

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación

Ingeniería en Ciencias de la Computación

Descubriendo conocimiento en redes sociales a partir de ontologías

Presenta:

Pablo Camarillo Ramírez

Director: Dr. Abraham Sánchez López

Puebla, Puebla, 2012

Dedicado a mis padres, hermanos y sobrinos

Resumen

Durante las últimas décadas, la forma en la que la humanidad comparte su conocimiento ha cambiado, de compartirlo a través de libros, revistas, periódicos, etc., ahora usa el Internet para compartir de forma instantánea cualquier tipo de conocimiento digitalizado, todo esto a través de las redes sociales u otras formas de comunicación. Las redes sociales poco a poco se destacan como la forma más común de interactuar con otras personas, personajes, instituciones e incluso políticos a través de la red de redes, el Internet. Sin embargo todo ese conocimiento inyectado a la red no tiene ninguna clase de relación semántica entre sí, es decir no hay forma que la computadora o dispositivo que lo consulta pueda conocer el conocimiento contenido en ese recurso publicado en las redes sociales.

En este trabajo de tesis se presenta una propuesta para relacionar semánticamente el contenido publicado en una red social. Se usará una semántica basada en la representación del conocimiento mediante una ontología de dominio. El dominio que abarcará esta ontología, es un dominio de conceptos científicos mencionados en las publicaciones realizadas en la red social que hemos desarrollado. Esta red social la llamamos *Moveek* y su objetivo es crear redes de usuarios que compartan y difundan publicaciones con cierto contenido científico. Cada publicación, de manera automática, resaltará los conceptos que el sistema ha detectado como conceptos científicos contenidos en la ontología. De esta forma el usuario puede consultar más publicaciones relacionadas semánticamente con el concepto resaltado.

Nuestro trabajo tiene como objetivo crear una plataforma basada en la composición de servicios Web capaz de brindar las funcionalidades de una red social y que al mismo tiempo otorgue diversas funcionalidades propias de la web semántica tal como la anotación semántica y consultas semánticas a la información publicada en esta red.

Índice general

| | |
|---|-----------|
| Índice general | I |
| Índice de figuras | III |
| Índice de cuadros | V |
| 1 Estado del Arte | 3 |
| 1.1. Redes Sociales | 3 |
| 1.2. Servicios Web | 7 |
| 1.2.1. ¿Qué es un servicio Web? | 7 |
| 1.3. ¿La Web 2.0 + Web Semántica = Web 3.0? | 10 |
| 1.4. Web Semántica | 11 |
| 1.5. Conocimiento | 13 |
| 1.6. Ontología para la Web semántica | 14 |
| 1.6.1. ¿Qué es una ontología? | 14 |
| 1.6.2. Ontología para la Web | 16 |
| 1.6.3. Ontología y servicios Web | 16 |
| 1.7. La anotación semántica | 17 |
| 1.7.1. Algunas definiciones | 17 |
| 1.7.1.1. La automatización en la creación de las anotaciones | 18 |
| 1.7.1.2. El almacenamiento de las anotaciones y de sus recursos | 20 |
| 1.7.1.3. La utilización de las anotaciones | 21 |
| 1.7.2. Los lenguajes de la anotación semántica | 21 |
| 1.7.2.1. La pirámide de los lenguajes de la Web semántica | 22 |
| 1.8. Aportación | 23 |
| 2 Ontología Propuesta | 25 |
| 2.1. Propuesta | 25 |
| 2.1.1. Desarrollo de la ontología | 27 |
| 2.1.1.1. Determinar los requerimientos | 28 |
| 2.1.1.2. Reutilización de ontologías | 29 |
| 2.1.1.3. Elaboración del modelo conceptual | 31 |
| 2.1.1.4. Implementación | 35 |
| 2.1.1.5. Validación de la ontología | 40 |
| 2.2. Conclusión | 41 |
| 3 La anotación en la Web semántica | 45 |
| 3.1. La anotación y la Web semántica | 45 |
| 3.1.1. Los recursos terminológicos u ontológicos (RTO) | 45 |

| | | |
|----------|--|------------|
| 3.1.1.1. | Las taxonomías | 46 |
| 3.1.1.2. | Los tesauros | 47 |
| 3.1.1.3. | Las ontologías | 47 |
| 3.1.2. | Los RTO y la anotación semántica | 49 |
| 3.2. | Herramienta de anotación | 51 |
| 3.3. | Discusión | 55 |
| 3.3.1. | Síntesis con respecto a la anotación semántica | 55 |
| 3.4. | Conclusión | 56 |
| 4 | La red social Moveek | 57 |
| 4.1. | Análisis y diseño de la red social Moveek | 57 |
| 4.1.1. | Análisis | 57 |
| 4.1.2. | Diseño | 60 |
| 4.2. | Servicios Web | 64 |
| 4.2.1. | ¿Por qué usar composición de Servicios Web? | 64 |
| 4.2.2. | Descubriendo el conocimiento: la combinación de servicios Web. | 66 |
| 4.3. | Conclusión | 73 |
| 5 | Resultados | 75 |
| 5.1. | Funciones de la red social | 75 |
| 5.2. | Consultas semánticas | 78 |
| 5.3. | Conclusión | 83 |
| 6 | Conclusiones y trabajo futuro | 85 |
| | Bibliografía | 87 |
| A | Glosario de términos | 91 |
| B | Clases de la Ontología | 99 |
| C | Diagramas | 101 |
| D | Herramienta GATE | 107 |
| E | WSDL | 111 |

Índice de figuras

| | | |
|-------|---|----|
| 1.1. | Representación gráfica de una red social. | 3 |
| 1.2. | Desarrollo de la Web social. | 5 |
| 1.3. | Las principales tecnologías de los servicios Web y sus relaciones. | 9 |
| 1.4. | Las tecnologías de los servicios Web y los principales actores. | 9 |
| 1.5. | Los contextos compartidos-ontologías. | 14 |
| 1.6. | Pirámide de los lenguajes de la Web semántica en 2005. | 22 |
| | | |
| 2.1. | Descripción de la metodología de desarrollo de la ontología. | 28 |
| 2.2. | Propuesta de jerarquización de los términos del Glosario de términos | 33 |
| 2.3. | Relaciones entre las clases de OntoScience. | 36 |
| 2.4. | Pantalla inicial del sistema Protégé | 37 |
| 2.5. | Clases de OntoScience | 37 |
| 2.6. | Ejemplo de configuración de la propiedad de anotación <code>rdfs:label</code> | 38 |
| 2.7. | Propiedades de Objeto de la Ontología OntoScience. | 38 |
| 2.8. | Ejemplo de consulta a OntoScience | 39 |
| 2.9. | Ejemplo de consulta de publicaciones realizadas por un autor en un ámbito. | 39 |
| 2.10. | Fragmento del código en Java que invoca a métodos del API de Protégé | 40 |
| 2.11. | Resultado de la validación de consistencia a través del API de Protégé | 41 |
| 2.12. | Resultado de la consulta realizada con el API de Protégé. | 41 |
| 2.13. | Modelo inferido por el razonador <i>Pellet</i> de la taxonomía de nuestra Ontología | 43 |
| | | |
| 3.1. | Extracto de la taxonomía de OntoScience. | 46 |
| 3.2. | Extracto del artículo "El clan Coppola". | 50 |
| 3.3. | Resultado del servicio de extracción de la información | 53 |
| 3.4. | Diagrama de actividad del proceso de anotación semántica | 54 |
| 3.5. | Ejemplo de publicación anotada en HTML | 54 |
| | | |
| 4.1. | Diagrama de Casos de Uso de Moveek | 58 |
| 4.2. | Diagrama de Clases de Análisis de Moveek. | 59 |
| 4.3. | Modelo de Navegación general de Moveek | 60 |
| 4.4. | Diagrama de Secuencia del Caso de Uso Crear Perfil | 61 |
| 4.5. | Diagrama de Secuencia del Caso de Uso Consultar Perfil | 61 |
| 4.6. | Diagrama de secuencia del caso de uso <i>Realizar Publicaciones</i> | 62 |
| 4.7. | Diagrama de secuencia del caso de uso de aceptar amistad | 63 |
| 4.8. | Diagrama de Secuencia del Caso de Uso de Enviar Solicitud | 64 |
| 4.9. | Diagrama de Clases de Diseño de Moveek. | 65 |
| 4.10. | Relación entre orquestación y coreografía de Servicios Web | 66 |
| 4.11. | Relación de los servicios Web. | 67 |
| 4.12. | Diagrama de la operación CreatePublicacion | 67 |

| | |
|--|-----|
| 4.13. Diagrama de la operación Anotar del servicio Web ServiciosAnotacion | 68 |
| 4.14. Diagrama de la operación <i>PublicacionCubreAmbito</i> | 69 |
| 4.15. Diagrama de la operación <i>AutorRealizaPublicacion</i> | 69 |
| 4.16. Diagrama de la operación <i>ObtenerMuro</i> | 70 |
| 4.17. Diagrama de la operación <i>ObtenerPublicacionTermino</i> | 71 |
| 4.18. Colaboración de los tres servicios Web | 72 |
| | |
| 5.1. Pantalla de entrada a Moveek | 75 |
| 5.2. Formulario de registro de Moveek. | 76 |
| 5.3. Pantalla principal del perfil de Moveek | 77 |
| 5.4. Proceso de publicación en Moveek. | 78 |
| 5.5. Proceso de la consulta semántica. | 79 |
| 5.6. Ejemplo de la petición por el método GET de la clase consultada. | 80 |
| 5.7. Ejemplo de la petición SOAP Request del servicio Web de la consulta semántica. | 80 |
| 5.8. Ontología cargada en el editor Protégé. | 81 |
| 5.9. Consulta semántica en Protégé. | 81 |
| 5.10. Navegación por una instancia devuelta por la consulta semántica. | 82 |
| 5.11. Resultado de la consulta semántica a la clase <i>Química</i> de la ontología. | 82 |
| 5.12. Resultado de la consulta semántica a la clase <i>Ciencia.política</i> de la ontología. | 83 |
| | |
| C.1. Prototipo de Interfaz para el ingreso a la red social. | 101 |
| C.2. Prototipo de Interfaz para el formulario de registro a Moveek. | 102 |
| C.3. Prototipo de Interfaz de la pantalla principal del usuario. | 102 |
| C.4. Prototipo de Interfaz para la acción de realizar publicación. | 103 |
| C.5. Prototipo de Interfaz para la acción de realizar publicación de tipo liga. | 103 |
| C.6. Prototipo de Interfaz para la acción de realizar publicación de tipo documental. | 103 |
| C.7. Prototipo de Interfaz de la visita a un usuario Amigo. | 104 |
| C.8. Prototipo de Interfaz de la visita a un usuario desconocido. | 104 |
| C.9. Modelo de navegación para autenticarse en la red social. | 105 |
| C.10. Modelo de navegación para registrar un usuario. | 105 |
| C.11. Modelo de navegación para realizar una publicación. | 106 |
| C.12. Modelo de navegación para invitar a un usuario a ser tu amigo. | 106 |
| | |
| D.1. Pantalla principal de GATE. | 108 |
| D.2. Asociación de la ontología al nomenclator. | 108 |
| D.3. Asociación de la ontología al nomenclátor. | 109 |
| D.4. Pipeline completo para la extracción de la información. | 109 |
| | |
| E.1. Contrato de servicio del servicio Web operacionesProtege | 114 |
| E.2. Contrato de servicio del servicio Web servicioAnotacion | 115 |

Índice de cuadros

| | |
|---|-----|
| 2.1. Clasificación planteada por el epistemólogo alemán Rudolf Carnap (1955). | 26 |
| 2.2. Caracterización de las ciencias según el esquema de Bunge | 27 |
| 2.3. Metadatos de Dublin Core. | 30 |
| 2.4. Descripción de WikiOnt | 30 |
| 2.5. Clases y subclases de Ontology of SCIENCE. | 30 |
| 2.6. Análisis comparativo entre WikiOnt y Ontology of SCIENCE | 31 |
| 2.7. Lista de sitios donde se extraerán los términos para realizar la anotación | 35 |
| 3.1. Términos del tesoro de OntoScience | 48 |
| A.1. Glosario de términos | 97 |
| B.1. Términos referentes a las redes sociales | 99 |
| B.2. Términos científicos de la ontología. | 100 |

Agradecimientos

En primer lugar debo agradecer a mi familia porque siempre han sido los que me han dado fuerzas y ánimos para seguir adelante y conseguir mis objetivos. A mis padres que son mi ejemplo a seguir en la vida y a mis hermanos que desde pequeño me cuidaron, me apoyaron y me enseñaron mucho de lo que ahora sé. A ellos les debo todo.

A mis profesores que me ayudaron a explotar mis habilidades y me otorgaron la visión y los conocimientos necesarios para desarrollar este trabajo. Al Dr. Abraham Sánchez López por su guía y su apoyo para la elaboración de esta tesis. A ellos toda mi admiración y respeto.

A mis amigos y compañeros que me acompañaron en esta aventura universitaria y que juntos hemos aprendido a vivir. Sería poco práctico mencionarlos a todos en estas líneas, pero de corazón les agradezco su compañía y su apoyo.

A la Vicerrectoría de Investigación y Estudios de Posgrado por el apoyo económico que me brindó para la producción de este trabajo.

