo	16	114	YouTube	No Relevante	No Relevante
"Clauzura" Curso extraordinario de ped diabetico	Alta	Baja	Media	Baja	Media	Media	Media	Media	V2Kojic_sZE	0	15	HealthT	No Relevante	Relevante
clínica del pa diabético gita ácido úrico segúnite p	Media	Alta	Media	Media	Alta	Media	Media	Media	BwoYk0ss2M	0	10	HealthT	Relevante	Relevante
Como bajar cifras de Glicemia en 10 minutos	Baja	Baja	Baja	Baja	Baja	Baja	Baja	Baja	aBj_U7vYbY	0	4	HealthT	No Relevante	No Relevante

#### 4.8.2 Evaluación de la Precisión

Para cada uno de las búsquedas se llevara una evaluación de la precisión (K) de tres listas de 5, 10, 20 videos respectivamente, con las mejores calificaciones de HealthTrust basadas en PageRank, y relevancia (YouTube). Donde K simboliza el número de videos para cada lista.

$$Precisión(K) = \text{videos reelevantes} / (K) \text{Videos obtenidos}$$

#### 4.9 Resultados

A continuación se presentan los resultados obtenidos de las evaluaciones descritas en el numeral anterior.

##### 4.9.1 Evaluación de Relevancia

En la Tabla 19 se presenta un análisis descriptivo de los resultados, comparando la evaluación de relevancia obtenida por parte de los expertos tanto del grupo de

los veinte (20) primeros videos obtenidos de la búsqueda en YouTube, como del grupo de los 20 primeros videos obtenidos con aplicando el algoritmo HealthTrust.

**Tabla 19.** Análisis descriptivo de los resultados

<b>Estadísticos</b>			Relevancia evaluador 1	Relevancia evaluador 2
YouTube	N	Válidos	20	20
		Perdidos	0	0
		Media	3,25	3,05
		Desviación. Típica (estándar).	,910	1,050
		Mínimo	2	1
		Máximo	4	5
HealthTrust	N	Válidos	20	20
		Perdidos	0	0
		Media	3,35	3,30
		Desviación. Típica (estándar).	1,089	1,031
		Error típ. de curtosis	,992	,992
		Mínimo	1	1
	Máximo	5	5	

De esta tabla se puede concluir que la media de la relevancia obtenida por el primer evaluador muy similar para los dos grupos (media HealthTrust =3,35 y la media de YouTube =3,25), mientras que para el evaluador 2, la media de la relevancia de los videos resultantes de la búsqueda con HealthTrust es más alta (Media = 3,3) que la media de la relevancia de los videos resultantes de la búsqueda con YouTube (Media =3,05). Del análisis de la desviación estándar, puede observarse que para el evaluador 1, la dispersión de los datos es mayor para el conjunto de videos obtenidos con HealthTrust, mientras que para el evaluador 2, la dispersión de los datos es mayor para el conjunto de videos obtenidos con YouTube. Sin embargo las diferencias entre las dispersiones de los datos entre grupos son mínimas. Del análisis de máximos y mínimos se observa que en ambos grupos (videos YouTube vs HealthTrust) el rango de valores es de 1 a 5, es decir que se obtuvieron valores de relevancia desde muy bajos a muy altos. La única excepción fue en la valoración de relevancia del evaluador 1 de los videos en YouTube, donde el rango fue de 4(Mínimo = 2, Máximo =4).

De las Tablas 20 hasta 23, se presenta un análisis de distribución de frecuencias, tanto para el evaluador 1, como el evaluador 2.

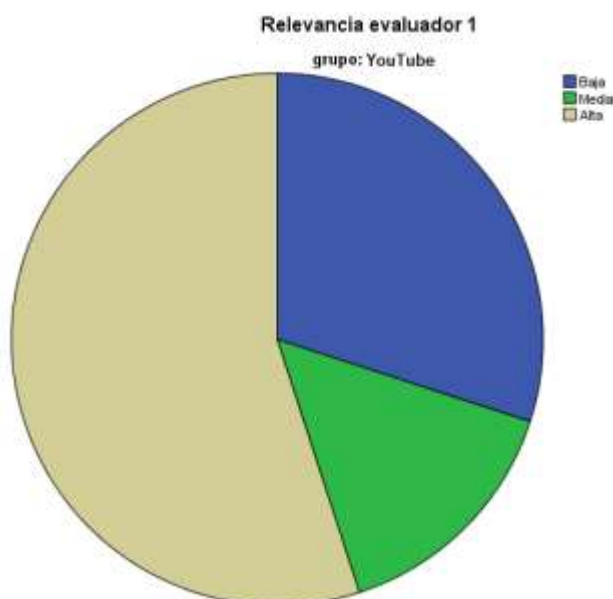


**Tabla 20** Distribución de frecuencias para la variable relevancia del Evaluador 1  
**Relevancia evaluador 1**

Grupo		Frecuencia	Porcentaje	Porcentaje válido	Porcentaje acumulado
YouTube	Baja	6	30,0	30,0	30,0
	Media	3	15,0	15,0	45,0
	Alta	11	55,0	55,0	100,0
	Total	20	100,0	100,0	
HealthTrust	Muy Baja	2	10,0	10,0	10,0
	Baja	2	10,0	10,0	20,0
	Media	4	20,0	20,0	40,0
	Alta	11	55,0	55,0	95,0
	Muy Alta	1	5,0	5,0	100,0
	Total	20	100,0	100,0	

