

# **CAPÍTULO 3 DISEÑO DE UNA ONTOLOGÍA PARA EL ANÁLISIS DE GRUPOS DE INVESTIGACIÓN**

Después de haber establecido las bases y describir posibles herramientas para el desarrollo proyecto, es momento de conceptualizar el diseño de la ontología, como ya se había establecido con anterioridad, la ontología es un elemento fundamental en este trabajo de investigación, por lo que es de vital importancia que esté bien diseñada para obtener un resultado satisfactorio.

Antes de plantear el diseño de la ontología, se mostrarán algunos ejemplos de aplicación que le han dado los grupos de investigación alrededor del mundo, para poder observar qué resultados obtuvieron con el uso de las mismas, y qué poder esperar después de diseñar la que se propondrá para este trabajo.

## **3.1 Aplicaciones de ontologías y su relación con grupos de investigación**

A continuación mostraremos algunos usos que se le han dado a las ontologías por grupos de investigación alrededor del mundo, como ya se dijo con anterioridad, esta actividad ayudará a establecer un mejor diseño para nuestra ontología, debido a que gracias a estudiar los siguientes casos de uso de las ontologías, nos dejará algunas ideas de cómo diseñar la nuestra.

### **3.1.1 Contenidos web accesibles a invidentes mediante ontologías**

En la actualidad la internet es el medio más popular para buscar información, la llamada “carretera de la información”, paulatinamente se ha venido convirtiendo en la biblioteca más visitada del mundo, además de contar con las grandes ventajas de que no se necesita una tarjeta para ingresar, ni para retirar material y mucho menos preocuparnos por regresarlo a tiempo. Sin embargo, este

gran almacén de información tiene, al igual que muchas de las bibliotecas físicas, una gran desventaja, no están diseñadas para personas con debilidad visual de cualquier tipo (miopía, hipermetropía, astigmatismo, entre otros), entonces para cualquier persona que padezca una discapacidad visual el utilizar la información que necesita resulta un verdadero reto, reto que en ocasiones no puede ser superado, sin embargo, existen instituciones que hacen que dicha tarea no sea tan difícil de llevar a cabo por estas personas.

De lo anterior, se plantea el reto de cómo ayudar a estas personas a encontrar lo que necesitan dentro de la red, sin que su discapacidad sea un problema. Entonces el Departamento de Informática de la Universidad de Extremadura, presenta una propuesta basada en ontologías para ayudar a que las personas con discapacidad visual superen este (Lozano-Tello et. al, 2002).

Para que las personas con discapacidad visual puedan superar esta situación, lo primero que observaron, y que ya se explicó en el capítulo anterior, la forma en que se realizan las búsquedas en la actualidad se realiza por palabras claves y la relevancia que estas tienen en el documento, pero la cuestión es que esta forma de búsqueda no es muy precisa, aunado al problema que tiene una persona que tiene una discapacidad visual, es una tarea maratónica y titánica el que estas personas puedan encontrar información útil para ellos sin ayuda alguna.

Considerando la situación anterior, consideraron factible apoyarse en el concepto de la web semántica, que es básicamente, como dice su nombre, la red basada en los significados, dando como producto el KAICO (Kit de Accesibilidad para Invidentes Con Ontologías). En la figura 3.1 se encuentra esquematizada la arquitectura KAICO.

