# Plataforma de Colaboración para la Investigación

Claudia Pérez Martínez
M. en Ciencias de la Computación
CU UAEM Valle de Chalco
claudia.uaem@hotmail.com
Area de conocimiento: Tecnología educativa
Disciplina: Lingüística computacional

Gabriel López Morteo
Dr. en Ciencias
UABC Instituto de Ingeniería
galopez@uabc.edu.mx
Area de conocimiento: Tecnología Educativa
Disciplina: Desarrollo de sistemas y telecomunicaciones

Juvenal Rueda Paz Dr. en Física CU UAEM Valle de Chalco juvenal.rueda@gmail.com Area de conocimiento: Física Disciplina: Física

Magally Martínez Reyes
Dra. en Matemática Educativa
CU UAEM Valle de Chalco
mmreyes@hotmail.com
Area de conocimiento: Educación
Disciplina: Enseñanza de las matemáticas

# Tabla de contenido

- I Introducción
- II Justificación y objetivos
- III Marco de referencia teórico-metodológico
- IV Análisis de Resultados
- V Conclusiones
- VI Bibliografía
- VII Anexos

#### I. Introducción

#### 1.1 Antecedentes

#### 1.1.1 Formas de representación del conocimiento

Según la UNESCO (2005a) la diversidad de las modalidades de acceso al conocimiento constituye una de las características más importantes de la sociedad del aprendizaje y supone también, entre otras cosas, el cuestionamiento de los conceptos de inteligencia en los que prevalecía el carácter estable de los procesos de evaluación y trasmisión de los conocimientos.

Así, en el contexto de la realidad actual, la tarea de discriminar información se vuelve un proceso casi imposible, la creciente información actual, principalmente por vía Internet, le resulta al pensamiento humano cada vez más difícil de filtrar, tratar y dominar; lo cual se traduce en barreras (número ilimitado de fuentes, desconocimiento de los mecanismos de filtrado, organización o apropiación cognitiva de la información) que impiden hacer un uso efectivo de ella. Así, el exceso de información no significa mejor comprensión de la realidad, sino que con frecuencia alarmante produce el efecto contrario (UNESCO, 2005a; Dudziacck 2003; Sahagún, 2004, citados por Peña, 2011).

Así que una de las tareas importantes dentro del quehacer científico es la búsqueda de mecanismos que nos permitan hacer uso efectivo de la información.

La organización del conocimiento tiene como propósito optimizar la utilización del conocimiento documentado, proveyendo acceso físico e intelectual, que a su vez amerita la utilización de normas, estándares, tecnologías y otros medios que coadyuven a la recuperación de los contenidos.

Para tal fin estudia la forma de representar el conocimiento de la manera más eficiente.

Sin embargo no existen formas únicas de organizar el conocimiento, sino que estas se diseñan atendiendo a diferentes criterios o posturas, tales como: la relevancia del tema, conexiones existentes entre los contenidos, demandas de los usuarios, actualidad, entre otros. No obstante, su propósito es viabilizar el uso del conocimiento por parte de todos sus usuarios reales o potenciales (Hjorland, 2003, citado por Peña, 2011).

En cuanto al uso de los conceptos para organizar el conocimiento, Hjorland indica que éstos son concebidos como unidades de pensamiento, y la relación interconceptual es un elemento fundamental ya que los paradigmas, con sus teorías y conceptualizaciones, son por naturaleza instrumentos clasificadores de la realidad con la que interactúa de continuo el individuo. De esta forma, las relaciones semánticas que se establecen entre los conceptos son básicas para sistematizar el conocimiento, pero su significado queda sujeto a la perspectiva teórica desde la cual sean considerados.

La representación del conocimiento incluye otros procesos tales como:

- La descripción
- Catalogación
- Indización y
- Condensación.