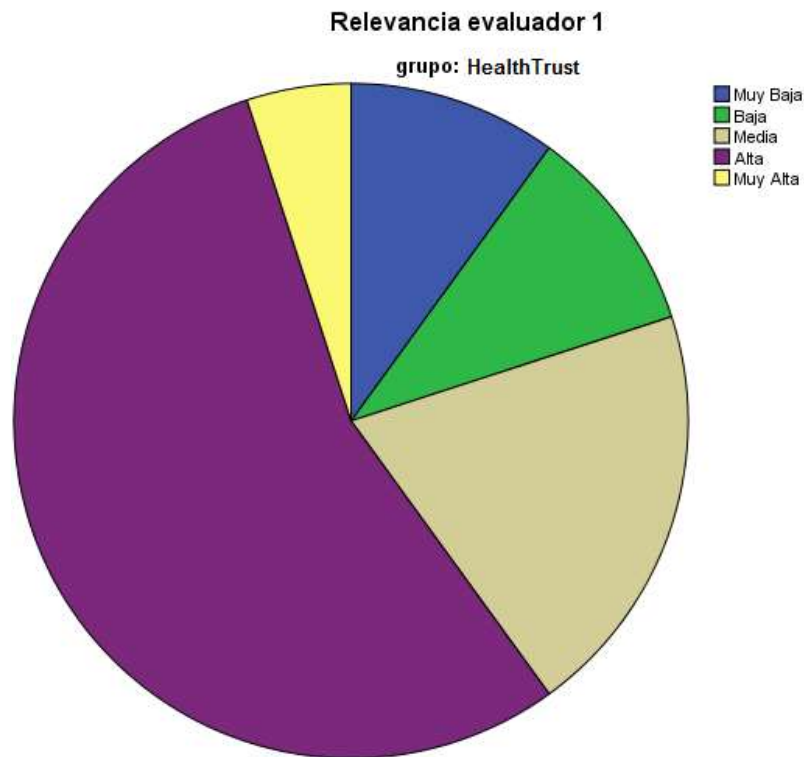
De la Tabla 20. Se puede observar que para el evaluador 1 no hubo diferencia en la calificación de relevancia entre el grupo de videos de YouTube y HealthTrust. El valor más frecuente en ambos grupos fue “Alto” (Relevancia =4) que corresponde al 55% de los videos evaluados (11 videos en cada grupo). A continuación se muestran las figuras 36, 37 donde se observan más claramente la distribución de relevancia para los servicios de busque de YouTube y HealthTrust respectivamente.

**Figura 36.** Calificación de Relevancia para YouTube por parte del evaluador 1





**Figura 37.** Calificación de Relevancia para HealthTrust por parte del evaluador 1

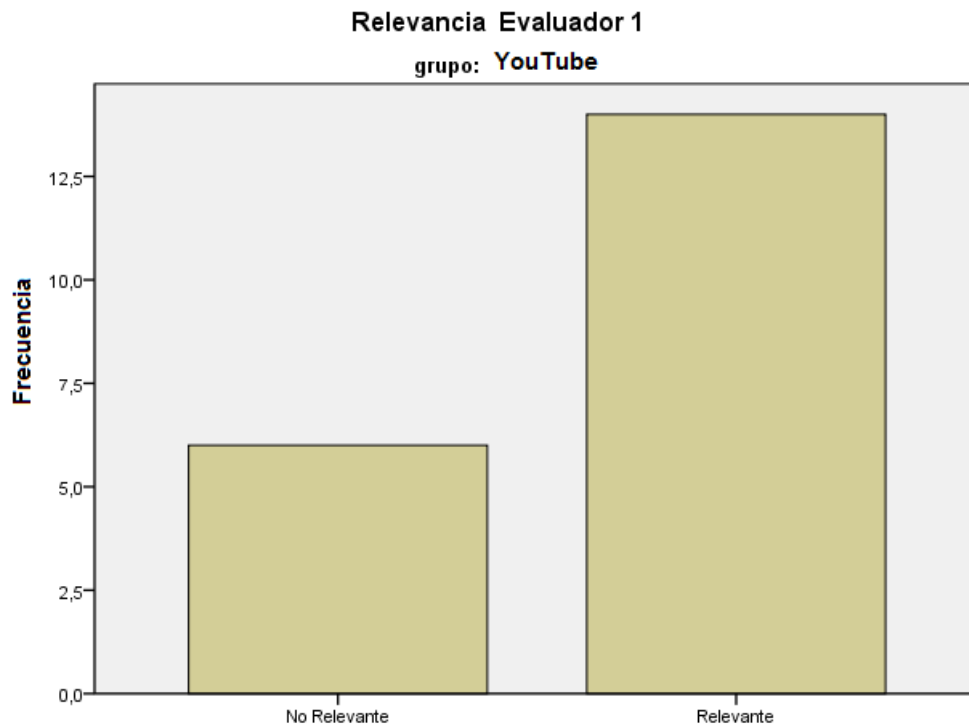


**Tabla 21** Distribución de frecuencias para la variable relevancia del Evaluador 1, con el valor de relevancia transformado de variable de escala, a variable en dos estados (0=No relevante, 1=Relevante).