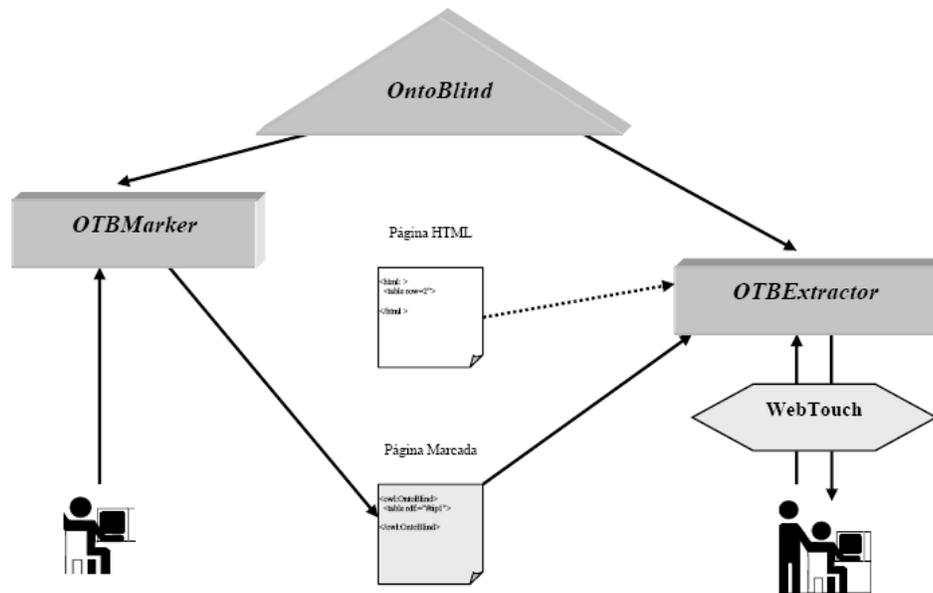


Figura 3.1 Arquitectura Kaico. Fuente: (Lozano-Tello et. al, 2002)

Para poder comprender la arquitectura y por ende el funcionamiento de Kaico, tenemos que analizar los componentes que la conforman, que son los siguientes:

- Ontoblind: La ontología del Kit, es la parte que se inyecta en la página web, contiene los conceptos y sus relaciones.
- OTBMarker: Editor web que permite hacer anotaciones semánticas en la página web basada en ontoblind
- OTBExtractor: esta aplicación interpreta las anotaciones referenciadas por ontoblind de las páginas web realizadas por otbmaker y la hace accesible a la persona con discapacidad visual, mediante los recursos que se tengan para hacerlo
- Webtouch: Son los recursos con los que se le da accesibilidad a la información a las personas con discapacidad visual.

Entonces el funcionamiento de Kaico de manera resumida sería de la siguiente forma: una persona crea las páginas web con otbmaker, para que después otbextractor interprete la información de las páginas web haciendo referencia de las mismas hacia ontoblind, para después mostrársela al usuario por medio de webtouch.

### **3.1.2 Ontologías para la Búsqueda Semántica en Laboratorios de Investigación y Desarrollo: OLID**

“La información manejada por un laboratorio de investigación y desarrollo debe estar organizada para facilitar las búsquedas y el intercambio de conocimiento. Una forma de representar tal conocimiento es por medio de una ontología. En este trabajo se propone *OLID*, la cual es una ontología que conceptualiza el conocimiento del dominio de un laboratorio de investigación y desarrollo dando el primer paso para poder implementar una aplicación semántica con la capacidad de inferir nuevos conocimientos.” (Hurtado y Sequeda, 2006)

En la actualidad, un gran número de organizaciones (empresas, universidades, corporativos, y más), están empezando a adoptar las tecnologías semánticas de la web para enfrentar de mejor forma la problemática que surge al tener información en enormes cantidades. Este proyecto se desarrolla en el contexto educativo y se supone que se cuentan con los siguientes datos: universidad X con un Y número de carreras y que cada carrera tiene un Z número de alumnos, cada alumno con su información personal, nombre, carrera, semestres cursados, créditos obtenidos. Para realizar una consulta sencilla como por ejemplo saber cuántos créditos tiene un alumno se tendría que realizar una búsqueda de la siguiente forma:

- Buscar la universidad correcta
  - Buscar la carrera correcta
    - Buscar al alumno
      - Buscar el dato que necesitamos

Dentro de un sistema local se puede dar el caso en que no sea tan difícil, pero pensando en una persona ajena al alumno (padres, tutores, fundaciones que otorgan becas, entre otros), necesitan obtener un dato del alumno, y sólo saben que el estudiante asiste a una universidad en el estado de Sonora, tardarían una cantidad de tiempo considerable en encontrar solamente la universidad, y aun ya con el conocimiento de qué universidad se trata, buscar los datos del alumno dentro de dicho sitio les tomaría también bastante tiempo si todo lo anterior se hace con el proceso de búsqueda actual. Sin embargo, con las tecnologías semánticas sería más rápido, hipotéticamente hablando, se tiene la siguiente ontología, para resolver la búsqueda de los créditos de un alumno dentro de una carrera de una universidad (ver figura. 3.2)