Capítulo 1

Estado del Arte

1.1. Redes Sociales

Las redes sociales, son actualmente una de las aplicaciones más utilizadas de la Web y se han consolidado poco a poco entre la comunidad de usuarios del Internet en los últimos años. Muchos negocios, organizaciones, comunidades y familias usan las redes sociales para promoverse, comunicarse e interactuar con su audiencia.

Las redes sociales dependen de que los usuarios construyan su propia red de contactos en el sitio Web. Esto los lleva a nuevos contactos y, a estar interrelacionándose en varias redes sociales, les permite ser encontrados por otros usuarios de manera más sencilla. También, les permite a nuevos contactos ser recomendados o introducidos a la red, ayudando a extender la red de usuarios.

En la Figura 1.1 se muestra un ejemplo de cómo se puede construir una red de contactos entre varios usuarios:

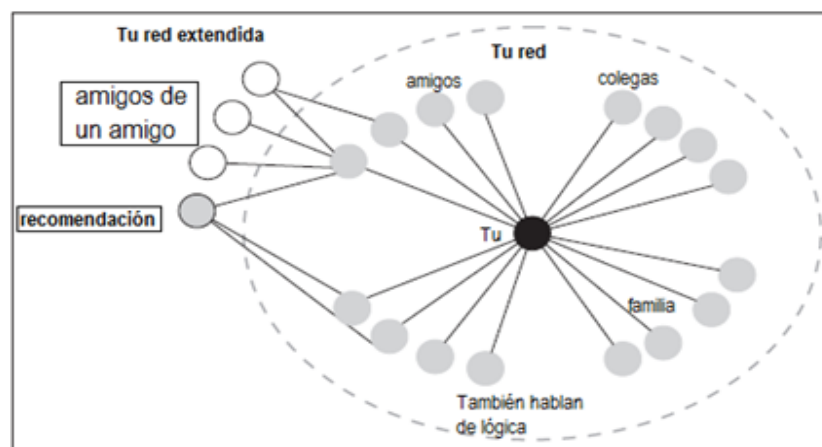


Figura 1.1: Representación gráfica de una red social.

Esta representación de una red social muestra las conexiones entre los contactos. También ilustra cómo un usuario puede ser capaz de descubrir amigos o la recomendación de un ami-

go (basado en amigos en común). Esto hace más fácil a los usuarios, construir su red social, comunicarse con nuevas personas o reconectarse con contactos perdidos.

Los servidores de redes sociales generalmente tienen dos funciones primarias. La función principal, los servidores permiten conectarse con cualquier otro usuario y construir una red de contactos, como ya se ha mencionado. Al proporcionar una comunidad con características de colaboración y contribución, así esto permite que el contenido y la información en la red social sea incrementada por los mismos usuarios [MIC10].

Aunque Tim Berners-Lee imaginó una Web de lectura/escritura (el primer navegador también funcionó como editor de HTML), la Web era un medio de sólo lectura para la mayoría de los usuarios. La Web de la década de los 90 era muy similar a la combinación de una libreta de teléfonos y a la sección amarilla (una mezcla de anuncios individuales, corporativos, catálogos) y esto pese al poder de conectividad de los hipervínculos que inspiró un poco de sentido entre la comunidad de usuarios. Esta actitud pasiva frente a la Web fue interrumpida por una serie de cambios en el uso de patrones y tecnología que ahora se conoce como Web 2.0, una palabra de moda acuñada por Tim O'Reilly.

A continuación, resumimos la historia y los aspectos más descriptivos de la Web 2.0. Los cambios que condujeron al actual nivel de compromiso social en línea no han sido radicales o individualmente significativos, lo que explica por qué el término Web 2.0 ha sido creado en gran medida después de los hechos. Sin embargo, este conjunto de innovaciones en la arquitectura y patrones de uso de la Web condujeron a un rol totalmente diferente del mundo virtual en línea y ahora es una plataforma para la comunicación y la interacción social. El incremento resultante en nuestra capacidad de obtener información en línea, así como el apoyo comunitario puede ser cuantificado. Una importante encuesta efectuada recientemente basada en entrevistas realizadas a 2,200 adultos, indica que el Internet mejora significativamente la capacidad de los estadounidenses para mantener sus redes sociales a pesar de temores iniciales acerca de los efectos de la disminución de contactos en la vida real. La encuesta confirma que no sólo las redes se mantienen y se extienden en línea, sino que también se activan exitosamente para hacer frente a situaciones de vida más importantes, como la obtención de apoyo en caso de una enfermedad grave, buscar trabajo, informarse acerca de las inversiones más importantes, etc. [BHWR06]

La primera ola de la socialización en la Web se debió a la aparición de los blogs, wikis y otras formas de la Web basada en la comunicación y la colaboración. Los blogs y los wikis atrajeron gran cantidad de usuarios en el 2003 (ver la Figura 1.2). Lo que tienen en común ambos tipos de aplicaciones, es que ambos disminuyen significativamente los requisitos para agregar contenido en la Web: la edición de blogs y wikis no requiere ningún conocimiento previo de HTML. Los blogs y los wikis permiten a individuos y grupos solicitar su espacio personal en la Web y llenar contenidos con relativa facilidad. Aún más importante, a pesar de que los Weblogs han sido reconocidos como la primer plataforma de publicación de carácter puramente personal (similar a los diarios), hoy en día la blogosfera es ampliamente reconocida como una red social interconectada densamente a través de la cual las noticias, ideas e influencias viajan rápidamente a medida que los bloggers referencian y reflexionan sobre sus propias publicaciones.

La fracción de páginas Web con el término blogs, wiki es medido contra el tiempo en el eje vertical izquierdo. La fracción de términos folksonomy, XmlHttpRequest y mashup se miden en el eje vertical de la derecha.

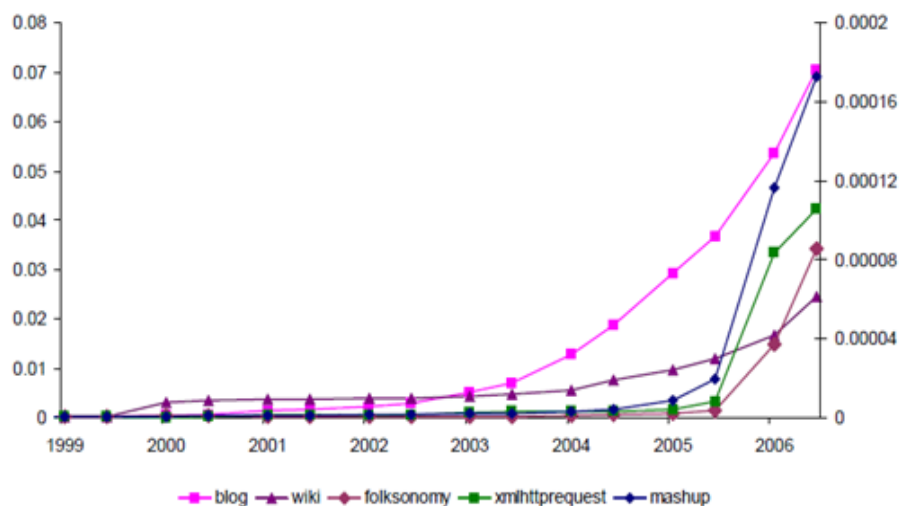


Figura 1.2: Desarrollo de la Web social.

Aunque el ejemplo de Wikipedia, la enciclopedia en línea es excepcional; los pequeños y grandes wikis se utilizan por grupos de diversas magnitudes, como una herramienta efectiva de gestión de conocimiento para el almacenamiento de registros, al describir las mejores prácticas o de forma conjunta el desarrollo de ideas. Independientemente del objetivo, la propiedad colectiva de un Wiki impone un sentido a la comunidad (o en el peor caso revela una falta de cohesión en los grupos sociales) a través de los debates necesarios sobre los contenidos compartidos. De manera similar, el significado de la mensajería instantánea, no es solo comunicación instantánea (el teléfono es instantáneo, y el correo electrónico es casi instantáneo), pero la capacidad de ver quién está conectado, una transparencia que induce a un sentido de responsabilidad social.

Las primeras redes sociales en línea (también conocidas como servicios de redes sociales) hicieron su entrada en el campo al mismo tiempo, como los blogs y wikis comenzaron a despegar. En el año 2003, la primera red social: Friendster, atrajo a más de cinco millones de usuarios registrados en el lapso de unos pocos meses [LEA03], que fue seguida por Google y Microsoft quienes anunciaron servicios similares. Aunque estos sitios ofrecen mucho del mismo contenido en las Web personales, ofrecen un punto de acceso central y llevan a establecer el proceso de intercambio de información personal y de socialización en línea. Después del registro, estos sitios permiten a los usuarios crear un perfil con información básica, invitar a otros a registrarse y ligar los perfiles de sus amigos. El sistema también permite visualizar y navegar por la red resultante con el fin de descubrir amigos en común, amigos que se habían olvidado o nuevas amistades potenciales basadas en intereses compartidos. (Los sitios temáticos atienden objetivos más específicos, tales como el establecimiento de un contacto comercial o la búsqueda de una relación romántica)

Estos sistemas populares permiten a los usuarios mantener redes de grandes cantidades de personal y contactos comerciales. Los miembros se han encontrado rápidamente, sin embargo, dado que la red es sólo un medio para la inmensidad del ciberespacio. Los servicios más recientes usan perfiles de usuario y a las redes para simular diferentes tipos de intercambios: las fotos se comparten en Flickr, los registros se intercambian en del.icio.us, los planes y objetivos unen a los

miembros en 43Things. La idea de la red basada en el intercambio se basa en la observación sociológica que crea similarmente la interacción social y viceversa, la interacción crea similarmente amigos, que son probablemente los que han adquirido o desarrollado intereses similares.

Muchos de estos sistemas se basan en la anotación de colaboración (Folksonomías) para conectar usuarios a los contenidos relevantes y otros a usuarios interesados en cosas similares. Al igual que los wikis, la nueva generación de sitios Web tiene el objetivo de brindar un papel activo en su comunidad de usuarios en la creación y gestión de contenidos, incluyendo la tarea de organizar la misma.

Los perfiles explícitos de los usuarios hacen posible a estos sistemas introducir un mecanismo de evaluación, es decir los usuarios o sus contribuciones se clasifican de acuerdo a la utilidad o la fidelidad.

Las evaluaciones son formas explícitas del capital social que regulan los intercambios en las comunidades en línea, en la misma forma que los moderadores de reputación conocida moderan los intercambios en el mundo real. Por ejemplo, en la plataforma de noticias Digg, los usuarios escogen *un pulgar para arriba* o *un pulgar hacia abajo* a las noticias proporcionadas por otros usuarios. Además, Digg proporciona una forma para que los usuarios vean lo que otros usuarios están *extrayendo* en tiempo real, atendiendo a la creación de un sentido similar de presencia social simultánea como en el caso de la mensajería instantánea.

El diseño e implementación de aplicaciones Web también se ha desarrollado con el fin de que la experiencia del usuario pueda interactuar con la Web, lo más cómodamente posible. En términos de diseño, los nuevos sitios Web ponen énfasis en que la apariencia limpia, accesible y atractiva interfiera lo menos posible con la funcionalidad de la aplicación. En términos de implementación, los nuevos sitios Web se basan en nuevas formas de usar algunas de las tecnologías pre-existentes. (Asynchronous JavaScript y XML, o AJAX, que se usa ampliamente en los últimos sitios Web, no es más que una mezcla de tecnologías que han sido soportadas por los navegadores desde hace varios años).

Siguiendo con la facilidad de uso, lo que también puede ser observado es una preferencia por los formatos, lenguajes y protocolos que son fáciles de utilizar y con los que es fácil desarrollar, en particular lenguajes de script, formatos, tales como JSON, protocolos como REST. Esto es para usar el paradigma del desarrollo ágil y el prototipado (Flickr, por ejemplo, es conocido por adaptar la interfaz de usuario varias veces al día). Además, con la adopción de la ideología del software libre, las aplicaciones Web 2.0 ofrecen su información y sus servicios para la investigación por parte de los usuarios: Google, Yahoo y un sinnúmero de sitios Web más pequeños exponen las principales características de sus sistemas a través de APIs ligeras, mientras que los proveedores de contenidos hacen lo mismo con la información en forma de canales RSS. Los resultados de la investigación de los usuarios con combinaciones de tecnologías son los llamados mashups, sitios Web basados en combinaciones de datos y servicios proporcionados por terceros. El mejor ejemplo de este desarrollo son los mashups basados en el servicio cartográfico de Google como HousingMaps.¹

¹Ver <http://housingmaps.org>

1.2. Servicios Web

La plataforma tipo red social que se propone en este trabajo de tesis se basa en una arquitectura orientada a servicios. Para ello, se han implementado las operaciones de interacción social mediante la invocación de Servicios Web. A continuación presentamos una breve descripción de los Servicios Web aunque se presentaran con más detalle en el Capítulo 4.

1.2.1. ¿Qué es un servicio Web?

Para saber que es un servicio Web, necesitamos primero definir el término Servicios Web. Definiremos de manera separada los términos Servicio y Web, para después presentar las definiciones más comunes utilizadas para identificar qué es un servicio Web.

Un servicio es un recurso abstracto que representa las posibilidades de realizar tareas que aseguran una funcionalidad coherente desde el punto de vista de las entidades proveedor y solicitante. Para ser utilizado, un servicio se debe realizar por un agente proveedor concreto. La noción de servicio existe desde la aparición de los servicios Web: de hecho se utiliza desde hace algunos años por la DCE (Distributed Computing Environment) de la OSF, por CORBA de la OMG, por Java RMI de SUN y por DCOM de Microsoft. Sin embargo la noción de servicio se ha vuelto de vital importancia con la aparición de las Arquitecturas Orientadas a Servicios (SOA).