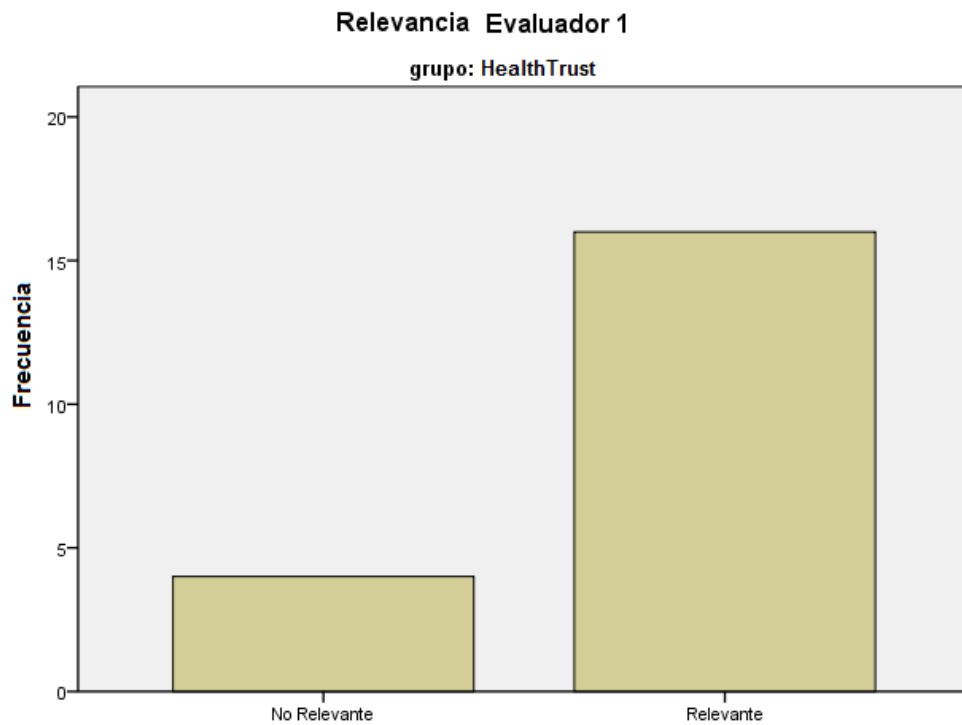
Grupo		Frecuencia	Porcentaje (%)	Porcentaje acumulado
YouTube	No Relevante	6	30,0	30,0
	Relevante	14	70,0	100,0
	Total	20	100,0	
HealthTrust	No Relevante	4	20,0	20,0
	Relevante	16	80,0	100,0
	Total	20	100,0	

De la Tabla 21 puede observarse que para el grupo de videos obtenidos con la búsqueda mediante HealthTrust, el porcentaje de videos relevantes fue del 80%, mientras que para el grupo de videos obtenidos con la búsqueda mediante YouTube fue del 70%. Para una visión más clara de estos valores se presentan las figuras 38 y 39.

**Figura 38.** Relevancia absoluta del Evaluador 1, usando el servicio de búsqueda de YouTube.



**Figura 39.** Relevancia absoluta del Evaluador 1, usando el servicio de búsqueda de HealthTrust.



De la tabla 22 podemos observar que la frecuencia predominante para el evaluador 2, es “Alta” para ambos servicios (HealthTrust vs YouTube) con un 35%, lo que es congruente con lo mostrado en la frecuencia del evaluador 1 (tabla 20). Además, es importante destacar que el servicio de YouTube, tuvo muchos más resultados “bajos”, mientras que el servicio de HealthTrust presento un resultado más con la valoración de “muy alto”, comparado con el grupo de videos de YouTube. En las figuras 40 y 41, se muestra una visión más clara por medio de graficas circulares.