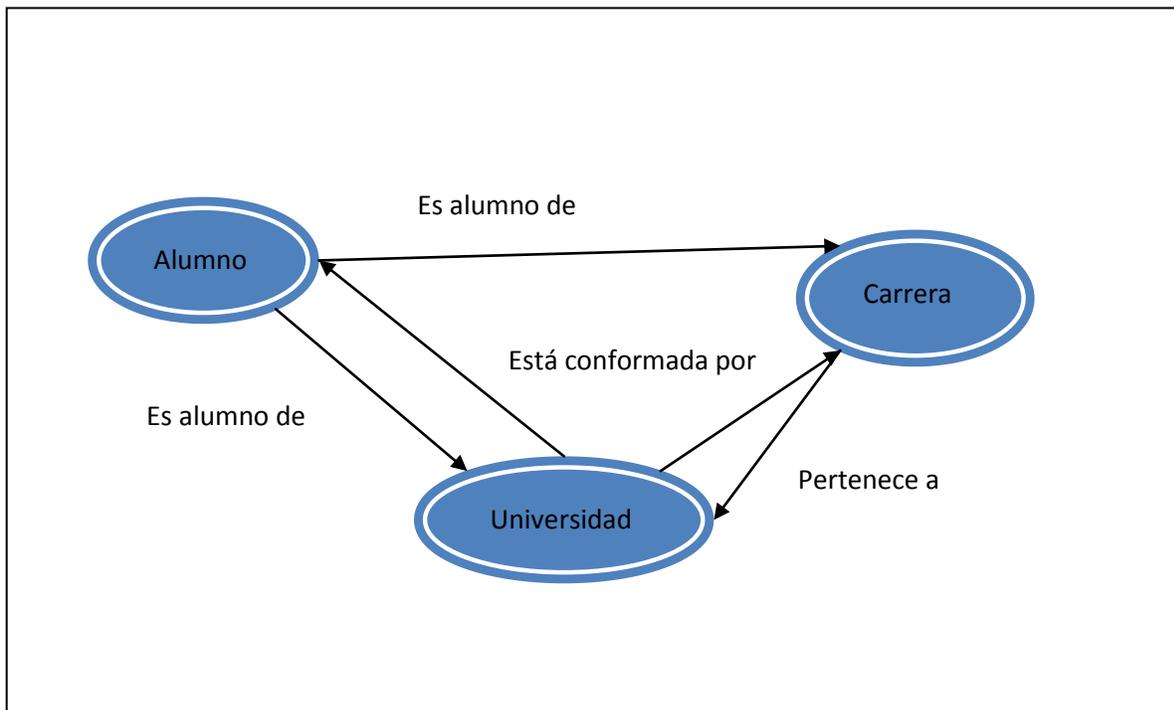


Figura 3.2 Ontología universidades

Una estructura muy sencilla para fines prácticos de la ejemplificación, además de suponer que cada clase ya contiene sus datos, sería posible preguntarle a la ontología ¿cuántos son los créditos que tiene Miguel Pérez en la Universidad de Sonora? Y la ontología obtendrá el siguiente resultado, Miguel Pérez es alumno de la Universidad de Sonora de la carrera *Ingeniería en Sistemas de Información* y tiene 270 créditos.

Como se observa, la aplicación muestra información que no fue solicitada, y por efectos de la ontología y las relaciones que existen entre los conceptos de la misma, nos la puede mostrar. En este caso existe una relación entre los conceptos Alumno y Carrera, la ontología es capaz de aprovechar dicha relación y obtener la información de la carrera en que cursa el alumno sin haberla solicitado.

Después de haber ejemplificado y colocarnos en el contexto educativo podemos seguir con OLID (Ontologías para la búsqueda semántica de Laboratorios de Investigación y Desarrollo) la cual es una herramienta que nace a raíz de satisfacer una necesidad parecida a la del ejemplo.

Al ser una tecnología parecida a la de nuestro proyecto, Olid es una de las referencias principales en esta investigación, la problemática que resuelve Olid es la siguiente: (Hurtado y Sequeda, 2006), la información generada por un laboratorio de Desarrollo e investigación es bastante amplia desde los integrantes que conforman dicho laboratorio y todas sus ramificaciones, proyectos, publicaciones, recursos y tópico.

Entonces cuando una persona ya sea que fuese parte del laboratorio o la universidad, o una persona completamente ajena a cualquiera de las dos instancias anteriores deseara encontrar información sobre cualquiera de los conceptos anteriormente mencionados, ocurría el problema principal que se ha venido mencionando reiteradamente, por lo que hay que facilitar los procesos de búsqueda dentro de la internet y además que dicha información sea útil.

Es por ello que se piensa en las tecnologías de la web semántica, las ontologías, lo que después dio paso a la creación de la ontología que forma la base de Olid (ver figura 3.3), la cual realiza la búsqueda de la información que el usuario solicita, como por ejemplo los créditos del alumno de una universidad. Esto se logra debido a las inferencias que pueden obtenerse por los conceptos y sus relaciones para poder obtener información que en un motor de búsqueda convencional pasaría desapercibida, lo cual es un lujo que un laboratorio de este tipo no puede costear.