La Web o World Wide Web es uno de los servicios más importantes del Internet. La Web se compone de servidores Web conectados al internet que alojan páginas Web. Estas páginas son documentos multimedia que contienen texto, imágenes, animaciones y videos. El internet es la red más grande del mundo, que se compone de más de 100 millones de computadoras en más de 100 países y realiza esfuerzos importantes para promover el comercio, actividades académicas y gubernamentales. Con el paso del tiempo, a pesar de las diferencias técnicas, los términos Web e Internet son casi sinónimos. Es válido utilizar de manera indistinta los términos *Internet* y para referirnos a uno u otro. Las tecnologías que soportan la Web son HTTP, HTML, y más recientemente XML.

El término servicios Web se utiliza frecuentemente pero no siempre con el mismo significado. Un servicio Web es más frecuentemente visto como una aplicación accesible a otras aplicaciones en la Web. Esta es una definición muy abierta, ya que permite considerar toda cosa que utiliza una URL (Uniform Resource Locator) como un servicio Web, por ejemplo los scripts CGI (Common Gateway Interface). La definición también se puede referir a un programa accesible en la Web con una API estable publicado con información adicional sobre un servicio de anuario, como CORBA y el Trading Service. Una definición más específica la proporciono el consorcio UDDI, que definió un servicio Web Como una aplicación de negocios modular que forma un todo, con interfaces basadas en estándares y orientada a Internet, pero sobre todo abierta.

Esta definición es más precisa, ya que enfatiza sobre los estándares de Internet y las interfaces abiertas; lo que permite las invocaciones de los servicios sobre Internet. Sin embargo, esta definición no es suficientemente precisa ya que no proporciona el significado de una aplicación de negocios modular y que forma un todo.

Algunas otras definiciones más precisas de servicios Web fueron propuestas por el W3C [W3C02]. El W3C define un servicio Web como una aplicación software identificada por una URI, cuyas interfaces y vínculos están definidos, descritos y descubiertos como objetos XML.

Un servicio Web soporta las interacciones directas con otros agentes de software utilizando el paso de mensajes basado en XML y utilizando los protocolos de Internet. Esta definición es más precisa que las anteriores, y evidencia la manera de identificar un servicio (utilizando las URI), los elementos de base para contemplar la posible interoperabilidad (utilizando una interfaz y los vínculos basados en XML), las interacciones a través de los protocolos basados en XML.

Sin embargo, esta definición evidencia únicamente el estándar XML, pero ¿Cualquier aplicación orientada a Internet que utiliza las tecnologías basadas en XML es también un servicio Web? [W3C04], el W3C define un servicio Web como un servicio concebido para soportar la interacción interoperable de máquina a máquina en Internet. Este posee una interfaz descrita en un formato explotable por la máquina, es decir descrita en WSDL. Otros sistemas interactúan con el servicio Web de una manera predefinida por su descripción utilizando los mensajes SOAP, típicamente utilizando HTTP con una serialización XML al mismo tiempo con otros estándares de la Web.

En esta última definición, la W3C pone en evidencia las tecnologías y sus roles para construir un servicio Web. Esta definición no menciona el descubrimiento de los servicios Web. A priori el descubrimiento se puede realizar de diferentes maneras, por ejemplo, utilizando el motor de búsqueda google para encontrar un proveedor de servicios Web y así los servicios buscados. Sin embargo, estos son insuficientes y no proporciona una forma normalizada de búsqueda de servicios. En este caso, UDDI es una respuesta más adaptada para registrar las informaciones de los proveedores de servicios y sus servicios [ACKM04], Los servicios Web proporcionan una manera de exponer las funcionalidades de un sistema de información y de hacerlo disponible utilizando las tecnologías estándares de la Web. La utilización de tecnologías estándares reduce los problemas ligados a la heterogeneidad, y es la llave para facilitar la integración de las aplicaciones. Además, los servicios Web proporcionan naturalmente, el soporte necesario para nuevos paradigmas de cómputo y para nuevas arquitecturas, entre las cuales podemos mencionar: SOC (Service Oriented Computing) y SOA (Services Oriented Architecture). Esta definición menciona solamente los estándares de la Web pero no los detalla. Una definición más precisa de los servicios Web la proporciona el diccionario Webopedia (<http://www.Webopedia.com>). Define un servicio Web como una manera estandarizada de integración de las aplicaciones basadas en la Web, utilizando los estándares abiertos XML, SOAP, WSDL, UDDI y los protocolos de transporte del internet. XML se utiliza para representar los datos, SOAP para transportar los datos, WSDL para describir los servicios disponibles, y UDDI para listar los proveedores de servicios y los servicios disponibles. En nuestro trabajo, adoptamos esta última definición para identificar un servicio Web, ya que no enfatiza sobre las tecnologías y su interdependencia. Estas tecnologías estándares son interdependientes ya que se han concebido de manera independiente. La Figura 1.3 presenta estas tecnologías de los servicios Web evidenciando su interdependencia.

La Figura 1.4 presenta la utilización de los servicios Web o los actores proveedores y los solicitantes.

De acuerdo a la Figura 1.4, los actores que solicitan servicios, los proveedores de servicios y el anuario mantienen una relación por medio de las tecnologías de los servicios Web:

1. El proveedor de servicios publica sus servicios Web en el anuario utilizando UDDI.
2. El solicitante busca un servicio Web con las características X, Y y Z.

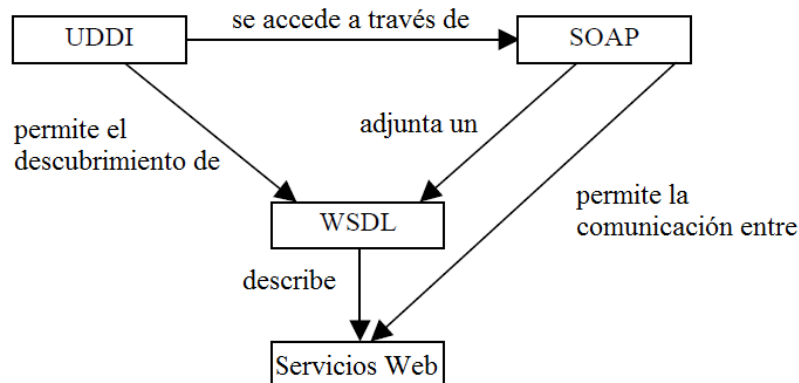


Figura 1.3: Las principales tecnologías de los servicios Web y sus relaciones.

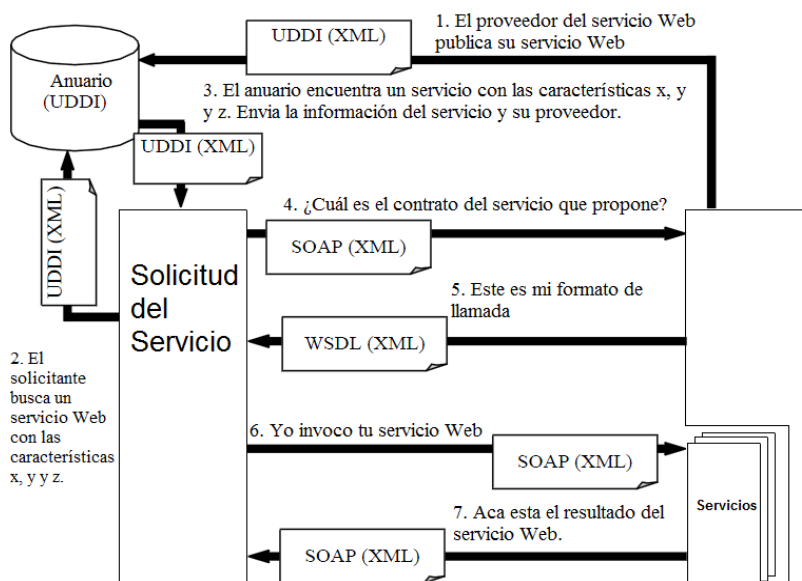


Figura 1.4: Las tecnologías de los servicios Web y los principales actores.

3. El anuario encuentra un servicio con las características X, Y y Z, y envía las informaciones sobre el proveedor de servicio y sobre el servicio.
4. El solicitante del servicio solicita el contrato de servicio Web al proveedor.
5. El proveedor envía el contrato de servicio (WSDL).
6. El solicitante del servicio invoca el servicio Web según su contrato (Utilizando SOAP).
7. El servicio Web regresa el resultado de la llamada (SOAP).

Los estándares SOAP, WSDL y UDDI constituyen la primera generación de los servicios Web. Otros estándares más recientes, como WS-Coordination, WS-Security, WS-Atomic Transaction y BPEL4WS (Business Process Execution Language for Web Services) constituyen la segunda generación de servicios Web.

1.3. ¿La Web 2.0 + Web Semántica = Web 3.0?

La Web 2.0 es a menudo contrastada con la Web Semántica, que es más “*humana*” y como sabemos es un esfuerzo cuidadosamente orquestado por la W3C para desencadenar una nueva era de la evolución en el uso de las tecnologías semánticas. En la práctica, las ideas de la Web 2.0 y la Web semántica no son alternativas excluyentes: mientras que la Web 2.0 considera fuertemente cómo los usuarios deben interactuar con la Web, la Web Semántica abre nuevas oportunidades tecnológicas para los desarrolladores de la Web ya que promueve la combinación de datos y servicios de diferentes fuentes. A continuación, señalaremos algunas de las oportunidades que surgen mediante la combinación de las ideas de estos dos desarrollos.

En primer lugar, notemos que la idea básica de la Web 2.0 es que los usuarios están dispuestos a proporcionar contenido, así como los metadatos. Esto quiere decir que los usuarios pueden hacerse cargo directamente del formato de los artículos, tal y como se organizan en tablas y categorías en Wikipedia, las fotos organizadas en grupos y de acuerdo a las etiquetas de Flickr o la información estructurada incrustada en páginas Web y publicaciones en blogs que utilizan micro-formatos; también llamados mini-vocabularios ya que codifican metadatos de todo tipo en las páginas HTML, por ejemplo, información sobre el autor o una entrada del blog ²).

Aunque esta observación puede parecer trivial, aborda una preocupación primordial para la comunidad de la Web Semántica, es decir, si los usuarios estarían dispuestos a proporcionar los metadatos para dar origen a la Web Semántica. Lo interesante desde el punto de vista histórico es que la Web Semántica fue creada originalmente para ser alimentada por los usuarios de recursos Web, al describir sus páginas personales y el contenido multimedia³ en ellas. Sin embargo, las primeras implementaciones de incorporación de RDF en HTML (como el proyecto ZAPATA [HH01]) han sido abandonados, ya que el diseño no fue realista al esperar que los usuarios habituales, que apenas puede codificar correctamente en HTML, pudieran dominar las complejidades de la codificación de metadatos en RDF, o incluso para codificar correctamente con XML.

Aunque todavía es dudoso que los usuarios dominen cotidianamente los lenguajes de la Web Semántica como RDF y OWL, a partir de las lecciones de la Web 2.0 parece claro que muchos de ellos están, de hecho, dispuestos a proporcionar información estructurada, siempre que puedan hacerlo en tareas de manera orientada y a través de interfaces de usuario amigables que oculten la complejidad de la representación subyacente. Los microformats, por ejemplo, demostraron ser más populares debido a su creación más sencilla usando atributos existentes de HTML. Además, las páginas Web creadas automáticamente a partir de una base de datos (como un blog o páginas de perfiles personales) pueden codificar metadatos en microformats, sin que el usuario este necesariamente consciente de ello. Al mismo tiempo, los microformats retienen todas las ventajas de RDF en términos de accesibilidad de la máquina. Por ejemplo, los motores de búsqueda de un blog son capaces de proporcionar la búsqueda sobre las propiedades del autor o de la noticia.

Conscientes de ello, la idea de ofrecer formas de codificar RDF en páginas HTML han reaparecido (ver microformats o RDF / A [AB06]). En el mismo contexto, también hay trabajos en

²Ver <http://www.microformats.org>

³Adopción del nombre por RDF, el marco de descripción de recursos (Resource Description Framework por sus siglas en inglés).

el desarrollo sobre la forma de extender MediaWiki, el software que está detrás de Wikipedia, para permitir a los usuarios codificar datos en el texto de los artículos al escribir el texto. Este marcado adicional computable de los hechos, permite fácilmente extraer, consultar y agregar el conocimiento de Wikipedia [VKV⁺06]. También hay trabajos similares sobre los nuevos sistemas Wiki que combinan la creación libre de texto con la edición colaborativa de información estructurada [ADR06].

En segundo lugar, debido a muchas colaboraciones en línea, es posible acceder a bastantes aplicaciones que tienen significativamente más metadatos sobre los usuarios. La información acerca de las elecciones, preferencias, gustos y redes sociales de usuarios significa que la nueva generación de aplicaciones es capaz de construir perfiles de usuario mucho más interesantes. Claramente, la tecnología semántica puede ayudar en la adaptación de los usuarios con intereses similares, así como la coincidencia con contenido disponible. La semántica, sin embargo, no es suficiente: como lo demuestra Golbeck, hay componentes de confianza que están más allá de lo que se puede inferir basándose solamente en la similitud del perfil [GH06]. Por lo tanto, los sistemas socio-semánticos que pueden ofrecer recomendaciones basadas tanto en la red social de usuarios como en sus perfiles personales es probable que superen a los sistemas tradicionales de recomendación, así como a los mecanismos puramente basados en la red de confianza.