**Tabla 22.** Distribución de frecuencias para la variable relevancia del Evaluador 2.

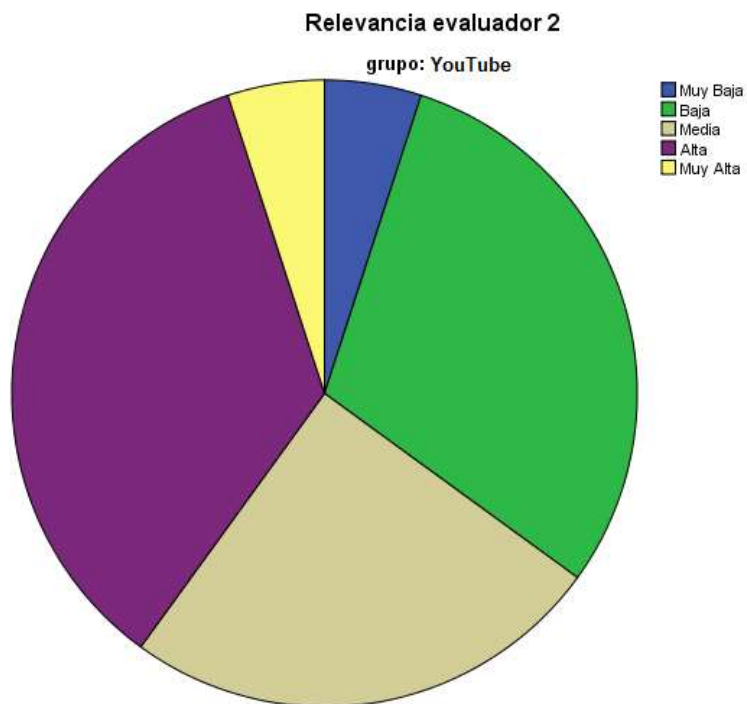
Grupo		Frecuencia	Porcentaje (%)	Porcentaje acumulado
YouTube	Muy Baja	1	5,0	5,0
	Baja	6	30,0	35,0
	Media	5	25,0	60,0
	Alta	7	35,0	95,0
	Muy Alta	1	5,0	100,0
	Total	20	100,0	
	HealthTrust	Muy Baja	1	5,0
Baja		3	15,0	20,0
Media		7	35,0	55,0
Alta		7	35,0	90,0
Muy Alta		2	10,0	100,0
Total		20	100,0	

En la tabla 23 se puede observar que el porcentaje de videos relevantes (calculado la relevancia absoluta) para el evaluador 2 con respecto al servicio de búsqueda de YouTube es de 65%, mientras que para el servicio basado en el algoritmo de HealthTrust es de 80%. A continuación se muestran las gráficas 40 y 41, donde se pueden observar más claramente estos valores.

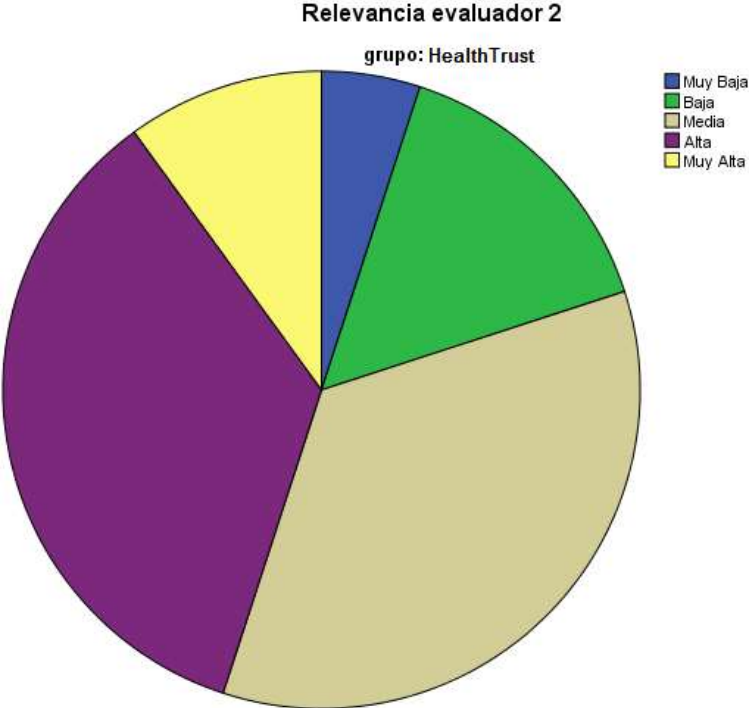
**Tabla 23.** Distribución de frecuencias para la variable relevancia del Evaluador 2, con el valor de relevancia transformado de variable de escala, a variable en dos estados (0=No relevante, 1=Relevante).

Grupo		Frecuencia	Porcentaje (%)	Porcentaje acumulado
YouTube	No Relevante	7	35,0	35,0
	Relevante	13	65,0	100,0
	Total	20	100,0	
HealthTrust	No Relevante	4	20,0	20,0
	Relevante	16	80,0	100,0
	Total	20	100,0	