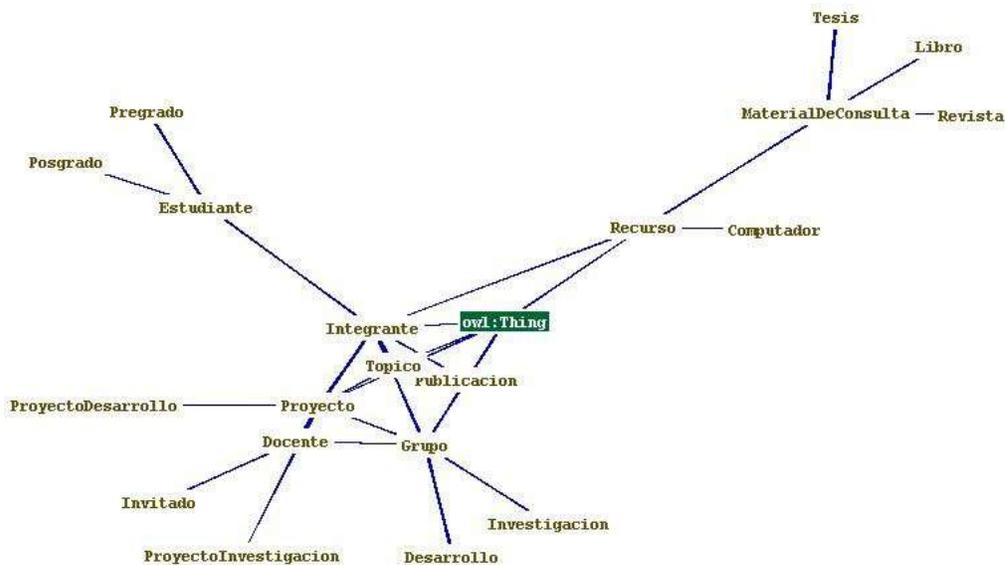


Figura 3.3 Ontología OLID. Fuente: (Hurtado y Sequeda, 2006)

La implementación de Olid tiene como esencia, el facilitar a cualquier persona que necesite información generada dentro ese laboratorio, trabajos, investigaciones en proceso y terminadas, memorias de congresos, resumiendo todas las publicaciones en sí.

### 3.1.3 Ontologías auditoría y modelado multi-perspectivo del conocimiento

Este es un trabajo cuyo fin es el de producir una ontología que represente todos los aspectos del modelado multi-perspectivo, el cual tiene que toda fuente de conocimiento puede ser representada por las cuestiones, ¿Qué es?, ¿Quién la posee?, ¿Cómo es usado?, ¿Dónde se encuentra?, ¿Por qué es necesitado? O ¿Por qué existe?, de esta forma es posible cuantificar la cantidad de conocimiento que tiene cada persona, o las actividades que pueden ser apoyadas con cierta fuente de conocimiento.

Para poder realizar lo anterior, auditar el conocimiento, es necesario representar el conocimiento, para poder llevar a cabo dicha tarea es necesario utilizar el sistema de Zachman (Kingston, 2001).

El sistema de Zachman consiste de 6 columnas representando, precisamente las diferentes perspectivas del modelado (quién, qué, cómo, cuándo, dónde y porqué). Con esto se pretende identificar los diferentes niveles de abstracción que pueda tener el conocimiento o lo que se esté estudiando, dándonos como su nombre lo describe una perspectiva múltiple del conocimiento para poder auditar de mejor forma el conocimiento.

Sin embargo, la medición del conocimiento es una tarea muy difícil al igual que es difícil medir la aportación de conocimiento de cada persona para poder realizar una tarea o proceso cualquiera. Es por esto que se empezó a idear herramientas que pudieran ayudar a medir el conocimiento, como lo son los mapas de conocimiento, los cuales funcionan de una forma muy sencilla partiendo de la siguiente idea “si se quiere saber o hacer B, primero tienes que saber A”.

Lo anterior fue implementado por Kingston (2001) en una ontología para poder realizar este trabajo de una forma más eficaz y eficiente. El diseño de la ontología contiene las clases principales: event, organization, person, publication y research topic. Al usar esta ontología deberá de ser posible, decodificar todas y cada una de las partes relevantes del conocimiento. Ya que almacena de una forma muy detallada todo el conocimiento generado dentro de una organización.

Sin embargo, cómo medir si una fuente de conocimiento es mayor que otra, la ontología aparte de almacenar de una forma muy detallada todas las fuentes de conocimiento, también tiene una serie de slots en las fuentes de conocimiento, que sirven para llevar una “bitácora”, por así decirlo, sobre las veces que esta fue consultada. Es decir, aquella que fue más veces consultada tiene mayor valor que aquella que no.