Todos ellos, orientados a tomar aquellos rasgos, datos o particularidades que permitan formular una imagen o idea de los contenidos en la mente de quienes

podrían requerir su consulta. Por ende, la ejecución de este proceso tiene una incidencia directa en el acceso, difusión y uso que pueda tener el conocimiento. La representación se efectúa mediante el empleo de un sistema de símbolos legibles para el ser humano (Fernández, 2004; Von der Becke, 1998). Su materialización tiene diversas formas y da lugar a diferentes productos como:

- Ontologías
- Mapas de conceptos
- Resúmenes
- Tesauros
- Catálogos,
- Índices
- Y otros

No obstante, la base de cada uno de ellos son los metadatos, es decir, los datos utilizados para describir otros; su finalidad es dar al usuario las pautas necesarias para ubicar las fuentes o para discriminar la utilidad o no que puedan tener éstas, de acuerdo con sus intereses intelectuales.

En líneas generales, la representación del conocimiento es el proceso que hace posible comunicar contenidos, sirviendo de puente entre el creciente universo de fuentes documentales, cognitivas y de información, y las necesidades intelectuales que poseen los individuos, vale decir, sus usuarios reales y potenciales.

García (1998a, citado por Peña, 2011) añade que la representación del conocimiento, además de mejorar la recuperación, hace posible la reestructuración arquitectónica de los contenidos.

Camaraza (2005, citado por Peña, 2011) indica que la capacidad de información de los humanos se ha vuelto cada vez más selectiva, es decir, que tiende más hacia la precisión. Pero el medio para expresar la necesidad de conocimiento y construir las representaciones de los contenidos queda limitado al lenguaje; éste tiene su raíz en procesos psíquicos y acuerdos sociales, que son complejos en su naturaleza.

Como puede observarse, es imprescindible seleccionar una buena forma de representación del conocimiento cuando se trata de recuperar información de cualquier tipo.

# 1.1.2 Wikipedia como forma de representación del conocimiento

El paradigma wiki se refiere a un sitio web cuyas páginas pueden ser editadas a través del navegador web por múltiples voluntarios de manera libre, ya que las restricciones son mínimas, y así los usuarios pueden crear, modificar o borrar un mismo texto que comparten. La aplicación más relevante ha sido la enciclopedia denominada Wikipedia, que surge en 2001, fundada por Jimbo Wales y Larry Sanger.

El primer uso y con el propósito con el que Wikipedia fue creada es como una enciclopedia. Irónicamente, esta aplicación es la que ha generado más duda. Como se ha señalado, la política de edición abierta ha provocado mucha incertidumbre. Denning (et. al, 2005, citado por Medelyan et. al., 2009). Sin embargo, esta enciclopedia ha ocupado un lugar muy especial en el área de Procesamiento de

Lenguaje Natural, ya que ha sido reconocida como un recurso lingüístico computacional muy importante y que actualmente, gracias a su uso, los resultados de los algoritmos resultan mejores que utilizando recursos tan reconocidos en el área como Wordnet.

Así, Wikipedia puede mirarse como un corpus, como un tesauro, como una base de datos, o como una estructura de red y se han logrado resultados interesantes en las formas de extracción de la información, así como formas de enriquecimiento semántico de contenidos (Pérez, 2011) (Pérez & Gelbukh, 2010). Existe, por ejemplo, una propuesta de generación de trayectorias de aprendizaje basado en esta enciclopedia, utilizando no su contenido, sino la semántica que encierra (Pérez, 2012).

Auxiliándonos en este poder semántico de la Wikipedia, se propone construir recursos de información basados en este paradigma, los cuales, cuando empiecen a crecer, serán susceptibles de ser abordados como un recurso, no solo informativo, sino también semántico, en los que la recuperación de la información podría llevarse a cabo de manera más eficiente (Pérez, 2012).

# 1.1.3. Modelo conceptual de espacios colaborativos basados en el uso de un CMS (Pérez, 2012)

En muchas instituciones, los procesos colaborativos no tienen un medio eficiente de comunicación, y de repositorio de información, así como tampoco cuentan con utilidades para que puedan visualizar de forma eficiente extractos importantes de la información que los miembros del grupo en colaboración generan (Figura 1).