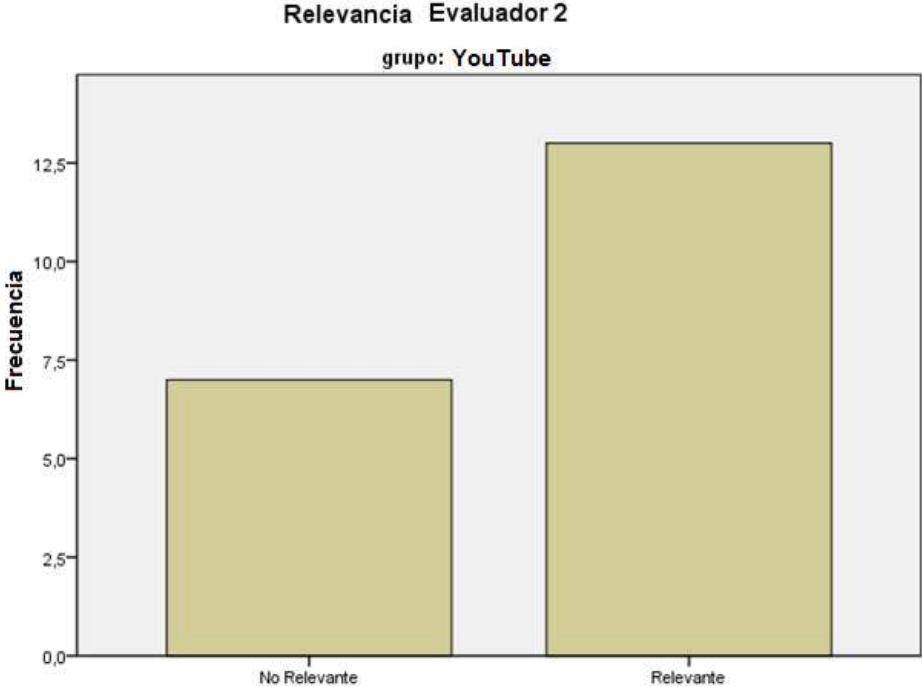
**Figura 40.** Calificación de Relevancia para YouTube por parte del evaluador 2.



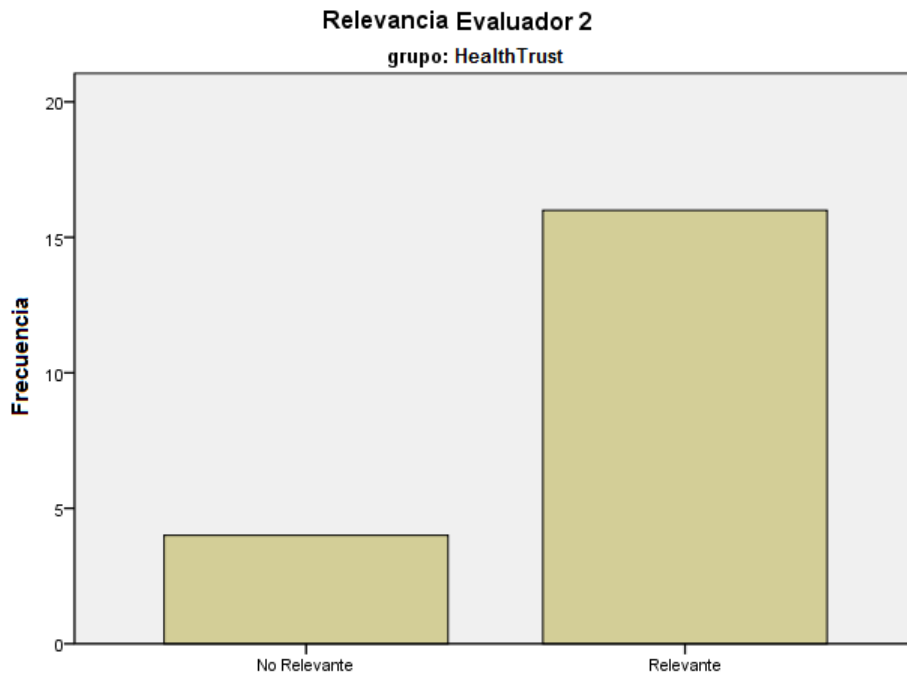
**Figura 41.** Calificación de Relevancia para HealthTrust por parte del evaluador 2.



**Figura 42.** Relevancia absoluta del Evaluador 2, usando el servicio de búsqueda de YouTube.



**Figura 43.** Relevancia absoluta del Evaluador 2, usando el servicio de búsqueda de HealthTrust.



#### 4.9.2 Evaluación de la Precisión

Estudiamos la precisión de tres listas de K canales, usando los valores (k=5, K=10 y K=20) con los valores de relevancia más altos para el algoritmo HealthTrust y para las búsquedas por defecto de YouTube. Para lo cual se plantean dos posibles escenarios, en los cuales un canal se puede considerar como relevante: 1) Ambos revisores recomendaron el canal y 2) Al menos un revisor recomendó el canal.