### **3.1.4 Consideraciones derivadas de las aplicaciones anteriores**

A pesar de que estas tecnologías son nuevas, útiles y muy avanzadas, aún se encuentran en desarrollo, lo cual es algo en lo que tenemos que hacer hincapié y es muy importante recordar. Hay que recordar que la tecnología está para ayudar a llevar al humano a llevar cabo tareas muy laboriosas, tardadas o que el realizaría de manera imprecisa y más cuando tiene que manipular gran cantidad de información.

Estos últimos ejemplos sirven para analizar cómo los grupos de investigación han aprovechado las ventajas que representa trabajar con ontologías, la flexibilidad que dan al momento de realizar una búsqueda, ya que las realiza por medio de inferencias entre las relaciones de los conceptos de la misma ontología, además, también tienen una gran utilidad en cuanto a la administración del conocimiento se refiere.

Una de las cosas más importantes que cabe destacar de los ejemplos que se analizaron, es el hecho de que dichas tecnologías terminaron por impactar en grupo demográfico, hecho que es importante observar, ya que en los escritos revisados para el diseño de una ontología, siempre hablan de que los conceptos tienen que estar definidos a la perfección, las reglas para las relaciones, la taxonomía, y muchos más detalles que con llevan el hecho de realizar una ontología, pero son pocos los que mencionan algo sobre analizar la población hacia la cual estará orientada la ontología. Algo que se debe de considerar como uno de los principales detalles que se deben de tomar en cuenta, sólo basta hacer preguntas muy sencillas como, ¿Quién va a hacer uso de la ontología?, ¿Cuál es el rango de edad al que pertenecen los futuros usuarios?, ¿Tienen alguna discapacidad?, ¿Tienen para pagar por un servicio como este?, Y las que se crean pertinentes para poder establecer un perfil de usuario para la ontología.

Es importante tomar en cuenta, que no importa si tenemos el mejor sistema del mundo, con el mejor diseño, que realice funciones mas allá de lo que nos pidió

el cliente, si al usuario no le gusta por cualquier motivo, que no le entienda, que se canse de esperar mucho por un proceso, o cualquier otra función que al usuario le desagrade, este último no querrá usarlo y nuestro sistema por mejor que sea de los demás no servirá de nada.

Es por esto que se recomienda darle un mayor énfasis a esta parte del diseño de la ontología, por la misma razón que veíamos anteriormente, ya que no sería lo mismo diseñar la ontología para que la utilicen sólo profesores e investigadores, a diseñarla para también la usen otros usuarios como alumnos o personas que sólo quieran informarse sobre el tema.

## **3.2 Ontología para el análisis de grupos de investigación**

A continuación, se procederá a diseñar la ontología que se utilizara para el caso de estudio del proyecto, debe de responder las preguntas expuestas, para poder tener un resultado satisfactorio.

Se necesita que la ontología tome los datos desde una base de datos, ya que esta información, la de los grupos de investigación, ya existe en esa tecnología (bases de datos).

### **3.2.1 Preguntas a responder por la ontología**

Para plantear el diseño de la ontología que se utilizará durante este proyecto, se tiene que establecer una serie de cuestiones que esta última deberá ser capaz de responder. Para poder llevar a cabo lo anterior, es necesario delimitar y analizar el dominio que interesa representar en la ontología y analizar las distintas relaciones que existen entre los conceptos para que la ontología pueda realizar inferencias.

En el caso de la ontología debe responder preguntas, que obviamente se encuentran dentro del contexto de los grupos de investigación, en particular los llamados cuerpos académicos, que contemplen algunos de los aspectos de interés de este proyecto como son: publicaciones, colaboradores en trabajos de

investigaciones terminadas o aquellas que se encuentren en proceso, quiénes colaboran con dicho grupo de investigación, por mencionar algunas.

Definido lo que tiene que responder, entonces es posible diseñar la ontología apropiada que será la base para el resto de las etapas de este proyecto. Esto es fundamental, ya que un mal diseño de la ontología terminaría en un resultado no satisfactorio para el mismo.

A continuación se plantean una serie de preguntas que interesan que responda nuestra ontología.

Las siguientes preguntas son las que responderá nuestra ontología, no son las definitivas, ya que puede darse el hecho de que no contemos con conceptos o las relaciones para que se les dé respuesta a las cuestiones siguientes:

- ¿Quiénes son los integrantes del grupo de investigación?
- ¿Quiénes colaboran al grupo de investigación?
- ¿Qué publicaciones ha producido el grupo de investigación?
- ¿Qué trabajos tiene en proceso el grupo de investigación?
- ¿Qué trabajos ha terminado el grupo de investigación?
- ¿Quiénes son los colaboradores que tienen licenciatura?
- ¿Quiénes son los colaboradores que tienen posgrado?
- ¿Quiénes son los integrantes que tienen el grado de maestro?
- ¿Quiénes son los integrantes que tienen el grado de doctor?
- ¿Quiénes son los integrantes que no tienen posgrado?