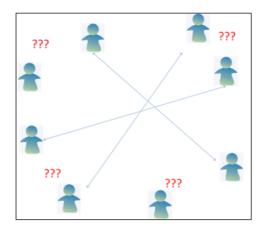


Figura 1. Equipo humano en un proceso colaborativo sin un medio de comunicación propicio

Actualmente las TICs nos proporcionan muchas formas de llevar a cabo procesos colaborativos más eficientes al brindar funcionalidades de comunicación y repositorio.

Sin embargo, como se ha señalado en secciones anteriores, una forma de representación del conocimiento adecuada, será un elemento importante para una extracción de la información eficiente. En el caso de los investigadores, quizá sería importante conocer quiénes colaboraron y en qué colaboró cada participante, conociendo además el momento en que realizó su aportación. También sería interesante extraer resúmenes de los documentos elaborados, etc. Esto no es posible realizarse en plataformas cuya base de datos es completamente cerrada al usuario final y solo puede hacer uso de las funciones preestablecidas. Además, en caso de poder lograr el acceso, quizá la estructura de su bd no sea la más idónea para extraer información particular.

Además de las desventajas mencionadas, el desarrollo de una plataforma informática a la medida resulta costoso. Sin embargo, existen CMS disponibles (p.

ej. joomla, mambo, os.commerce, drupal, etc.), en donde se puede implementar de forma relativamente fácil y no tan costosa, un entorno para colaboración. La limitante en este tipo de plataformas, es que su estructura de base de datos tiene una estructura basada en el modelo relacional, lo cual no permite almacenamientos de información más eficientes que permitan una extracción de la información.

Basado en estudios relacionados con la lingüística computacional, se ha encontrado a Wikipedia como un repositorio de información que tiene una estructura de orden topológico (Pérez, 2011), y tal configuración, permite realizar operaciones que favorecen los resultados en la extracción de la información (Pérez & Gelbuk, 2010). Por tal motivo se propone la modelo conceptual de un espacio de colaboración en el que se utiliza una wiki para el almacenamiento de la información y una interfaz construida a la medida que brinde acceso directo al estado en tiempo real del objeto de estudio.

La wiki nos permite ver a la información bajo una estructura de orden topológico, y se ha documentado ampliamente su uso como un recurso importante dentro del área de lingüística computacional, el cual permite realizar procesos de extracción de la información de forma eficiente.

# 1.2 Definición del problema

El problema es cómo ayudar a mejorar el seguimiento de proyectos de investigación, que coadyuve a la administración de la información generada, desde el punto de vista de sistemas de información.

# Propuesta de solución

Se propuso la implementación de una <u>Plataforma de Colaboración para la</u> Investigación, basado en un modelo conceptual, como se muestra a continuación

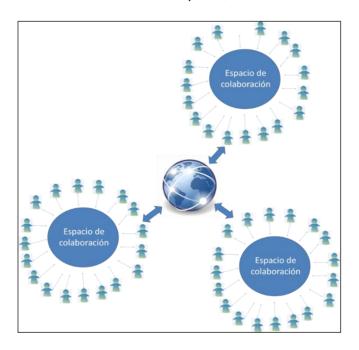


Figura 2. Espacios colaborativos basados en el uso de la web (Pérez, 2013)

# 1.3 Justificación

La implementación de la propuesta de implantación de la <u>Plataforma de</u> <u>Colaboración para la Investigación</u> resulta en dos beneficios principales, uno a corto plazo y uno a largo plazo. El primero, es un beneficio práctico de apoyo informático

a actividades académicas. El segundo, tiene que ver con que la información generada por el uso de esta plataforma, se constituirá como una fuente de información estructurada en orden topológico, y que será motivo de estudio para llevar a cabo distintas tareas de investigación en el área Procesamiento de Lenguaje Natural:

### Primero, a corto plazo, brindar a la institución en donde se implementa:

- a) Un soporte para que el colaborador edite de manera distribuida y compartida, los diferentes documentos resultado de su proceso de investigación, bajo una estructura pre-construida susceptible de ser modificada de acuerdo a las necesidades particulares del equipo de trabajo.
- b) Un repositorio de documentación, importante para un equipo de trabajo, que pueden almacenar y categorizar de forma asíncrona y distribuida.
- c) Una interfaz para poder visualizar en tiempo real, el estado del objeto de estudio.

  Segundo, a largo plazo, el contenido que se genere como producto de su uso a través del tiempo:

Brindará un repositorio de <u>información estructurada en forma topológica</u>, que fungirá como objeto de estudio en la investigación de actividades propias de la Lingüística computacional, particularmente en:

- Procesos de extracción de la información basada en el uso de recursos léxicos grandes (Pérez, 2010).
- Procesos de enriquecimiento semántico a través de una wikificación automática (Pérez & Gelbukh, 2010)

Como podrá observarse en este documento, existe innovación en la forma en la que se pueden adecuar los Sistemas Gestores de Contenidos tales como Wikipedia, no sólo para la creación de materiales con usos prácticos, sino también para transformarse en fuente de investigación en el área de Procesamiento de Lenguaje Natural. Y cuyos resultados seguramente propiciarán el desarrollo científico.

# 1.1.4 Objetivos

# Objetivo general

Implantación del "Modelo conceptual para espacios colaborativos basado en el uso de un CMS" (Pérez, 2012) a través de una aplicación web, en la UAEM CU Valle de Chalco en colaboración con la UABC Instituto de Ingeniería, en el Departamento de Tecnología Educativa.

## Objetivos específicos

- Depurar el Modelo conceptual
- Llevar a cabo el proceso de desarrollo de software para la construcción de la Plataforma de Colaboración para la Investigación.
- Documentar el proceso de desarrollo de software.
- Documentar un modelo de procesos para el uso eficiente de la plataforma.
- Implantar la plataforma

# II. Marco de Referencia Teórico-Metodológico.

Se llevará a cabo una descripción de los fundamentos conceptuales o categoriales de la investigación, así como la exposición de metodología seguida.

Análisis de Resultados. Se integrarán los hallazgos de la investigación, incluyendo los cuadros y figuras elegidas, así como la interpretación de éstos.

### 2.1 Exposición de la metodología

La metodología de trabajo que se propone es de tipo mixta:

- Inicia con una <u>investigación exploratoria</u>, con el propósito de identificar las actividades realizadas en procesos colaborativos de investigación; esta tarea tiene el fin de hacer una depuración del modelo conceptual para espacios colaborativos basado en el uso de un CMS (Pérez, 2012).
- A continuación, sigue un proceso de <u>experimentación</u>. Consiste en el proceso de desarrollo de software para la implementación del modelo conceptual, su implantación y observación de los resultados de su utilización.
- El siguiente paso es medir la eficiencia de la plataforma construida,
   fundamentada en la satisfacción del usuario final.
- Finalmente se documentarán los resultados obtenidos.

Es importante resaltar que esto corresponde a la primera fase de esta propuesta.

En la segunda fase, será necesario esperar a que existan usuarios utilizando la

## **III Resultados**

Se concluyó exitosamente el sistema denominado Plataforma de Colaboración para la Investigación el cual se implantó en el servidor de la UABC bajo la supervisión del Dr. Gabriel López Morteo, quien es colaborador de este proyecto.

http://eduwiki.edumat.org



Figura 3 Ventana principal de la Plataforma

Cada proceso dentro de una investigación tendrá un espacio para que sea identificado por el usuario y pueda insertar información específica.