**Tabla 24.** Evaluación de precisión de los servicios para al menos un revisor.

Recomendado por / Precisión	Al menos un revisor			
	YouTube		HealthTrust (PageRank)	
	proporción	porcentaje	proporción	porcentaje
Precisión a 5	4/5	80%	4/5	80%
Precisión a 10	7/10	70%	9/10	90%
Precisión a 20	16/20	80%	18/20	90%

**Tabla 25.** Evaluación de precisión de los servicios para ambos revisores.

Recomendado por / Precisión	Ambos revisores			
	YouTube		HealthTrust (PageRank)	
	proporción	porcentaje	proporción	Porcentaje
Precisión a 5	4/5	80%	4/5	80%
Precisión a 10	5/10	50%	8/10	80%
Precisión a 20	14/20	70%	14/20	70%

Se observa que la precisión usando el servicio basado en HealthTrust es igual para pocos videos (K=5), pero mejora para más videos (K=10 y K=20), comparado con la precisión de las búsquedas en YouTube. En el caso de tener al menos un revisor, la precisión del servicio basado en HealthTrust aumenta de 80% a 90%, cuando se evalúan más videos (precisión a K=10 y K=20). Sin embargo, en el caso de los dos revisores, la precisión del servicio basado en HealthTrust disminuye a mayor número de videos (K=20). Este mismo fenómeno se observa en la evaluación obtenida con los videos de YouTube, con lo que se presume que a mayor número de canales evaluados, la percepción de relevancia de los canales por parte de los evaluadores es menor.

# Capítulo 5

## Conclusiones y trabajo futuro

En este trabajo de grado se aborda la necesidad de gestionar la calidad de los contenidos que se publican en el entorno de las Redes Sociales Virtuales en Salud. Inicialmente la investigación se centró en la búsqueda de atributos y métricas de calidad por medio de la investigación documental planteada por Serrano[16]. Este estudio permitió generar una base de conocimiento que fue posteriormente plasmada por medio de un servicio web que permite consultar las valoraciones de cada uno de los videos o canales con respecto al algoritmo HealthTrust y sus dos métricas HITS y PageRank.

### 4.1 Conclusiones del Estado arte.

- Las Redes Sociales Virtuales en Salud (RSVS) se ha convertido en una nueva forma de promoción de la salud donde los pacientes pueden retroalimentar sus experiencias, compartir sus opiniones sobre la efectividad de medicamentos, tratamientos, y brindar apoyo a personas que tienen su misma condición como pacientes.
- El modelo de descripción jerárquico de los parámetros de calidad propuesto en este trabajo de grado constituye un marco de referencia relevante para analizar las diferentes iniciativas de calidad presentes en la actualidad.
- Las métricas de confianza son ampliamente usadas en la Web 2.0, y constituyen un punto de partida importante para la gestión de la calidad de la información.
- Existen varios atributos que pueden ser utilizados para gestionar la calidad de la información en sitios de Redes Sociales Virtuales en Salud (RSVS).
- Del análisis y selección realizada se puede concluir que los criterios más adecuados para ser aplicados en un RSVS son los de Autoría, Confianza, Precisión, Completitud. La única métrica que existe y ha sido implementada mediante herramientas software es la métrica HealthTrust, que sirve para los atributos de autoría y confianza.

### 4.2 Diseño e implementación de servicio

- La métrica PageRank presento un mejor desempeño que la métrica HITS en los dos escenarios de pruebas planteados.
- Con base a los dos escenarios de pruebas planteados se puede concluir que el tiempo requerido para el análisis de las relaciones sociales es muy alto y requiere máquinas muy potentes. El tiempo para la generación del grafo es directamente proporcional al número de canales y videos en la base de datos, por lo tanto este análisis se dificulta para grafos de redes sociales de gran tamaño.



- La API de YouTube de datos en su Versión 2 es más eficaz retornando resultados en español que el buscador presente en la página principal de YouTube, cuando se realizan búsquedas de videos.
- Las métricas HITS y PageRank constituyen una herramienta importante para evaluar la relevancia de los nodos dentro de un grafo social.

### **4.3 Conclusiones de la Evaluación**

- En general la evaluación de la relevancia de los videos resultantes de la búsqueda con HealthTrust fue más alta que la evaluación de relevancia de los videos resultantes de la búsqueda con YouTube.
- El valor más frecuente de relevancia fue “Alto” en ambos grupos. Sin embargo, para los dos evaluadores, el promedio de relevancia de HealthTrust fue de 80%, mientras que para el grupo de videos obtenidos con la búsqueda mediante YouTube fue del 70%.
- El servicio basado en el algoritmo de HealthTrust mediante la utilización de la métrica PageRank, es más preciso que el servicio de YouTube en los dos escenarios planteados.

### **4.4 Trabajo futuro**

- Implementar un componente semántico a las búsquedas que complemente la labor del componente sintáctico, por medio de la inclusión de ontologías de dominio en salud.
- Enriquecer los rankings del algoritmo de HealthTrust con el uso de otras interacciones sociales como, “numero de reproducciones”, “likes”, “dislikes”.
- Implementar otras métricas basadas en los atributos seleccionados en el objetivo 1, que permitan en conjunto con HealthTrust gestionar la calidad del contenido dentro de una RSVS.
- Probar la efectividad de la métrica HealthTrust con otros tipos de contenido como infografías, publicaciones en wikis, y otro tipo de contenido publicado por medio de aplicaciones de social media.
- Modelar una Red de contenido con información existente, dentro de una RSVS como estudiabetes.org, teniendo en cuenta que la información existente en este sitio ya se encuentra moderada y usar ese modelo de referencia para evaluar contenido similar usando el algoritmo HealthTrust.