En lo que respecta a lo que la tecnología de la Web Semántica puede ofrecer a la comunidad de la Web 2.0 es una infraestructura estándar para las combinaciones creativas de construcción de los datos y servicios. Los formatos estándar para el intercambio de datos y el esquema de la información, ayudan en la integración de datos, junto con los lenguajes de consulta estándar y los protocolos para realizar consultas a fuentes de datos remotas que proporcionan una plataforma para el fácil desarrollo de aplicaciones Web híbridas. San Francisco basado en MetaWeb está desarrollando en FreeBase⁴, una especie de "datos en común" que permite a los usuarios compartir, relacionar, y editar conjuntamente ontologías y datos estructurados a través de una interfaz basada en la Web. Una API pública que permite a los desarrolladores construir aplicaciones utilizando el conocimiento combinado de FreeBase.

De acuerdo a [MIK07] el término Web 3.0 se entiende como la siguiente evolución de la Web que involucra, justamente los conceptos abordados en esta sección: la Web 2.0 y la Web semántica. Mika sugiere en [MIK07] que el uso de estas dos tecnologías en su conjunto forma parte de la Web 3.0, sin embargo no se debe entender que el uso de la Web 2.0 o de la Web semántica por separado se trata de la Web 3.0.

1.4. Web Semántica

Tim Berners-Lee, el creador de la Web, ha asegurado que la Web semántica es la próxima evolución de la Web. Es decir, que vamos a llegar a una Web inteligente donde la información este almacenada de manera comprensible por las computadoras con el fin de aportar al usuario lo que busca verdaderamente. Actualmente, los humanos son los únicos que tienen la capacidad de comprender lo que encontramos y de decidir en que se parece lo que nosotros deseamos buscar verdaderamente. ¿Por cuales medios? Son los motores de búsqueda quienes nos ayudan.

⁴Ver <http://www.freebase.com>

Pero estos buscadores sólo son capaces de responder solamente a dos preguntas: ¿Cuáles son las páginas que contienen un término? Y ¿Cuáles son las páginas más populares del término?

Es decir, la idea de la Web semántica, no es hacer que las computadoras comprendan el lenguaje humano o que funcionen en lenguaje natural. Esto no es una Web de reflexión con una inteligencia artificial, sino más bien, agrupar la información de manera útil. Y, como en la Web actual, que está construida principalmente en base al identificador URI, el protocolo HTTP y el lenguaje HTML, la Web semántica se basa también en el URI, HTTP y el lenguaje RDF.

Por otro lado, encontramos que las capacidades de automatización de la Web actual están limitadas ya que el internet se concibió para publicar documentos no estructurados. La Web no nos facilita el acceso a la información. Por ejemplo: si se desea buscar un fabricante de puertas y ventanas para construir una casa, escribiendo las palabras "gates" y "windows" en google, se tienen los resultados no satisfactorios ya que la mayor parte se referiría a M. Bill Gates y Microsoft Windows. ¿Por qué? Como todo el mundo lo sabe, HTML es un lenguaje no estructurado que no puede distinguir la presentación en la información. Es por esto que el motor de búsqueda no tiene la capacidad de comprender este problema.

Por lo tanto, uno de los objetivos de la Web semántica es afinar la búsqueda sobre internet. Para hacerlo, se debe agregar a la información existente una capa de metadatos para que las computadoras puedan explotarla.

La Web semántica se apoya en tres etapas complementarias. De inicio, se agregan los metadatos a cada recurso Web. Después, se certifica su autenticidad. Finalmente, se corrigen los errores propios de HTML.

Se esperaría que los recursos de XML y RDF (Resource Description Framework) debieran corregir este problema. Con XML, se puede estructurar un documento o un recurso para indicar que efectivamente, es la puerta de un fabricante o se trata de Bill Gates de Microsoft. En este caso de encontrar una puerta, se puede excluir de los resultados en los recursos que se refieren a una persona ya que se indica por lo tanto: `fabricante:object:gates`. Si verificamos el árbol XML de los documentos, en el caso de los índices, encontramos las ramas que contienen "gates". ¿Esto es posible?, no, porque la construcción de un índice como este tomaría mucho tiempo. Solución: se va agregar a cada página HTML un archivo para describir su contenido, respetando una estructura estándar. La W3C propone por lo tanto, enriquecer la información existente (y la futura) por medio de metadatos RDF.

En primer lugar, se puede decir que el formato RDF permite definir metadatos para precisar las características de una información, los RDF son tripletas que van a asociar los metadatos definidos por grupos de tres. Se puede describir una triplete como tres URIs. Por lo tanto se puede componer estos conceptos de base gracias a las tripletas combinadas por tres. ¿Cuál es la diferencia con la Web actual? La Web actual utiliza los vínculos en forma de pare, por ejemplo la asociación "Señor Tim Berners-Lee" y "W3C". Entonces, la Web semántica busca tipear la información agregando el tercer término "fundador".

`<Tim Berners-Lee><fundador><W3C>`

Se puede ver que hay una relación entre Señor Tim Berners-Lee y la organización W3C. En este caso, la computadora puede perfectamente decidir qué debe hacer lógicamente. Actualmente existen lenguajes experimentales de lógica como RDFS [PLC09], DAML+OIL [PLC09], y OWL

[PLC09]. Para describir estos lenguajes se necesita una ontología. En resumen, la ontología es la descripción precisa de los términos y las relaciones de un tema preciso, más adelante explicaremos el concepto de ontología más ampliamente). Por lo tanto, parecería que las computadoras comprenden la información trabajando gracias a las ontologías precisas y completas. El objetivo de la Web semántica es proporcionar la información en un sentido que las computadoras puedan comprender.

Además, es común formular otra pregunta: ¿Cómo se puede construir un sistema de comunicación de máquina a máquina para compartir la información? La respuesta es el servicio Web semántico que combina las URI, HTTP y RDF. Gracias a esto, estaríamos en la capacidad de automatizar completamente el proceso, de almacenar un seguimiento preciso de la transacción y de disponer de una abstracción de esta.

1.5. Conocimiento

Para que la Web semántica funcione, las máquinas deben tener acceso a colecciones estructuradas de información y a conjuntos de reglas de inferencia [LHL01]. A menudo las máquinas utilizan esta forma de representación para conseguir un razonamiento automatizado. Es lo que conocemos como la representación del conocimiento. El reto de la Web semántica es proporcionar un lenguaje que exprese a la vez los datos y las reglas. Enseguida se debe razonar sobre los datos. Después las reglas de cualquier sistema de representación del conocimiento pueden exportarse a la Web semántica.

Es decir, se agrega lógica a la Web para darle la posibilidad de utilizar las reglas y efectuar las correspondientes inferencias, escoger el tipo de solución y responder a las preguntas del usuario. ¿Cómo se puede hacer esto?

Existen dos importantes tecnologías de desarrollo para la Web semántica: XML y RDF. Brevemente, XML nos permite agregar una estructura arbitraria a nuestros documentos sin preocuparse del significado de las estructuras. Por otro lado, RDF se utiliza por la Web semántica y permite a las máquinas comprender los documentos y los datos semánticos.

Se puede decir que el sentido está dado por RDF. Si se utiliza RDF en documento partimos del principio de que las cosas particulares tienen propiedades con ciertos valores. Por ejemplo: **El señor Tim Berners-Lee es el fundador W3C**, o **Soy alumno de la FCC**, etc. Esta estructura es una manera para describir la mayor parte de los datos tratados por las máquinas.

¿Y cómo se puede describir mejor la información que se publica y por ende interpretar la información que se recibe? En un contexto compartido, es necesario interpretar la información [BER03]. Por lo tanto, se necesita también de modelos formales para la representación del conocimiento.

En la Figura 1.5, se supone que se desea acceder a varias bases de datos que nos proporcionan información sobre personas, con sus respectivas direcciones. Se desea encontrar igualmente, el nombre de las personas con un código postal específico. Así pues, se tiene la necesidad de conocer los campos en cada base de datos que representan los nombres y los códigos postales. Por ejemplo: RDF puede especificar que “(el campo 5 en la base de datos) (es un campo de tipo) (código postal)”.

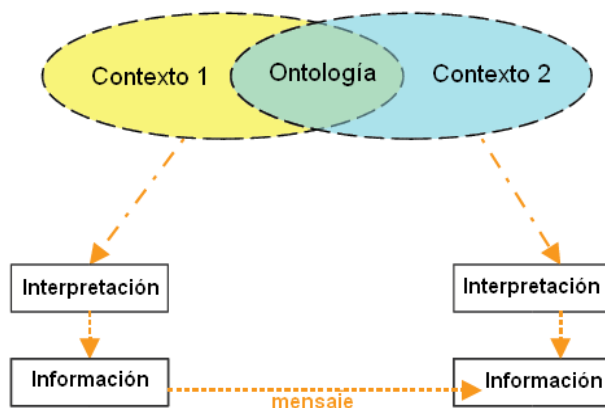


Figura 1.5: Los contextos compartidos-ontologías.

Pero en este caso, se supone que hay dos bases de datos que utilizan identificadores diferentes para el concepto “*código postal*” para dos contextos distintos. Es decir, el programa debe saber que estos dos términos tienen la misma información cuando se desea compararlos y combinarlos. Es difícil descubrir los sentidos idénticos en las bases de datos. En este caso, un conjunto de información denominado ontología nos podría ayudar. Brevemente, una ontología es una descripción formal de conocimiento y se van a utilizar los lenguajes de ontologías para definir los conocimientos, es decir, definir los vocabularios compartidos bajo la forma de conceptos y de relaciones para la interpretación de la información.

1.6. Ontología para la Web semántica

Como ya se había planteado anteriormente, uno de los principales objetivos de la Web semántica es el compartimiento y la reutilización de información para los usuarios y las máquinas.

Por lo tanto, se debe agregar el conocimiento a la información para mejorar la comprensión para la máquina. Para hacerlo, simplemente, se anotan los recursos con el significado de los conceptos. Se tiene a menudo necesidad de los mismos conceptos para representar la misma cosa en la indexación de diferentes recursos. En este caso, necesitamos de las ontologías.

La ontología ha jugado un rol muy importante para la Web semántica ya que va a representar la semántica de los documentos y por ende su explotación de las aplicaciones y los agentes inteligentes. Igualmente es muy útil para estructurar y definir el significado de los términos de los metadatos [PHA05].

Así pues, con la ayuda de las ontologías, las aplicaciones Web del futuro podrán ser inteligentes, en el sentido que estas podrán funcionar a nivel conceptual humano.

1.6.1. ¿Qué es una ontología?

De la enciclopedia Wikipedia [WF23], en computación, una ontología es un conjunto estructurado de conceptos. Los conceptos están organizados en un grafo cuyas relaciones pueden ser:

- Relaciones semánticas;

- Relaciones de composición y de herencia (en el sentido de los objetos).

La estructuración de los conceptos en una ontología permite definir los términos, unos en relación a los otros; cada término es la representación textual de un concepto. Por ejemplo, para describir los conceptos que componen el diseño de las tarjetas electrónicas, se podría definir la siguiente ontología (simplificada):

- Una *tarjeta* electrónica es un conjunto de componentes.
- Un *componente* puede ser ya sea un *condensador*, una *resistencia* o un *chip*.
- Un *chip* puede ser una *unidad de memoria* o una *unidad de cálculo*.
- Una *tarjeta electrónica* que contiene una unidad de cálculo contiene también al menos una *unidad de memoria*.

Simplemente, se puede construir una ontología a partir de un corpus de textos. Se va a recorrer el texto para buscar términos recurrentes o definidos por el usuario, después se analiza la forma de cómo estos términos están relacionados en el texto (por la gramática, y por los conceptos que abarcan, y además de una definición que puede encontrarse en un léxico proporcionado por el usuario). El resultado es una ontología que representa el conocimiento global que contiene el corpus del texto en el dominio de aplicación que este abarca.

Encontramos que el término “*ontología*” tiene varias definiciones una de las más simples y populares es la definición de M. Tom Gruber [GRU23]. El autor define que “*una ontología es una especificación formal explícita de una conceptualización compartida*”. Los atributos “*formal*” y “*explícito*” significan que una ontología permite una interpretación automatizada de la conceptualización efectuada por la máquina. De otra forma, una ontología define un vocabulario común para los investigadores que necesitan compartir la información en un dominio. Las ontologías pueden utilizarse por personas, bases de datos y aplicaciones que necesitan compartir información en un dominio particular. Estas incluyen definiciones, información explotable, conceptos elementales en el dominio y sus relaciones. También codifican un conocimiento en un dominio y también un conocimiento que se puede extender a otros dominios. De esta manera, las ontologías permiten la reutilización de los conocimientos.

A continuación se detallan algunas razones para desarrollar una ontología:

- *Compartir la comprensión común de la estructura de la información entre las personas o los fabricantes de software.* Por ejemplo: se cuenta con un cierto número de sitios Web que contienen la información médica o proporciona servicios de e-commerce en medicina. Si estos sitios comparten y publican la misma ontología, que es la base de los términos que ellos utilizan, entonces los agentes pueden extraer y agregar información de estos diferentes sitios. Los agentes pueden utilizar esta información agregada para poder responder a las preguntas de los usuarios o como datos de entrada para otras aplicaciones.
- *Permitir la reutilización del conocimiento en un dominio.*
- *Detallar lo que está considerado como implícito en un dominio.*
- *Distinguir el conocimiento en un dominio del conocimiento operacional.*
- *Analizar el conocimiento en un dominio.* La Web semántica necesita ontologías de un cierto grado de estructura significativa. Es por lo tanto que se van a definir las descripciones para los siguientes conceptos

- Las clases (de cosas generales) en los diversos dominios de interés;
- Las relaciones que puedan existir entre las cosas;
- Las propiedades (o los atributos) vinculados a estas cosas.