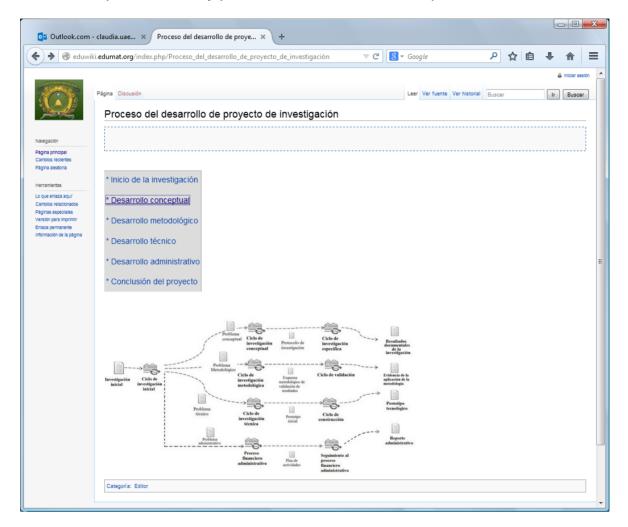


Figura 4 Elementos principales de información que soportará cada proyecto de investigación.

Por ejemplo, desarrollo conceptual tendrá los elementos de información que se muestran a continuación.

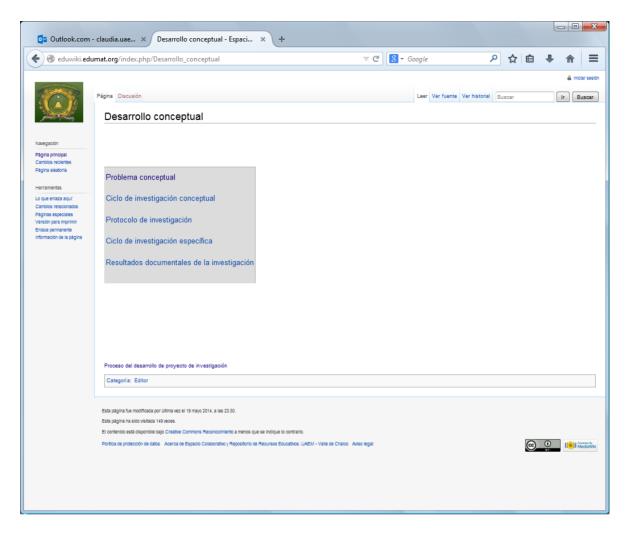


Figura 7 Elementos de información que soporta Desarrollo Conceptual

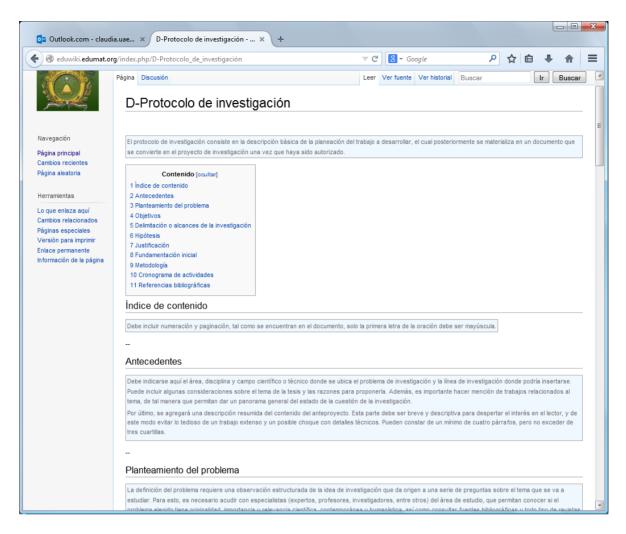


Figura 8 Vista inicial del espacio para registrar el protocolo de investigación.