## REFERENCIAS

- [1] Gunther Eysenbach, John Powell, Oliver Kuss, y Eun-Ryoung Sa, «Empirical Studies Assessing the Quality of Health Information for Consumers on the World Wide Web (A Systematic Review)». *Journal of the American Medical Association*, may-2002.
- [2] A. Risk y J. Dzenowagis, «Review Of Internet Health Information Quality Initiatives», *J Med Internet Res*, vol. 3, n.º 4, dic. 2001.
- [3] «Our commitment to reliable health and medical information». [Online]. Available: <http://www.hon.ch/HONcode/Patients/Visitor/visitor.html>. [Accessed: 04-jun-2012].
- [4] Pudliszek, «Assessment of the quality of medical wikis», 16-jun-2009.
- [5] «Advogato - Trust Metric». [Online]. Available: <http://www.advogato.org/trust-metric.html>. [Accessed: 14-jun-2012].
- [6] «FilmTrust». [Online]. Available: <http://trust.mindswap.org/FilmTrust/>. [Accessed: 15-jun-2012].
- [7] «Trustmail :: Solutions». [Online]. Available: <http://trustmail.in/spanish/solutions/index.html>. [Accessed: 15-jun-2012].
- [8] presidencia de la Republica, «Decreto 1011 de 2006», 03-abr-2006. [Online]. Available: <http://www.alcaldiabogota.gov.co/sisjur/normas/Norma1.jsp?i=19975>. [Accessed: 13-nov-2012].
- [9] Standards Coordinating Committee of the Computer Society of the IEEE, «IEEE Standard Glossary of Software Engineering Terminology». 28-sep-1990.
- [10] Diego M. López, Bernd Blobel, y Carolina González, «Architectural Approach for Quality and Safety Aware HealthCare Social Networks». .
- [11] E. Torrente, J. Escarrabill, y T. Martí, «Impacto de las redes sociales de pacientes en la práctica asistencial», *Revista de Innovación Sanitaria y Atención Integrada*, vol. 2, n.º 1, p. 1, 2010.
- [12] T. O'Reilly, «Web 2.0: compact definition», *Message posted to http://radar.oreilly.com/archives/2005/10/web\_20\_compact\_definition.html*, 2005.
- [13] Wikipedia, «Servicio web - Wikipedia, la enciclopedia libre». [Online]. Available: [http://es.wikipedia.org/wiki/Servicio\\_web](http://es.wikipedia.org/wiki/Servicio_web). [Accessed: 19-may-2013].
- [14] R. Lempel y S. Moran, «The stochastic approach for link-structure analysis (SALSA) and the TKC effect», *Computer Networks*, vol. 33, n.º 1, pp. 387–401, 2000.
- [15] «JMIR--Infodemiology and Infeveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet | Eysenbach | Journal of Medical Internet Research». [Online]. Available: <http://www.jmir.org/2009/1/e11/>. [Accessed: 27-nov-2012].
- [16] Carlos Enrique Serrano, «Modelo Integral para el Profesional en Ingeniería». Universidad del Cauca, 2008.
- [17] Luis Fernandez-Luque1,, Randi Karlsen, y Genevieve B Melton, «HealthTrust: a social network approach for retrieving online health videos.» .
- [18] A. Y. S. Lau, K. A. Siek, L. Fernandez-Luque, H. Tange, P. Chhanabhai, S. Y. W. Li, P. L. Elkin, A. Arjabi, L. Walczowski, C. S. Ang, y others, «The Role of Social Media for Patients and Consumer Health», *IMIA Yearbook*, vol. 6, n.º 1, pp. 131–138, 2011.
- [19] J. M. Moreno, J. M. Morales del Castillo, C. Porcel, y E. Herrera-Viedma, «A quality evaluation methodology for health-related websites based on a 2-tuple fuzzy linguistic approach», *Soft Computing*, vol. 14, n.º 8, pp. 887-897, jul. 2009.
- [20] Health Summit Working Group, «QUALITY OF HEALTH INFORMATION ON THE INTERNET». [Online]. Available: <http://www.ieee.org/organizations/pubs/newsletters/npss/march2000/health.htm>. [Accessed: 19-may-2013].