■

Dicho de otra forma, una ontología es un modelo de organización de los conocimientos en un cierto dominio. Encontramos en la ontología clases de objetos que se deben organizar (personas, estudiante, profesor, tesis, ...) y los tipos de relaciones entre los objetos (un objeto "estudiante" puede vincularse por una relación "supervisado por" a un objeto de tipo "profesor", etc).

1.6.2. Ontología para la Web

Para la Web, simplemente, una ontología es un documento o archivo que define de manera formal las relaciones entre los términos [LHL01]. Este tipo de ontología poseerá una taxonomía y un conjunto de reglas de inferencias. La taxonomía define las clases de objetos y las relaciones entre ellos. Las clases, las subclases y las relaciones entre las entidades son una herramienta muy poderosa para el usuario en la Web. Se pueden expresar esas relaciones asignándole propiedades a las clases y permitiendo a las subclases heredar las propiedades.

Por ejemplo: Una *dirección* puede estar definida como un tipo de *lugar*, y los *códigos postales* de la ciudad pueden estar definidos para aplicarse solamente a los *lugares*, etc. Las ciudades tienen sitios Web y los *códigos postales* deben tener un tipo de *ciudad*. Por lo tanto, se puede asociar un sitio Web al *código postal* de una *ciudad* (aquí, no hay una base de datos que vincule directamente este código a un sitio Web). Las reglas de inferencia son por lo tanto muy expresivas una ontología puede expresar la siguiente regla "Si un *código postal* de la *ciudad* está asociado a un *código de estado* y que una *dirección* utiliza este código de la ciudad, **entonces** esta *dirección* está asociada al *código del estado*". Las computadoras pueden manipular los términos de manera más eficaz y significativa.

1.6.3. Ontología y servicios Web

Hasta ahora, la ontología es una de las aplicaciones más importantes para los agentes en la Web y para los servicios Web. Recientemente muchas empresas dedicadas al e-commerce B2B han discutido el rol de las ontologías para controlar las interacciones de máquina a máquina. Sin embargo, se debe suponer que las ontologías solo se utilizan en las aplicaciones para asegurarse que son convenientes a los términos utilizados, a los tipos y a las restricciones. En la Web semántica, se pueden crear ontologías comprensibles por una máquina y utilizables por los agentes [GAR04]. Por lo tanto se pueden utilizar los servicios Web y automatizar su uso.

Actualmente, existe un problema para encontrar los diferentes servicios Web disponibles en la Web. Por ejemplo: ¿Puedes enviar una tarjeta de felicitación para alguien? Pero no sabes exactamente el nombre de una de las compañías que ofrecen este tipo de servicios. Como lo vimos anteriormente, el resultado de esta búsqueda no es correcto ya que la mayor parte de los motores de búsqueda utilizan las palabras claves que no devuelven al proveedor más popular de este servicio. Si se hubiera proporcionado otra palabra clave, por ejemplo "e-card", se tendría

un mejor resultado. Cada vez más, el número de servicios que se desarrollan crece de acuerdo a nuestras necesidades, por lo tanto, la capacidad de los motores de búsqueda comunes para encontrar los servicios es más apropiada y extendida al límite. Existen, un buen número de esfuerzos que pueden invocar una “*publicidad de servicio*” que nos puede ayudar a mejorar esta situación, por ejemplo UDDI, ebXML, eSpeak [HEN00]. Se ha creado un vocabulario de anuncios de servicios que ayudan a los motores de búsquedas a tener la capacidad de encontrar estos servicios. Con este ejemplo, esta página Web de tarjetas de felicitación puede registrar cualquier cosa como “*un servicio personalizado, felicitaciones, comunicaciones, etc.*” y el usuario podrá más fácilmente obtenerla sin necesidad de conocer el término “*e-card*”. El primer uso de las ontologías para la búsqueda de servicios en la Web, es crear los anuncios de servicios en un lenguaje de ontologías. ¿Cómo hacerlo? Las herramientas de búsqueda utilizan la jerarquía y las restricciones de propiedad para encontrar el mayor número de correspondencias gracias a las propiedades de las clases, de las subclases o de otros vínculos semánticos, es decir otra ontología apropiada. Además, se va a utilizar una combinación de lenguajes ontológicos para mejorar los anuncios de servicios. Se puede incluir una descripción de un servicio comprensible por una máquina de acuerdo a la manera que este funciona y una lógica explícita que describa las consecuencias de uso de este servicio.

1.7. La anotación semántica

El objetivo de esta sección del trabajo de tesis es elaborar la descripción de la anotación semántica apoyándose en una visión amplia de este dominio.

1.7.1. Algunas definiciones

El diccionario Larousse define el término **anotación** como una *nota crítica o explicativa que acompaña a un texto –una nota de lectura que se añade a los libros*. Por otro lado, el diccionario de la Real Academia de la lengua Española, precisa que el término *anotación* es un derivado del término latin *annotare*, que significa “*notar; anotar*”. La anotación corresponde por lo tanto a “*la acción de anotar, en el resultado de esta acción*”. El término *anotar* por lo tanto se define como “*acompañar un texto de notas, de comentarios, de observaciones*”.

Así como el término *anotación* se refiere a una nota, una crítica, una explicación o a un comentario. Es decir, escribimos una nota sobre un tema o bien lo criticamos, explicamos, comentamos un tema. Una anotación por sí misma no tiene sentido, está siempre asociada al objeto que ha sido anotado. Es por esto que las anotaciones se consideran como metadatos. Como lo señala Handschum [HAN05], si un metadato es un dato sobre un dato, una anotación constituye un caso particular de un metadato ya que este representa un nuevo dato vinculado a un recurso documental. Prié & Garlatti distinguen un metadato como una descripción normalizada referenciada a un recurso identificado (notablemente para el caso de la Web) y una anotación como un comentario libre situado en el interior del recurso documentado [GP04]. Es importante en este caso precisar el significado de recurso documental: Este puede corresponder al conjunto de un documento o bien solamente a un fragmento de este y contener texto, imagen, sonido, video o una combinación de estos contenidos. En el estudio de la Web, este será identificado y accesible mediante una dirección URL [LAU07].

La anotación de recursos documentales es una vieja tradición en el mundo de la documentación y de las bibliotecas. La DLF (Digital Library Federation), una asociación constituida por las quince bibliotecas americanas más importantes en los Estados Unidos, definió tres clases de anotaciones que pueden aplicarse a los recursos documentales de una biblioteca numérica [HAN05]:

- La anotación administrativa, o anotación documental [LAU07], indica la información asociada a la creación y al mantenimiento del recurso documental como *"que, cual, o cómo"*. Desde la aparición de la Web, el lenguaje DublinCore es el estándar para la anotación con los descriptores como autor, título, fuente, editor, fecha de publicación, idioma, etc.
- La anotación estructural conecta las partes de los recursos documentales entre ellos con el fin de constituir una representación lógica de un documento [RDH⁺03].
- La anotación descriptiva describe un recurso documental en relación a su contenido, es decir que esta va a separar los conceptos mencionados en el recurso documental, las relaciones entre estos conceptos así como sus instancias [LAU07] [RDH⁺03].

Nos interesamos principalmente por la **anotación descriptiva**. El contenido de un documento puede ser analizado según diferentes ángulos y cada uno de estos puede ser útil para un objetivo bien preciso de la aplicación que utiliza las anotaciones descriptivas resultantes de estos análisis. Euzenat [EUZ05] distingue particularmente tres estructuras que pueden constituir los diferentes enfoques para la anotación descriptiva: *"La estructura gramatical para analizar las relaciones entre sintagmas, la estructura retórica para dejar en claro la argumentación de un texto o también la estructura lógica el sentido de un documento"*.

En el estudio de la Web semántica, una anotación descriptiva, particularmente las que interesan a la estructura lógica del contenido de un documento, se denominan frecuentemente anotación semántica [GP04]. Como se señala en [LAU07], el término "semántico" es ambiguo pero indica una voluntad de resaltar el sentido de un contenido y este, de manera más o menos formal según los preceptos de la lógica. Las anotaciones semánticas tienen por lo tanto como objetivo expresar la "semántica" del contenido de un recurso con el fin de mejorar su comprensión, su búsqueda y por lo tanto su reutilización por parte de los usuarios finales [COR06]. Por consecuencia definimos la anotación semántica como *una representación formal de contenido, expresada con la ayuda de conceptos, relaciones e instancias descritas en una ontología, y conectadas con el recurso documental fuente*.

1.7.1.1. La automatización en la creación de las anotaciones

Fuchs define un tratamiento automático como aquel que opera por *"medios mecánicos"* mientras que un tratamiento manual o instrumental, es el que opera con la intervención humana [FUC93]. Los medios mecánicos corresponden a las computadoras, y más precisamente a los agentes de software. Existen tratamientos realmente automatizados es decir que la intervención humana sea requerida de forma preliminar para la preparación del tratamiento o solo como parte de la verificación y validación de los resultados producidos.

En lo que se refiere a la creación de las anotaciones semánticas, el tratamiento manual consiste simplemente en utilizar una interfaz de usuario en la cual el usuario humano pueda seleccionar el recurso que se va a anotar, escoger el modelo formal que sirve para la creación de las anotaciones

semánticas y, respetar las restricciones impuestas por el modelo formal, crea las anotaciones deseadas en el recurso seleccionado [RMEK01].

El tratamiento automatizado para la creación de las anotaciones semánticas es de hecho, más semiautomático. Es decir, los tratamientos realmente automatizados utilizan algoritmos basados en los modelos estadísticos o en la explotación de la redundancia en un *corpus* de recursos [SSC04]. Pero su perspicacia está limitada y sus anotaciones son frecuentemente no notables en una aplicación concreta, particularmente en las empresas. Las herramientas de anotación que se basan en este método no son viables más que en proyectos de laboratorio, es decir de forma experimental. Por el contrario, los tratamientos semiautomáticos han probado en el transcurso de los últimos años que pueden aportar una ayuda no despreciable al anotador humano. Estos tratamientos semiautomáticos, se apoyan generalmente en un motor de extracción de información. La construcción de un motor se efectúa según dos métodos: ya sea la creación manual de sus patrones de extracción [FEJ⁺02, MAD⁺03] o mediante aprendizaje supervisado [CSH02, WCD03].

- **Creación manual de patrones:** Este primer método permite el funcionamiento de un motor de extracción de información a partir de un conjunto de reglas definidas por un usuario experimentado, generalmente un lingüista. Estas reglas son, ya sea expresiones regulares (por ejemplo la expresión regular de una fecha bajo la forma "dd/mm/aaaa"), o resultados del análisis lingüístico más detallado (por ejemplo si el análisis sintáctico encuentra un sujeto que representa una entidad nombrada de tipo persona y un verbo que significa el nacimiento y un complemento del lugar que representa una entidad nombrada de tipo lugar, entonces se puede anotar esta frase como el lugar de nacimiento de la persona del sujeto). El conjunto de estas reglas es enseguida compilado en un autómata de estados finitos. El autómata recorre el conjunto del corpus documental y lo anota en función de las reglas que se activan. Las anotaciones son enseguida propuestas al anotador humano usando una interfaz adecuada. Esta misma interfaz los verifica, los valida, los modifica o los suprime. El análisis lingüístico puede ser masomenos fino, pero es también conveniente en la anotación de textos semi o no estructurados. Por el contrario la definición de las reglas se apoya en las expresiones lingüísticas particulares en un dominio dado, este método es fuertemente dependiente de este dominio y frecuentemente difícil de adaptarse a nuevos dominios sin necesidad de redefinir las reglas de extracción.
- **Aprendizaje supervisado de patrones:** Este método permite desarrollar un sistema capaz de aprender a anotar un corpus dado. El anotador humano inicia este aprendizaje, anotando manualmente un subconjunto del corpus. Estas anotaciones sirven entonces de ejemplos al motor de extracciones quien deduce un conjunto de reglas: expresiones simples regulares del lenguaje o expresiones regulares que se apoyan en la estructura misma de los documentos, por ejemplo la utilización de marcas HTML o XML contenidas en el documento analizado. El anotador humano puede también iniciar el aprendizaje al proporcionar un conjunto de reglas base. El motor de extracción es por lo tanto capaz de aprender a anotar recursos al redefinir iterativamente estas expresiones regulares, ya sea por generalización o por especificación, a partir de correcciones proporcionadas a medida por el anotador humano y esto, hasta la obtención de un umbral de éxito aceptable para este último. Una vez alcanzada la fase de aprendizaje, el motor puede entonces automáticamente anotar el resto del corpus documental. Este método es más adaptado a los documentos muy estruc-

turados (provenientes de base de datos) o semi-estructurados (como las páginas Web) en los cuales las construcciones son suficientemente simples y fijas para que las expresiones regulares lingüísticas o estructurales puedan aprenderse fácilmente sin generar muchas anotaciones no pertinentes en el dominio en estudio. El problema con este método es que no solamente se debe proporcionar al motor de aprendizaje un número suficiente de ejemplos de anotaciones, sino que además estos deba ser pertinentes en relación al dominio considerado. También no es siempre fácil constituir o recuperar este tipo de corpus de anotaciones [HAB05].