#### 3.1 Descripción de la construcción del proyecto

Se construyó un sistema con la arquitectura que se muestra en la siguiente figura:

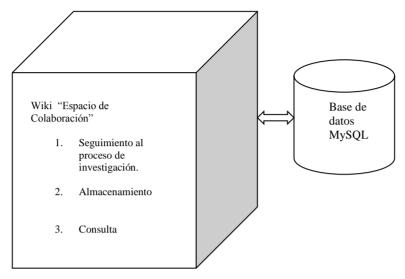


Fig. 3.. Arquitectura propuesta para el modelo conceptual [5]

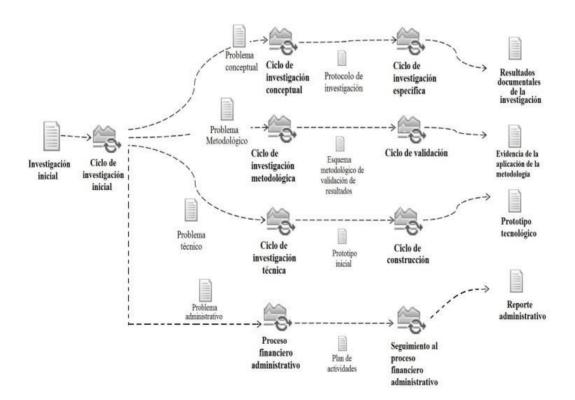
Se conformará un nodo para cada grupo de investigación, que consiste para el alcance del actual proyecto: Un wiki para almacenamiento de información, medio de interacción de forma asíncrona y a distancia para editar, eliminar, crear y visualizar las aportaciones del grupo. La interacción de los integrantes del equipo irá construyendo poco a poco una fuente de información de orden topológico acerca de tema especializados como lo serán los temas del proyecto en el que están involucrados, que proporcionará elementos suficientes al módulo de consulta. Deben establecerse previamente:

- a) Políticas generales de uso,
- b) Espacio para poder especificar políticas de nodos particulares, los usuarios de este módulo las podrán establecer de manera independiente. La estructura de la

Plataforma de Espacios Colaborativos podrá crecer tanto como se requiera, pudiendo haber un nodo central administrativo, el cual podría generar otro tipo de consultas globales (ver figura 2).

## 3.2 Implementación

Se hizo una adaptación del modelo *Action-Research*, el cual permite describir el proceso de investigación como un proceso con tres flujos de trabajo: conceptual, metodológico y técnico. De acuerdo a las necesidades de una institución educativa, es deseable documentar de igual manera el proceso administrativo, en el que se detallan cuestiones de presupuestos y gastos, así como de la documentación requerida para concluir el desarrollo de una tesis. La figura siguiente muestra cómo fue adecuado el modelo.



**Figura 9** Adaptación del modelo *Action-research* para esquematizar el proceso de Investigación en la UAEM

Para su implementación en una wiki, se diseñó un modelo de plantillas en tres capas, la primera es una capa de plantillas de uso general, como tablas, y formatos de texto; la segunda capa es el diseño del contenido de cada una de las páginas que conforman los flujos de trabajo del modelo; finalmente la tercera capa es la capa del usuario, quien solamente asignará el texto deseado y lo asigna a una variable y la página se creará ya bajo la estructura y formatos pre-definidos (ver figura 10).