- [21] «eEurope 2002: Quality Criteria for Health related Websites», *Journal of Medical Internet Research*, vol. 4, n.º 3, nov. 2002.
- [22] Health On the Net Fundation, «Certification for collaborative Websites / Websites with Web 2.0 elements». [Online]. Available: [http://www.hon.ch/cgi-bin/HONcode/guidelines\\_comments\\_en.pl](http://www.hon.ch/cgi-bin/HONcode/guidelines_comments_en.pl). [Accessed: 20-may-2013].
- [23] F. Viegner, «El derecho a la Intimidación y los límites a la injerencia estatal».
- [24] Congreso de Colombia, «Ley Estatutaria No. 1581 Disposiciones generales para la Protección de datos personales». .
- [25] «Guide to Healthy Web Surfing: MedlinePlus». [Online]. Available: <http://www.nlm.nih.gov/medlineplus/healthywebsurfing.html>. [Accessed: 22-may-2012].
- [26] «Quality Guidelines: healthfinder.gov - Your Source for Reliable Health Information». [Online]. Available: [http://healthfinder.gov/aboutus/content\\_guidelines.aspx](http://healthfinder.gov/aboutus/content_guidelines.aspx). [Accessed: 04-jun-2012].
- [27] Department of health and human services, «Social Media Guidelines and Best Practices». Centers for disease control and preventions.
- [28] T. G. Muñoz, «EL CUESTIONARIO COMO INSTRUMENTO DE INVESTIGACIÓN/EVALUACIÓN.», *España,[Documento en línea]. Disponible: http://personal.telefonica.terra.es/web/medellinbadajoz/sociologia/El\_Cuestionario.pdf*, 2003.
- [29] B. Stvilia, M. B. Twidale, L. C. Smith, y L. Gasser, «Assessing information quality of a community-based encyclopedia», in *Proceedings of the International Conference on Information Quality*, 2005, vol. 11.
- [30] F. Ojeda, F. J. Perales, y J. Gutierrez-Perez, «Quality evaluation of websites and blogs for environmental education», *Cultura y Educaci&#243;n*, vol. 24, n.º 1, pp. 77–93, 2012.
- [31] Donald W. Kemper, «Hi-Ethics:Tough Principles for Earning Consumer Trust». .
- [32] E. R. Weitzman, E. Cole, L. Kaci, y K. D. Mandl, «Social but safe? Quality and safety of diabetes-related online social networks», *Journal of the American Medical Informatics Association*, vol. 18, n.º 3, pp. 292-297, ene. 2011.
- [33] Osma Suominena, Eero Hyvöna, Kim Viljanena, y Eija Hukkab, «HealthFinland—A national semantic publishing network». .
- [34] E. Hyvönen, K. Viljanen, y O. Suominen, «HealthFinland—Finnish health information on the semantic web», *The Semantic Web*, pp. 778–791, 2007.
- [35] «DISCERN - The DISCERN Instrument». [Online]. Available: [http://www.discern.org.uk/discern\\_instrument.php#feedback](http://www.discern.org.uk/discern_instrument.php#feedback). [Accessed: 04-jun-2012].
- [36] «Apache Lucene - Apache Solr». [Online]. Available: <http://lucene.apache.org/solr/>. [Accessed: 20-jun-2012].
- [37] M. R. Prajapati, «A Survey Paper on Hyperlink-Induced Topic Search (HITS) Algorithms for Web Mining», *International Journal of Engineering*, vol. 1, n.º 2, 2012.
- [38] Osmar R. Zaiane, «Hyperlink Induced Topic Search (HITS)», *Web information Systems*. [Online]. Available: <http://webdocs.cs.ualberta.ca/~zaiane/courses/cmput499/slides/Lect14/sld053.htm>. [Accessed: 25-nov-2013].
- [39] «JUNG - Java Universal Network/Graph Framework». [Online]. Available: <http://jung.sourceforge.net/>. [Accessed: 16-dic-2013].
- [40] «HITS (jung2 2.0 API)». [Online]. Available: <http://jung.sourceforge.net/doc/api/edu/uci/ics/jung/algorithms/scoring/HITS.html>. [Accessed: 16-dic-2013].

- [41] A. Borodin, G. O. Roberts, J. S. Rosenthal, y P. Tsaparas, «Finding authorities and hubs from link structures on the world wide web», in *Proceedings of the 10th international conference on World Wide Web*, 2001, pp. 415–429.
- [42] L. Page, S. Brin, R. Motwani, y T. Winograd, «The PageRank citation ranking: bringing order to the web.», 1999.
- [43] S. White y P. Smyth, «Algorithms for estimating relative importance in networks», in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 266–275.
- [44] Resnick, Marc, «Social Networking in Health Care Part II: Managing Trust in Peer to Peer Support Groups on Social Networks | World Usability Day». [Online]. Available: <http://www.worldusabilityday.org/social-networking-health-care-part-ii-managing-trust-peer-peer-support-groups-social-networks>. [Accessed: 29-jul-2013].
- [45] M. F. S. Sarasty, «AMIR-ST: Propuesta de una Aproximación Metodológica para la Ingeniería de Requisitos de Sistemas Telemáticos», *Jornadas de Ingeniería de Requisitos Aplicada, Sevilla*, 2001.
- [46] D. E. Bouchard, G. S. Knight, y A. Y. Chikhani, «Representing power distribution systems using objects and patterns», in *Electrical and Computer Engineering, 1998. IEEE Canadian Conference on*, 1998, vol. 2, pp. 208-214.
- [47] Barbara Kitchenham, «DESMET: A method for evaluating Software Engineering methods and tools». Department of Computer Science, ago-1996.