1.7.1.2. El almacenamiento de las anotaciones y de sus recursos

Las anotaciones pueden ser ya sea “*embarcadas*” o ya sea “*descargadas*” en relación con la fuente de recursos documental [HAB05]. Una anotación se denomina “*embarcada*” cuando esta se agrega directamente al contenido del documento de origen. Este último se podrá actualizar en el sitio Web de origen, en el sistema de gestión de contenidos documentales de las empresas, o simplemente registrados localmente en una computadora para su posterior utilización. Por el contrario, la anotación se denomina “*descargada*” cuando esta se almacena en el exterior del documento fuente. No solamente la anotación misma debe ser almacenada sino que también se debe preservar el vínculo con el recurso anotado. Es lo que permitirá enseguida encontrar todas las anotaciones correspondientes a un recurso. Generalmente, estas anotaciones se almacenan en los servidores de anotaciones que pueden ser consultados con el fin de encontrar las anotaciones de un recurso dado. La ventaja con la anotación descargada es que es posible anotar todo recurso documental, incluidos aquellos en los cuales la aplicación no es propietaria. Esto es particularmente cierto en el estudio de la anotación de páginas Web. El inconveniente reside en el hecho de que si el documento fuente se modifica o suprime, las anotaciones se vuelven obsoletas (también denominadas huérfanas). Las anotaciones pueden ser ya sea “*embarcada*” o ya sea “*descargadas*” en relación con la fuente de recursos documental [HAB05]. Una anotación se denomina “*embarcadas*” cuando esta se agrega directamente al contenido del documento de origen. Este último se podrá actualizar en el sitio Web de origen, en el sistema de gestión de contenidos documentales de las empresas, o simplemente registrados localmente en una computadora para su posterior utilización. Por el contrario, la anotación se denomina “*descargada*” cuando esta se almacena en el exterior del documento fuente. No solamente la anotación misma debe ser almacenada sino que también se debe preservar el vínculo con el recurso anotado. Es lo que permitirá enseguida encontrar todas las anotaciones correspondientes a un recurso. Generalmente, estas anotaciones se almacenan en los servidores de anotaciones que pueden ser consultados con el fin de encontrar las anotaciones de un recurso dado. La ventaja con la anotación descargada es que es posible anotar todo recurso documental, incluidos aquellos en los cuales la aplicación no es propietaria. Esto es particularmente cierto en el estudio de la anotación de páginas Web. El inconveniente reside en el hecho de que si el documento fuente se modifica o se suprime, las anotaciones se vuelven obsoletas (también denominadas huérfanas).

Recordemos que nos centramos en las anotaciones semánticas y que no suponemos que estas sean estructuradas y descritas formalmente para representar el conocimiento bajo la forma de instancias de conceptos y de relaciones de la ontología de referencia. Estas instancias pueden ser almacenadas en unas bases de conocimientos, independientemente del recurso anotado, con

el fin de ser reutilizadas para nuevas anotaciones de otros recursos o simplemente para ser consultadas por los usuarios de esta base de conocimientos. En este preciso caso, hablamos del detallado de la ontología o del enriquecimiento de la base de conocimiento. Estas nociones describen la acción de agregar nuevas instancias a una base de conocimientos restringida por una ontología, dicho de otra manera detallar una ontología con instancias de conceptos y de relaciones [GP04]. Las herramientas de anotación semántica pueden ser capaces de detallar una ontología existente a partir de anotaciones semánticas creadas para un recurso dado, al menos que la ontología no se haya ya detallado con todas las instancias posibles del dominio considerado. Es también interesante notar que ciertas herramientas presentadas anteriormente no tiene como primer objetivo la tarea de anotación. Estas consideran más bien a la anotación, como un medio de capturar el conocimiento de un dominio para detallar una ontología.

Comentario: Es importante distinguir la noción del detallado de la ontología y del enriquecimiento de la ontología. En este último caso, se trata no solo de agregar nuevas instancias a conceptos existentes, sino más bien a agregar nuevos conceptos o relaciones al modelo formal de la ontología. Al menos es necesario prever la manera de cómo se almacena la ontología con el fin de que las herramientas de anotación puedan consultar y explotar su modelo formal, es decir razonar a partir de éste.

1.7.1.3. La utilización de las anotaciones

La anotación semántica se utiliza en muchas aplicaciones como la búsqueda de información semántica, la categorización, la composición de documentos, etc. [GP04]. La anotación semántica se aplica a cualquier tipo de contenido: páginas Web, documentos textuales no estructurados, campos de una base de datos, documentos de audio o video, etc. Finalmente, mientras más se formalice el modelo de anotación, más los servicios propuestos a partir de esta anotación se vuelven inteligentes. Es decir, los agentes de software podrán inferir del nuevo conocimiento, razonar sobre este conocimiento y así mejorar los resultados de la búsqueda de información o bien despejar un sentido implícito contenido en el documento de origen [LAU07].

Acabamos de definir que es la anotación y más particularmente la anotación semántica de recursos textuales, a partir de las dimensiones propuestas inicialmente por Marshall, y posteriormente por Prié y Garlatti. Como lo explicamos, estas anotaciones semánticas se deben expresar de manera formal. Vamos por lo tanto a estudiar cuáles son los lenguajes y los recursos disponibles que permiten crear anotaciones semánticas, más particularmente en el estudio de un proyecto de Web semántica.

1.7.2. Los lenguajes de la anotación semántica

Se han definido varios lenguajes de representación del conocimiento, y de ontologías, por los investigadores en inteligencia artificial. Citamos notablemente OntoLingua, LOOM, OCML y F-Logic. Estos lenguajes se han concebido a partir de diferentes modelos resultantes de las teorías de la lógica como las lógicas de descripción, los predicados de primer y segundo orden, los frames, etc. Pero estos no pueden utilizarse como tal en la Web semántica sin una cierta adaptación, notablemente sintáctica [BCEH04]. Además de HTML, se han creado nuevos lenguajes para la Web como XML y RDF. Después, con la perspectiva naciente de la Web semántica,

estos dos conjuntos se han fusionado para dar paso a los nuevos lenguajes para la especificación de ontologías orientadas a la Web, como RDF Schema, DAML+OIL y OWL. Pero es evidente que la Web semántica no puede existir si una estandarización de sus lenguajes, tal y como fue el caso para HTML en la Web actual. Es por esto que el W3C ha participado activamente en la elaboración de estos nuevos meta-lenguajes para la Web mediante la publicación de las recomendaciones alrededor de XML, RDF(S) y OWL. Otros organismos, como la DARPA (Defense Advanced Research Projects Agency) o la ISO (International Standard Organization) han igualmente patrocinado proyectos relevantes alrededor de estos meta-lenguajes, respectivamente DAML y XML Topic Maps. Vamos a estudiar algunos de estos lenguajes en la perspectiva de la anotación semántica.

1.7.2.1. La pirámide de los lenguajes de la Web semántica

La arquitectura de la Web semántica se compone de un conjunto de lenguajes, generalmente representados bajo la forma de una pirámide. Cada nivel se apoya en los resultados definidos en el nivel inferior, es decir que cada nivel es progresivamente más especializado y más complejo que su nivel precedente. Por otro lado, todo nivel es independiente de los niveles superiores con el fin de pueda desarrollarse y hacerse operacional de manera autónoma en relación a los desarrollos de los niveles superiores. Esta pirámide de los lenguajes, ver Figura 1.6, se presentó originalmente por Tim Berners-Lee [PRI00] durante la conferencia XML.

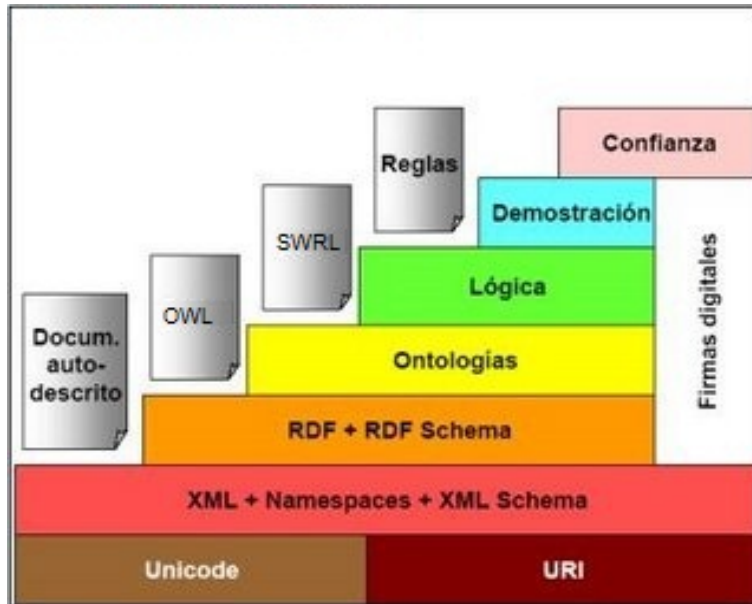


Figura 1.6: Pirámide de los lenguajes de la Web semántica en 2005.

En la base, los recursos se codifican en el formato Unicode y utilizan el sistema de direccionamiento URI (Universal Resource Indicator) que sirve igualmente de identificador en la Web. Estos recursos deben estar conforme al formato XML que utilizará un espacio de nombres particular y que será validado en relación a una gramática definida en un DTD o en un XML Schema. Después, tenemos que RDF permite describir los meta-datos de los documentos bajo la

forma de tripletas como se especifica en RDF Schema. Permite implementar un primer nivel de ontologías, relativamente simples. Con el fin de representar ontologías más complejas, el lenguaje OWL que define clases, atributos, relaciones y axiomas puede combinarse en un lenguaje de reglas que permite razonar sobre los recursos y de inferir del nuevo conocimiento. El nivel lógico (Logic) se utiliza para establecer la coherencia de las anotaciones y poder inferir conclusiones no explícitamente enunciadas. El nivel Prueba (Proof) podría proporcionar los medios para trazar y explicitar las diferentes etapas del razonamiento lógico con el fin de poder acordarle un nivel de confianza. El último nivel (Trust) tiene como objetivo autentificar la identidad y la veracidad de los datos y servicios disponibles sobre la Web semántica. [AMA07]

1.8. Aportación

En este trabajo de tesis se propone una red social que se centra en el intercambio de noticias, artículos y contenido científico en general; al mismo tiempo que el contenido de esta red se anota semánticamente para que, mediante servicios Web, se realicen consultas semánticas al contenido publicado en la red por los usuarios.

En este capítulo se expuso brevemente la forma en la que las redes sociales incrementan el número de contactos de un usuario mediante la recomendación de amigos o gustos en común entre los usuarios de la misma, sin embargo nuestra propuesta se centra en realizar esas tareas de interacción social otorgando más importancia a la semántica del contenido que publican, es decir que la búsqueda de información, creación de muros de publicaciones, etc. se centre en el contenido explícito que publica el usuario. Esto se realiza mediante el uso de una ontología que concentra una colección de términos de un dominio científico y mediante la propiedades de anotación descubren el conocimiento que se publica en la red social. Para lograr la interacción planeada con el usuario final de la red social, las consultas semánticas y las operaciones con la ontología en general se realizan a través de un conjunto de servicios Web que permiten la operación de la red social.

En el Capítulo 2 se propone la ontología que se usará para las tareas de la Web semántica. En el Capítulo 3 se abordan los aspectos teóricos de la anotación semántica y se hace énfasis en su importancia en la arquitectura de nuestra plataforma. En el Capítulo 4 se presenta el desarrollo de nuestra red social mediante las técnicas de composición de Servicios Web. En el Capítulo 5 se exponen los resultados obtenidos. Finalmente en el Capítulo 6 se presentan las conclusiones y el trabajo futuro para esta plataforma.

Capítulo 2

Ontología Propuesta

En este capítulo se describen los aspectos necesarios para la creación de la ontología que se utilizará para las tareas de Web semántica que nos interesan. Primero abordaremos brevemente el dominio que deseamos modelar, después desarrollaremos la ontología mediante la metodología propuesta por [Alv10].

2.1. Propuesta

Como ya se ha discutido brevemente en el capítulo del estado del arte, la idea de la Web semántica es extender la información no estructurada con la descripción de significados (semántica) que sean procesables por una computadora. El reto principal de la Web semántica es asegurar una interpretación compartida de la información. Las fuentes de la información relacionada deberían usar los mismos conceptos para referirse a las mismas entidades del mundo real o al menos debería haber una manera de determinar si dos fuentes se refieren a las mismas entidades, pero posiblemente usando diferente vocabulario. Las ontologías y el lenguaje de ontologías están creando la tecnología que resuelve estos problemas. Una ontología, es una conceptualización formal de un dominio, la descripción de sus conceptos y sus relaciones [GRU93b, BAT97]. Las ontologías son modelos de dominio con dos especiales características, que le conducen a la noción de significado o semántica compartidos.

1. Las ontologías son expresadas en lenguajes formales con una semántica bien definida.
2. Las ontologías se basan en una comprensión compartida con la *comunidad*.