Todas las anteriores son preguntas un tanto sencillas para la ontología, la siguiente es una pregunta que forzara a la ontología a realizar más de una inferencia entre conceptos para que le sea posible responder al usuario, es en este tipo de preguntas donde en realidad vemos el fuerte y el propósito de utilizar una ontología para nuestra investigación:

- Dime las publicaciones de x autor publicada por x medio

Es sencillo notar que la pregunta anterior tiene un grado más alto de complejidad para un buscador convencional, debido a que daría un resultado impreciso y posiblemente erróneo al deseado, debido a la sencilla razón de las relaciones entre los conceptos que solo es posible manejar por medio de la ontología. Lo anterior aunado a que en la mayoría de las preguntas se relacionan 3 o más conceptos e incluso subclases de ellos mismos, que para un buscador convencional resultaría una tarea completamente imposible el poder realizar las inferencias, de hecho, un buscador convencional no hace estas inferencias por lo que resultaría inútil querer que las hiciera. Es por esto último que las ontologías son la base para este tipo de sistemas que permiten realizar este tipo de búsquedas que un buscador convencional no es capaz de realizar.

### **3.2.2 Diseño de la ontología**

Ya estableciendo las preguntas que se desea sean respondidas por nuestra ontología y analizando como otros grupos de investigación han sacado provecho de las ontologías, es posible empezar a establecer un diseño para la ontología de nuestro proyecto. Se tuvo la gran oportunidad de analizar la ontología de Olid porque es muy similar a lo que se desarrolla en este proyecto, motivo por el cual se tomarán algunas ideas y fundamentos de dicha ontología.

Como ya fue mencionado en capítulos anteriores, los conceptos son los cimientos de toda ontología, y es una parte en la cual hay que poner una atención muy especial para obtener buenos resultados. Es importante mencionar, que para

el diseño de la ontología, se deben tomar en cuenta las preguntas a responder, y para este proyecto en particular, también se está tomando en cuenta la estructura de una base de datos que contiene información de los grupos de investigación. En esta base de datos se encuentra almacenada información que también deberá reflejarse en los conceptos y relaciones de la ontología.

A continuación se listarán los conceptos de la ontología en el dominio específico de interés para este proyecto:

- Grupo de investigación: Como su nombre lo indica, es el concepto que alberga el objeto de análisis en este proyecto.
- Integrantes: contiene a todos los miembros del grupo de investigación
  - Profesores: Subclase que contiene a todos aquellos miembros sin posgrado
  - Maestros: Subclase que contiene a todos aquellos miembros que tienen el grado de maestría, pero no tienen doctorado
  - Doctores: Subclase que contiene a todos aquellos miembros que tienen el grado de doctorado, y que por consecuencia tienen maestría.
- Colaboradores: Contiene a todos aquellos individuos que participan en diversas actividades, pero que no pertenecen al grupo de investigación.
  - Grupos de investigación: Subclase que contiene a todos los colaboradores que pertenecen a otro grupo de investigación.
  - Alumnos: Subclase que contiene a todos los alumnos que realizaron una colaboración con el grupo de investigación

- Licenciatura: Subclase de la subclase Alumnos que contiene a todos aquellos alumnos de licenciatura que participaron en proyectos y/o publicaciones.
  - Postgrado: Subclase de la subclase Alumnos que contiene a todos aquellos alumnos de nivel posgrado que participaron en proyectos y/o publicaciones.
- Investigaciones: Contiene a todos los trabajos de investigación del grupo.
  - En proceso: Subclase que contiene a todos los trabajos de investigación que aún no se encuentran terminados.
  - Terminados: Subclase que contiene a todos los trabajos de investigación que se encuentran terminados.
- Publicaciones: Contiene a toda producción bibliográfica que el grupo de investigación haya realizado.
  - Memorias: Subclase que contiene a todas las memorias de congreso a las cuales hayan asistido miembros del grupo de investigación.
  - Tesis: Subclase que contiene todo trabajo de tesis realizado dentro de alguna investigación por el grupo de investigación.
  - Libros: Subclase que contiene todo material literario producto de alguna investigación o experiencias de miembros del grupo.

- Artículos: Subclase que contiene todo artículo publicado en revista arbitrada, no arbitrada o de difusión, producto del grupo de investigación.