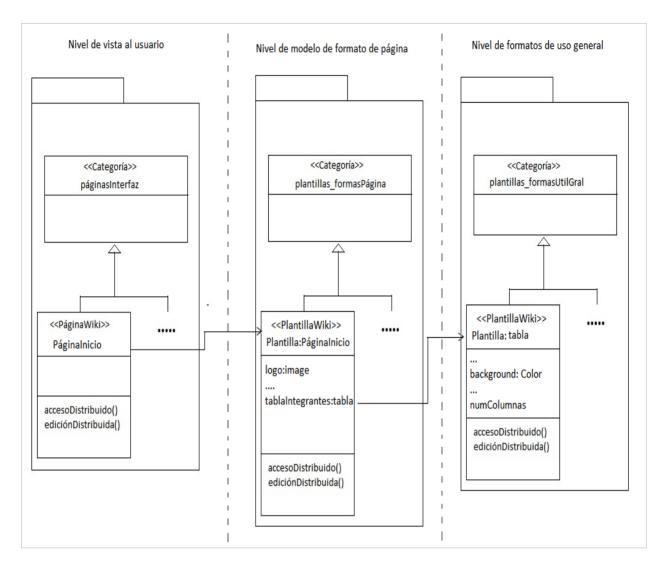


Figura 10. Estereotipo de arquitectura de tres niveles

Una vez generada la estructura para las plantillas se procedió a determinar cuáles y cuántas páginas constaría cada nodo de investigación, a continuación la tabla siguiente describe las páginas que conforman cada flujo de trabajo. Cada una de estas se construyó bajo la arquitectura que se mostró en la figura 3.

**Tabla 1.** Estructura de plantillas para la implementación de un nodo en el espacio colaborativo.

Flujo de trabajo	Plantillas que la constituyen
<ul> <li>Investigación inicial</li> </ul>	Investigación inicial
	Problema conceptual
	Problema metodológico
	Problema técnico
	Problema administrativo
<ul> <li>Desarrollo</li> </ul>	Problema conceptual
conceptual	Ciclo de investigación conceptual
	Protocolo de investigación
	Ciclo de investigación específica
	Resultados documentales de la investigación
<ul> <li>Desarrollo</li> </ul>	Problema metodológico
metodológico	Ciclo de investigación metodológica
	Esquema metodológico de validación de resultados
	Ciclo de validación
	Evidencia de la validación metodológica
<ul> <li>Desarrollo técnico</li> </ul>	Problema técnico
	Ciclo de investigación técnica
	Prototipo inicial
	Ciclo de construcción
	Prototipo tecnológico
<ul> <li>Desarrollo</li> </ul>	Problema administrativo
administrativo	Proceso financiero
	Plan de actividades
	Seguimiento del proceso financiero administrativo
	Reporte administrativo
<ul> <li>Conclusión del</li> </ul>	Resultados documentales de la investigación
proyecto	Evidencia de la validación metodológica
	Prototipo tecnológico
	Reporte administrativo

Una de las páginas que se diseñaron fue el protocolo de investigación, la figura 6 muestra una vista de cómo se visualiza.



Figura 11 Vista de la plantilla Protocolo de investigación

El usuario solamente deberá llenar información vaciando la información en un conjunto de variables, para llenar el protocolo de investigación, lo que se editará se muestra a continuación:

```
{{D-Protocolo de investigacióna | indice = -- | antecedentes = -- | planteamiento = -- | objetivos = --
```

```
|descripción = --
|Hipótesis = --
|justificación = --
|fundamInic = --
|metodología = --
|cronograma = --
|referencias = --
}}
```

Otra plantilla también es la del seguimiento financiero, la cual se muestra en la figura 7 y 8.



Figura 12 Vista de la plantilla Seguimiento administrativo.

# **IV Conclusiones**

El paradigma wiki ofrece una base sólida para el seguimiento y control de la edición de documentos en co-autoría, así como también, permite realizar extracción de información semántica sobre los documentos creados. Por lo que se ha propuesto utilizar este paradigma para la construcción de un espacio virtual de colaboración. Las implicaciones de este desarrollo pueden vislumbrarse desde diversos ángulos:

- Desde el punto de vista del usuario final el investigador, la relevancia de este sistema es que le proporcionaría dos funcionalidades importantes: podrá llevar a cabo tareas de coautoría y administración de documentos en línea de manera asíncrona; y podrá realizar consultas especializadas que difícilmente serían posibles bajo otra estructura de almacenamiento de su información.
- Para poder brindar al usuario final una real satisfacción, el proceso de desarrollo de software deberá contemplar una fase de análisis minuciosa que permita seleccionar una adecuada estructura inicial de categoría, tanto para satisfacer las necesidades actuales (tipos de consultas) como para necesidades futuras.
- No puede dejar de mencionarse un punto de posible conflicto entre la naturaleza de libre acceso de un wiki, y la confidencialidad de procesos de investigación en curso. Es necesario llevar a cabo un proceso de análisis minucioso para establecer de la manera más adecuada las Políticas de usuario.
- La estructura de las categorías en las que se almacenen la información representan una forma de pensamiento de los usuarios. Lo relevante es que al igual que cuando el humano adquiere un nuevo concepto -e inmediatamente

reconstruye las relaciones entre los conceptos almacenados y el concepto nuevo-, la estructura de representación de la información también permite vincular las estructuras ya existentes con las nuevas. La deficiencia estribaría en que no podría reconstruir las antiguas estructuras de la base de datos al menos no de manera automática, como lo podemos hacer los humanos.

Se espera que este prototipo llegue a ser una herramienta utilizada por más de un grupo de investigadores, y por más de una sola universidad, en la medida que se vaya especializando en las funcionalidades que requieren los interesados en esta propuesta.

Se espera que en futuros proyectos se lleve a cabo la fase 2, que consiste en la extracción semántica de información acumulada por muchos proyectos de investigación.

# IV Bibliografía

- (Cunningam, 2013) Cunningham & Cunningham, Inc. http://c2.com/cgi/wiki?WikiWikiOrigin. Accedido el 25 de Octubre de 2013.
- (Harrison, 20011) Harrison, Chris. (2011). Blog de Chris Harrison. http://www.chrisharrison.net/projects/clusterball. Extraido el 25 de Octubre de 2013.
- (Manning, 1999) Manning, C.; Schütze, H. Foundations of Statistical Natural Language Processing.Ed. MIT Press Cambridge Massachusets. (1999). the management knowledge, an international study Delphi.
- (Mediawiki, 2013). MediaWiki.org. MediaWiki. MediaWiki. MediaWiki.org. http://www.mediawiki.org/wiki/MediaWiki. Accedido el 25 de Octubre de 2013.

- (Medeylan et.al., 2009) Medelyan, O.; Milne, D.; Legg, C.; and Witten, I. Mining Meaning from Wikipedia. Cornell University Library. Accedido el 19 de enero de http://arxiv.org/abs/0809.4530. (2009).
- (Milne & Witten, 2008) Milne, D., and Witten, I.H. (2008) An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. Proceedings of the AAAI 2008 Workshop on Wikipedia and Artificial Intelligence (WIKIAI 2008). (2008).
- (Pérez & Gelbukh, 2010) Pérez, C. & Gelbukh, A.. 2010. Wikificación automática de contenido para objetos de aprendizaje. LACLO 2010, 5<sup>th</sup> Latin American Conference on Learning Objects, São Paulo, Brazil, 27 September to 1 October, 2010.
- (Pérez, 2011) Pérez, C. 2010. "Sistema multiagente para la búsqueda de respuestas". Tesis de Maestría en Ciencias de la Computación. Universidad Autónoma del Estado de México. México
- (Pérez, 2013) Pérez, C. 2012. Un modelo conceptual para espacios colaborativos basado en el uso de un CMS. Manuscrito no publicado. Universidad Autónoma de Baja California. México.
- (Peña, 2011) Peña, T.(2011). Organización y representación del conocimiento: Incidencia de las tecnologías de la información y comunicación (1ª ed.). Buenos Aires : Alfagrama.
- (Russell, 1996) Russel, S.; Norving, P. Inteligencia artificial, un enfoque moderno. Ed. Prentice Hall. México. (1996).
- (Scholl, König, Meyer y Heisig 2004) Scholl, S. & Köing, C. & Meyer, B. & y Heisig, P. 2004. The future of
- (Wikipedia, 2013) Wikipedia. Wiki. Wikipedia.org. http://es.wikipedia.org/wiki/Wiki. Accedido el 25 de Octubre de 2013.