En este estudio, se ha propuesto el desarrollo de una ontología que modele el **conocimiento científico** que se publica en la red social *Moveek*. A continuación se aborda brevemente el dominio que se modela en nuestra ontología. Según la Real Academia Española, la ciencia se define como el *conjunto de conocimientos sistemáticamente estructurados, y susceptibles de ser articulados unos con otros. La ciencia surge de la obtención del conocimiento mediante la observación de patrones regulares, de razonamientos y experimentación de ámbitos específicos, a partir de los cuales se generan preguntas, se construyen hipótesis, se deducen principios y se elaboran leyes generales y sistemas metódicamente organizados*. Claramente, un dominio es un excelente candidato a ser modelado por una ontología. Comencemos el estudio de nuestro dominio con

una clasificación general de las ciencias. Entre la literatura que se estudió para esta propuesta metodológica se evaluaron dos clasificaciones. La primera clasificación, propuesta por Rudolf Carnap en 1955, separa en tres grandes grupos las ciencias: las ciencias formales, las ciencias naturales y las ciencias sociales. En el cuadro 2.1 se muestran las características de cada grupo, así como algunas ciencias que forman parte del grupo.

| | |
|---------------------------|---|
| Ciencias formales | Estudian las formas válidas de inferencia: lógica - matemática. No tienen contenido concreto; es un contenido formal, en contraposición al resto de las ciencias fácticas o empíricas. |
| Ciencias naturales | Son aquellas disciplinas científicas que tienen por objeto el estudio de la naturaleza: astronomía, biología, física, geología, química, geografía física y otras. |
| Ciencias sociales | Son aquellas disciplinas que se ocupan de los aspectos del ser humano - cultura y sociedad- El método depende de cada disciplina particular: administración, antropología, ciencia política, demografía, economía, derecho, historia, psicología, sociología, geografía humana, trabajo social y otras. |

Cuadro 2.1: Clasificación planteada por el epistemólogo alemán Rudolf Carnap (1955).

Mario Bunge (1972) considera el criterio de clasificación de la ciencia en función del enfoque que se da al conocimiento científico: por un lado, el estudio de los procesos naturales o sociales (el estudio de los hechos) y, por el otro, el estudio de procesos puramente lógicos (el estudio de las formas generales del pensar humano racional), es decir, postuló la existencia de una ciencia factual (o ciencia fáctica) y una ciencia formal.

Las ciencias factuales se encargan de estudiar hechos auxiliándose de la observación y la experimentación. La física, la psicología y la sociología son ciencias factuales porque se refieren a hechos que se supone ocurren en la realidad y, por consiguiente, tienen que apelar al examen de la evidencia científica empírica. [Ste83]

El objeto de estudio de la ciencia formal no son las cosas ni los procesos, sino las relaciones abstractas entre signos, es decir, se estudian sus relaciones sintácticas y sus posibles inferencias. Son ciencias formales la lógica y las matemáticas.

La ciencia experimental se ocupa del estudio del mundo natural. Por mundo natural se ha de entender todo lo que pueda ser supuesto, detectado o medido a partir de la experiencia. En su trabajo de investigación, los científicos se ajustan a un cierto método, un método científico general y un método específico al campo concreto y a los medios de investigación. Las ciencias formales, en cambio, crean su propio objeto de estudio; su método de trabajo es puro juego de la lógica, en cuanto formas del pensar racional humano, en sus variantes: la lógica y las matemáticas.

En el Cuadro 2.2 se resume el trabajo de Bunge para la especificación de las categorías que propone para clasificar la ciencia [LAR02].

A partir de lo anterior, se decidió tomar en cuenta para establecer la taxonomía de nuestra ontología la clasificación propuesta por Mario Bunge. Sin embargo una ontología también debe tomar en cuenta las propiedades no esenciales o no definidas de las entidades de su competencia. En nuestro estudio se modelan las propiedades en base a las necesidades de consulta que se harán a la ontología.

| | Formales | Fácticas |
|------------------------------|--|--|
| Objeto de estudio | <ul style="list-style-type: none"> ▪ Estudian entes formales, ideales o conceptuales ▪ Dichos entes son postulados hipotéticamente (construidos, propuestos, presupuestos o definidos) por los científicos que los estudian. | <ul style="list-style-type: none"> ▪ Estudia el mundo de los hechos (Desde las galaxias a las partículas subatómicas; nubes, elefantes, alegrías y tristezas). ▪ Tales hechos se asumen que tienen existencia con independencia de los científicos y de las comunidades que los estudian, aunque puedan tener interacciones con ellos. |
| Modo de validación | <ul style="list-style-type: none"> ▪ Parten de axiomas o postulados y a partir de ellos demuestran teoremas ▪ Los axiomas son relativos al contexto en el cual se opera. ▪ No requieren de cotejo empírico o de experimentación. ▪ Sus conclusiones adquieren grado de certeza | <ul style="list-style-type: none"> ▪ Trabajan a partir de las consecuencias observacionales que se derivan de las conjeturas o hipótesis propuestas. ▪ Juzgan sobre su adecuación al pedazo de realidad que pretenden describir o explicar. ▪ El resultado favorable es provisional sujeto a corrección y revisión. |
| Objetivo que persigue | <ul style="list-style-type: none"> ▪ Buscan la coherencia interna. ▪ Buscan la verdad lógica y necesaria. | <ul style="list-style-type: none"> ▪ Procuran describir y explicar hechos y realidades ajenas a ellas mismas. ▪ Persiguen la verdad material o contingente. |

Cuadro 2.2: Caracterización de las ciencias según el esquema de Bunge

Por la naturaleza de nuestro dominio, es claro que sería muy costoso y virtualmente imposible modelar completamente el campo de las ciencias, es por ello que hemos decidido modelar de manera general la mayor parte de las ciencias y disciplinas existentes, pero nos concentraremos en modelar las ciencias formales específicamente las Ciencias de la Computación. Con lo anterior pretendemos abarcar los conceptos, que consideramos, más referenciados en las publicaciones realizadas por los usuarios de Moveek. Llamaremos a nuestra ontología *OntoScience*.

2.1.1. Desarrollo de la ontología

En este punto retomamos el trabajo de [Alv10] para desarrollar nuestra ontología, basándonos implícitamente en el trabajo de Noy & McGuinness Methodology. En la Figura 2.1 se presenta el diagrama de las etapas de la metodología a emplear para desarrollar nuestra ontología.

Como se puede observar en la Figura 2.1, esta metodología hace énfasis en la elaboración de un modelo conceptual robusto y en la determinación clara y concisa de los requerimientos de la ontología a construirse.

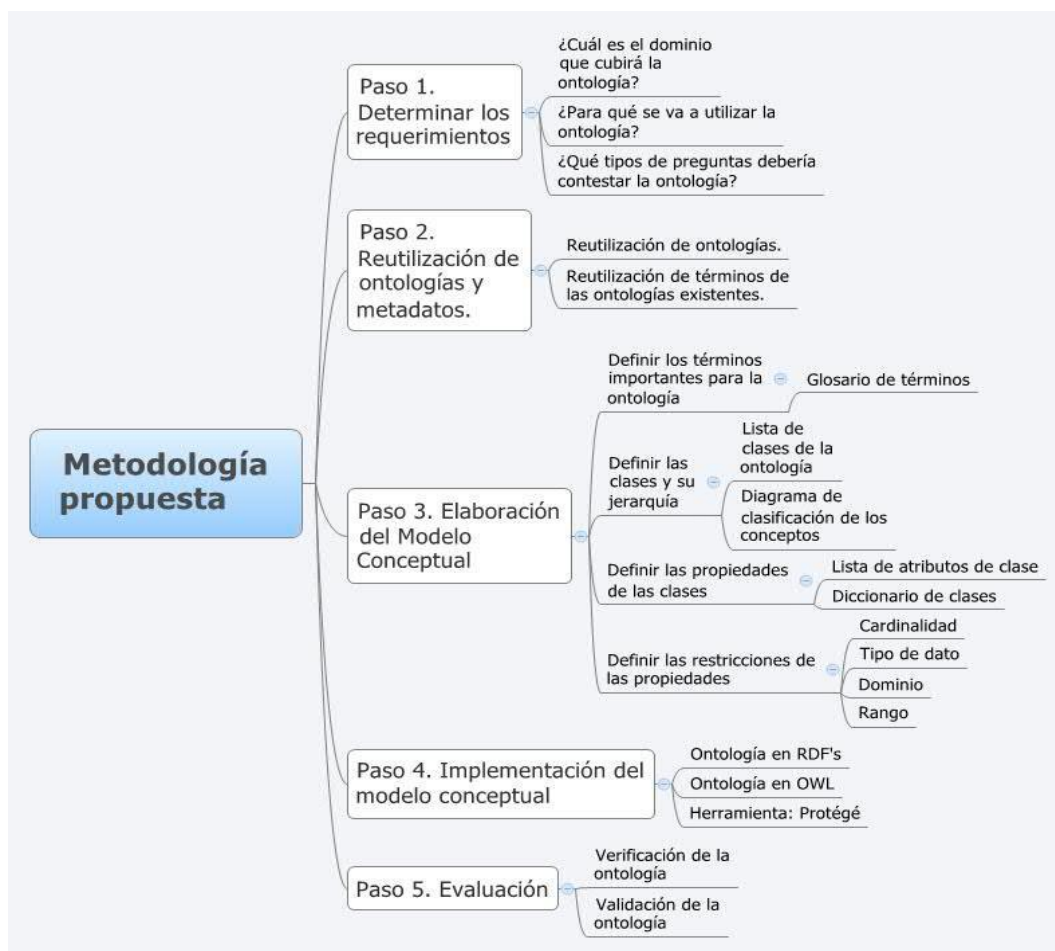


Figura 2.1: Descripción de la metodología de desarrollo de la ontología.

2.1.1.1. Determinar los requerimientos

Como todo componente de software, siempre es recomendable saber qué es lo que se pretende solucionar, es por ello que en este punto se van a detallar los requerimientos que va a cubrir OntoScience mediante unas " *preguntas significativas o de competencia*" [FOX95] , es decir utilizar una lista de preguntas que el sistema debería ser capaz de contestar.

En este apartado se darán respuesta a las siguientes cuestiones básicas:

- **¿Qué dominio cubrirá la ontología?** OntoScience representa la clasificación de la ciencia según Mario Bunge. De manera muy específica se modelará el dominio de las ciencias formales y su relación con los conceptos propios al desarrollo tecnológico y aspectos relativos a una red social.
- **¿Qué preguntas debe contestar la ontología?** En el contexto de este trabajo de investigación, se pretenden hacer tres consultas a la ontología que se resumen en las siguientes preguntas:
 - *¿Qué publicaciones realizadas por algún amigo contienen términos referentes a los que he publicado?*

- ¿Qué usuarios de Moveek realizan publicaciones de temas que yo también menciono en mis publicaciones?
- ¿Qué publicaciones se refieren a un término en específico consultado?
- **¿Quién utilizará y mantendrá la ontología?** Los usuarios de OntoScience serán los estudiantes y/o profesores de la Facultad de Ciencias de la Computación de la BUAP que deseen realizar algún estudio similar para proyectos de investigación de esta naturaleza. Nosotros como autores de esta ontología daremos el mantenimiento y enriqueceremos la ontología según observemos el crecimiento y variedad de los términos que se publiquen en *Moveek*.

2.1.1.2. Reutilización de ontologías

Como se mencionó en la parte de la propuesta de este capítulo, la Web actualmente es un conjunto de información alojada sin estructura ni orden alguno que facilite la consulta de información o conocimiento de esta, parece una biblioteca desordenada y caótica. Los metadatos son datos que describen datos y ayudan a solucionar este problema, permitiendo describir los recursos en la Web y de esta forma ayudando a su consulta.

Para que los metadatos permiten mantener un registro sobre el significado, contexto y propósito de un objeto de información para poder comprenderlo y administrarlo de la mejor manera posible y de esta forma poder cumplir los objetivos de significados compartidos de la Web semántica, mediante el uso de sus sinónimos más usados o conocidos. Existen diversos organismos internacionales encargados de estandarizar la definición de los metadatos. Entre las más importantes se encuentra Dublin Core.

2.1.1.2.1. Metadatos de Dublin Core

Dublin Core es una organización que promueve un estándar de propósito general, sencillo y descriptivo de los recursos de la Web sobre cualquier materia. Promueve la iniciativa Dublin Core Metadata Initiative (DCMI) que provee una infraestructura operativa para la red semántica y el intercambio de la información por medios electrónicos.

Los metadatos universales DC están diseñados específicamente para indexar documentos y recursos. Para ello contiene quince elementos básicos agrupados según el tipo de información: siete elementos de contenido, cuatro elementos con información relativa a la propiedad intelectual del recurso y cuatro elementos relativos a la identificación, formato y temporalidad del documento. [Alv10]. En el Cuadro 2.3 se clasifican los metadatos que Dublin Core establece para los recursos en la Web.

Para nuestro trabajo de investigación utilizaremos algunos de los metadatos que contiene *Dublin Core* como parte de los nombres que asignaremos a las clases de *OntoScience*.

2.1.1.2.2. WikiOnt

Es una ontología desarrollada por los estudiantes de DERI de la Universidad Nacional de Irlanda (National University of Ireland). Fue creada para tener una definición común de la estructura de la información contenida en Wikipedia, lo que la vuelve interpretable por una computadora y de esta forma la computadora pueda procesar y reutilice datos. Los detalles completos de WikiOnt se encuentran en:

| Contenido | Propiedad Intelectual | Instanciación |
|---------------------------|---------------------------|----------------------------|
| Título (tiitle) | Autor (creator) | Fecha (date) |
| Tema (subject) | Editor (publisher) | Formato (format) |
| Descripción (description) | Colaborador (contributor) | Identificador (identifier) |
| Tipo (type) | Derechos (rights) | Idioma (language) |
| Fuente (source) | | |
| Relación (relation) | | |
| Ámbito (coverage) | | |

Cuadro 2.3: Metadatos de Dublin Core.