Como se puede observar, una vez definidos los conceptos de nuestra ontología, ahora se debe establecer las relaciones entre ellos, ya que estas serán los principios sobre los cuales el motor de nuestra ontología realizará sus inferencias; esto se puede ver en la figura 3.4 que aparece a continuación.

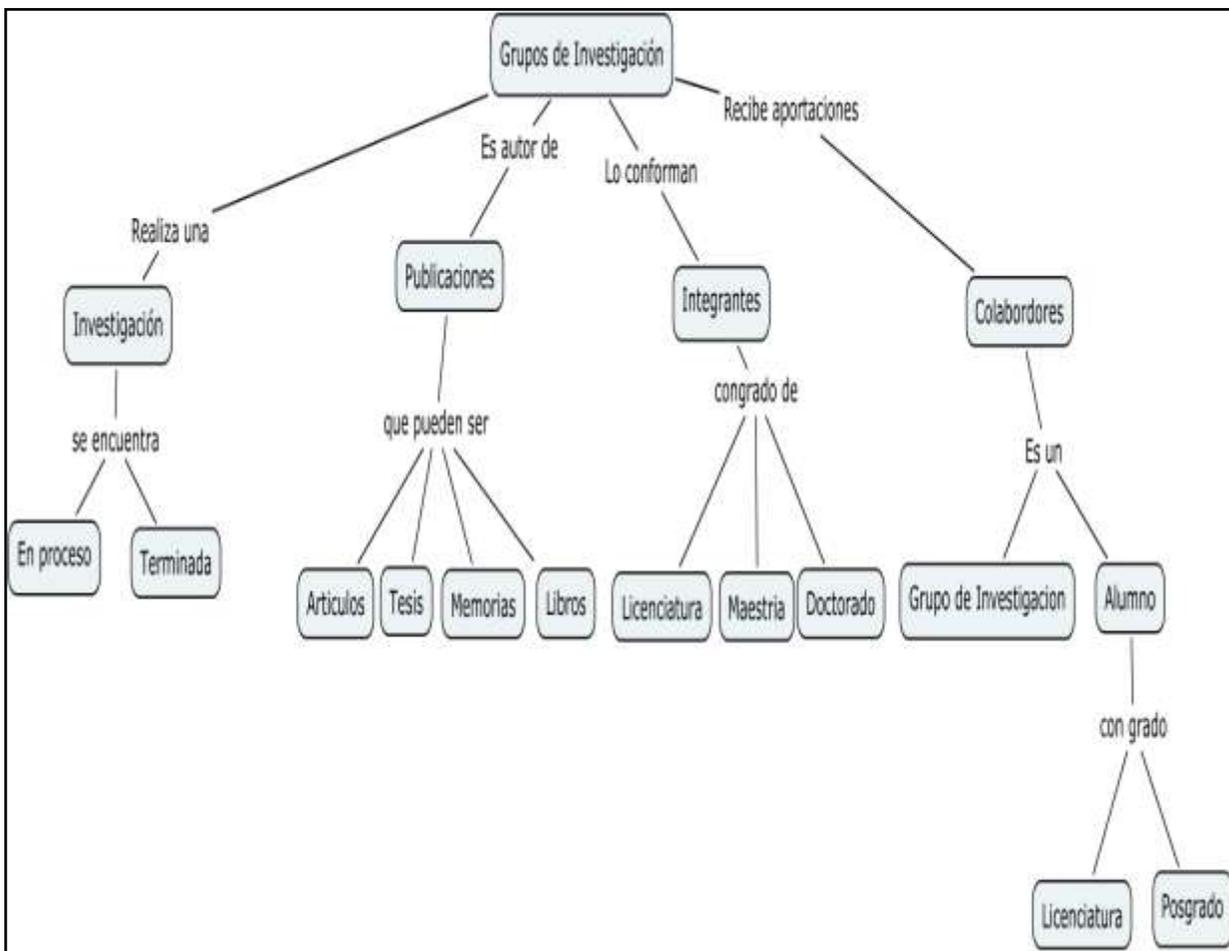


Figura 3.4 Primer prototipo de OAGI

Como se mencionó previamente, se debe tomar en cuenta que nuestro diseño de conceptos y relaciones de la ontología parte también de la base de datos con la cual se alimentará la misma (instancias). Es por esto que se tiene que tomar en cuenta la estructura de la base de datos para poder diseñar la ontología lo más cercano posible a la fuente de información para que la transición de la misma, hacia la ontología no represente mayor problema.

Entendiendo esto, se tiene que nuestra base de datos, donde está guardada la información que se desea manipular con la ontología tiene la siguiente estructura (ver figura 3.5):

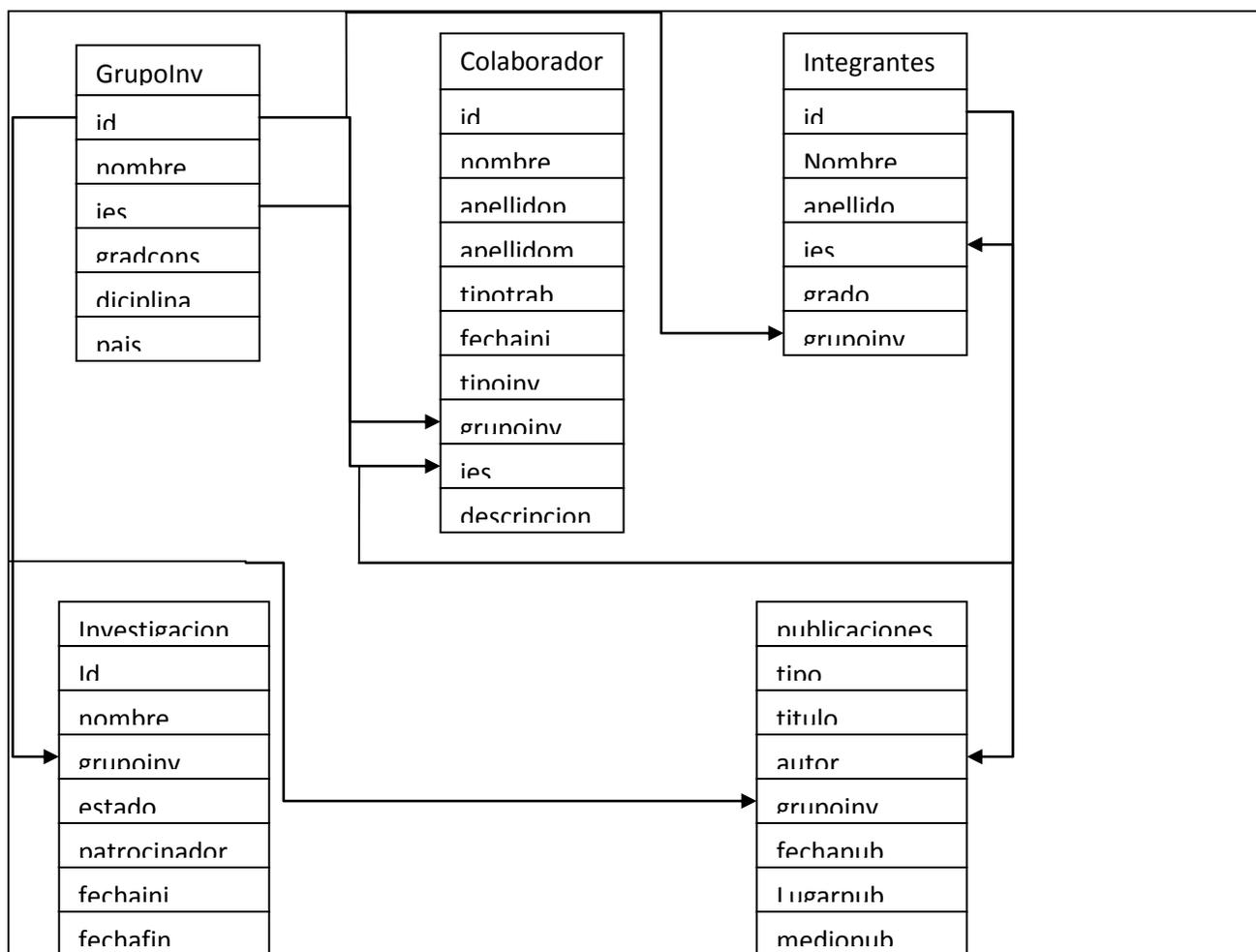


Figura 3.5 Estructura de Base de Datos "CA"

Entonces, si consideramos el diseño de la figura 3.4 (que contempla las preguntas que interesan que responda nuestra ontología) y es cotejado con la estructura de la base de datos de la figura 3.5, es posible ver que el diseño es muy parecido con la estructura que tiene la base de datos, por lo que el diseño de la ontología que se había considerado inicialmente, tiene factibilidad de ser utilizado para los fines del proyecto.

Entonces, ya teniendo la ontología diseñada se tiene que proceder al siguiente paso, elegir las herramientas con la que se desarrollará el proyecto y que se explica de manera detallada en el siguiente capítulo.