<http://sw.derri.org/2005/04/wikipedia/wikiont.html>.

En el Cuadro 2.4 se muestran los términos que modela WikiOnt.

| Clases | Propiedades | | |
|----------|--------------|-------------|---------------|
| Article | contentType | Contributor | Sioc views |
| Category | externalLink | DC date | Skos narrower |
| Image | internalLink | DC title | Skos subject |
| Stub | Image height | DC creator | Text |
| | Image width | Link | redirectsTo |

Cuadro 2.4: Descripción de WikiOnt

2.1.1.2.3. Ontology of SCIENCE

Esta ontología fue desarrollada por Fred Freitas y se desarrolló con el propósito de buscar, clasificar las páginas y extraer información de estas páginas que correspondan a las clases de la ontología. Se encuentra disponible en [Uni12].

Esta ontología modela el contenido científico que se publica por la comunidad científica del mundo en internet. Se menciona en (Universidad de Standorf) que las clases que contiene tienen la jerarquía que se muestra en el Cuadro 2.5.

| Dividable Publication | Book | Magazine | Journal |
|-----------------------|------------------|--------------------|--------------|
| | Proccedings | | |
| Part-Publication | Journal-Article | Conference-Article | Book-Chapter |
| Events | Scientific-Event | Education-Event | |

Cuadro 2.5: Clases y subclases de Ontology of SCIENCE.

2.1.1.2.4. Análisis comparativo

Se han presentado dos ontologías que si bien no cumplen completamente las necesidades que se requieren para *OntoScience*, nos dan la pauta para construir nuestra ontología a partir de los criterios con las que esas ontologías se han desarrollado. En el Cuadro 2.6 se muestra un análisis comparativo de las ontologías WikiOnt y Ontology of SCIENCE que nos servirán como base para *OntoScience*.

Una vez hecho el análisis comparativo entre las dos ontologías estudiadas, es momento de concluir este análisis en algunos puntos:

| Característica | WikiOnt | Ontology of SCIENCE |
|----------------------------|--|---|
| Documentación | Abundante | Escasa |
| Nivel de especialización | Bajo. Solo cubre el dominio de artículos en la Wiki. | Bajo. Solo cubre el dominio de trabajos de investigación, no de contenido científico en general, es decir, fuera de estos trabajos. |
| Lenguaje de representación | OWL | OIL, DAML, FLO |
| Aplicaciones | Media Wiki | La arquitectura de un sistema multiagente publicado en: [Uni12] |
| Licencia | Licencia pública general (GNU –GPL) | Licencia pública general (GNU –GPL) |

Cuadro 2.6: Análisis comparativo entre WikiOnt y Ontology of SCIENCE

- Ambas ontologías tienen licencia pública (GNU –GLP) lo cual nos permite utilizar y extender sus elementos.
- WikiOnt nos proporciona mucha documentación para poder realizar las tareas de anotación semántica de los recursos en la wiki, sin embargo excluye la información científica contenida en publicaciones de investigación, pero Ontology of SCIENCE si lo hace.
- Ya que WikiOnt está representada en OWL, podemos reutilizar directamente los elementos que necesitemos de esta ontología, ya que utilizaremos OWL como lenguaje de representación con el fin de facilitar las tareas de interoperabilidad para las tareas de la anotación semántica.
- Ontology of SCIENCE está representada en lenguajes que no utilizaremos para representar nuestra ontología, solo usaremos la taxonomía que propone para extraer información de las publicaciones de Moveek que puedan estar en alguna publicación modelada.

2.1.1.2.5. Elementos a reutilizar

Ya hemos estudiado las ontologías que cubren alguna parte del dominio que modela la ontología OntoScience. Es momento de especificar las clases que reutilizaremos de las ontologías ya estudiadas.

- De *WikiOnt* se tomarán las clases: Article y sus respectivas propiedades y relaciones.
- De *Dublin Core* se tomarán las especificaciones de los metadatos: descripción (description), autor (creator), fecha (date) y ámbito (coverage).

2.1.1.3. Elaboración del modelo conceptual

2.1.1.3.1. Definición de los términos de la ontología

En el apéndice A se listan los términos que tienen relación con el dominio que deseamos modelar. Se muestran los términos recolectados relativos al dominio de una red social de divulgación científica como lo es Moveek. Para ello hay que distinguir dos tipos de términos: los referentes a las ciencias y los referentes a la red social.

Para que la ontología pueda resolver las consultas de la mejor manera, es conveniente representar los dos grupos de términos (red social y ciencias) en la misma ontología.

2.1.1.3.2. Definición de las clases y jerarquías

Ahora que tenemos el glosario de términos podemos extraer las clases y jerarquías que modelaremos en nuestra ontología. Se seleccionarán aquellos conceptos que describen objetos independientes para construir clases. El resto de los términos se considerarán para un tratamiento posterior.

En el caso de las jerarquías de las clases de ciencias, el glosario de términos especifica explícitamente la jerarquía que tiene una ciencia respecto a la otra.

Algunas reglas que sugiere [FOX95] para la formulación de una nueva clase son las siguientes:

- Una nueva subclase de una clase generalmente
 - Debería tener nuevas propiedades que no posee la clase.
 - Debería tener diferentes valores para las propiedades que los de la clase.
 - Debería participar en diferentes relaciones de la clase.
- Las nuevas clases no tienen por qué introducir nuevas propiedades. Si un factor es importante en el dominio y pensamos en los objetos con diferentes valores para ese factor como diferentes clases de objetos, entonces deberíamos crear una nueva clase o clases considerando dicho factor.
- Las instancias son los conceptos más específicos representados en una ontología. Si los conceptos de un dominio forman una jerarquía natural (como es el caso de nuestro dominio), debemos representarlos como clases, aunque sean clases abstractas. Considerando ciertas reglas para la definición de las clases.

A continuación se presenta nuestra propuesta de jerarquización de los términos desglosados en el apéndice A. Esta jerarquización de las clases se realiza de esta forma para que la ontología realice las consultas semánticas necesarias. En la Figura 2.2 se presenta la propuesta.

Ahora que tenemos un panorama más claro de la jerarquía de nuestros términos, podemos formular la lista de clases de la ontología. En nuestro caso, estos términos son considerados conceptos (a partir de los cuales se pueden crear instancias) ya que cada publicación se puede referir a algún término listado en el Cuadro B.2, es decir cada publicación puede ser clasificada en algún término de esta ontología. Además las consultas semánticas para las que utilizaremos la ontología requieren de esta jerarquía para realizarse.

2.1.1.3.3. Definición de las propiedades de la ontología

Ahora que ya hemos definido las clases que formarán parte de nuestra ontología, es momento de definir sus propiedades y las relaciones que existen entre ellas mediante un diagrama de relaciones binarias.

Para nuestro objetivo de realizar anotaciones semánticas al contenido que se publique en *Moveek*, es necesario hacer uso de una poderosa herramienta que nos proporciona el lenguaje de representación que estamos usando: las propiedades de anotación de OWL.

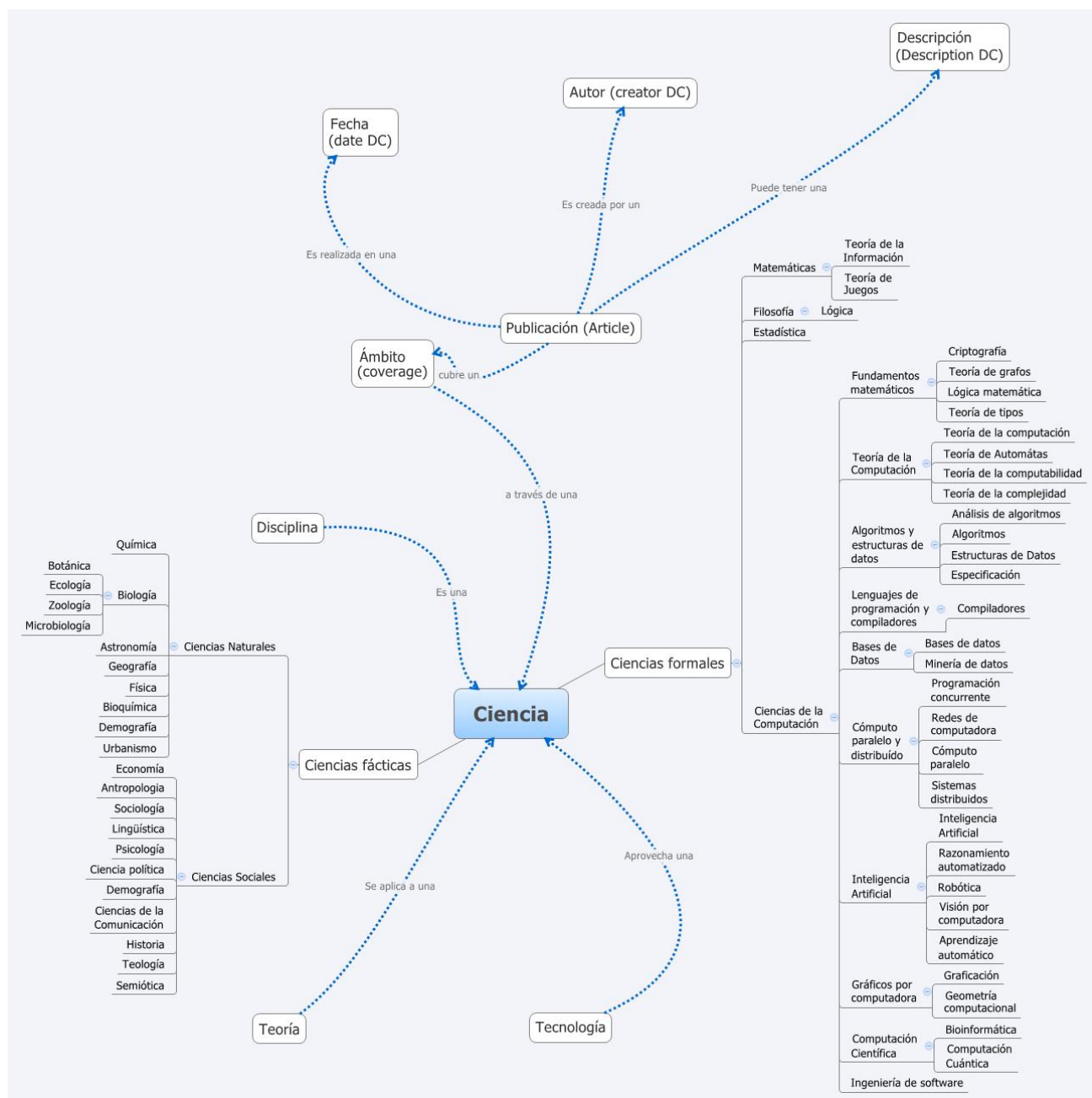


Figura 2.2: Propuesta de jerarquización de los términos del Glosario de términos

OWL Full no establece ninguna restricción o anotación en una ontología. OWL DL permite realizar anotaciones sobre clases, propiedades, instancias y cabeceras de la ontología, pero únicamente bajo las siguientes condiciones:

- El conjunto de propiedades de objeto, propiedades de tipo de datos, propiedades de anotación y propiedades de ontología deben ser mutuamente disjuntos. Así, en OWL DL `dc:creator` no puede ser al mismo tiempo una propiedad de tipo de datos y una propiedad de anotación.

- Las propiedades de anotación deben tener una forma explícita de tripleta como sigue:

```
AnnotationPropertyID rdf:type owl:AnnotationProperty
```

- Las propiedades de anotación no deben ser usadas en axiomas de propiedades. Así, en OWL DL uno no puede definir subpropiedades o restricciones de dominios/rangos para propiedades de anotación.
- El objeto de una propiedad de anotación no debe ser ni una literal, ni una referencia URI, ni una instancia.

Existen 5 propiedades de anotación que están definidas por OWL, se llaman:

- `owl:versionInfo`
- `rdfs:label`
- `rdfs:comment`
- `rdfs:seeAlso`
- `rdfs:isDefinedBy`

A continuación se presenta un ejemplo de un uso correcto de una propiedad de anotación en OWL DL:

```
1 <owl:AnnotationProperty rdf:about="&dc;creator">
2
3 <owl:Class rdf:about="#Ciencia">
4   <rdfs:label>Ecologia</rdfs:label>
5   <dc:creator>Mark Zuckemberg</dc:creator>
6 </owl:Class>
```

Este ejemplo asume que `&dc;` y `dc` apuntan respectivamente al URI de Dublin Core y a su espacio de nombres. Así, usando las propiedades de Dublin Core como propiedades de anotación en OWL DL requiere una tripleta explícita. Esto asegura que las anotaciones son interpretadas con una semántica correcta por los razonadores de OWL DL. [CON09]

Ahora que hemos establecido las propiedades de anotación que nos proporciona OWL DL, definiremos la propiedad que usaremos para nuestros fines. La propiedad de anotación `rdfs:label` nos permite conectar un término en forma de texto (cadena) contenido en un recurso textual cualquiera con un término de la ontología mediante un proceso de *Anotación Semántica*, el cual detallaremos más adelante.

Los términos utilizados para realizar la anotación son tomados de una serie de sitios en internet en los cuales transitan publicaciones con contenido científico. En el Cuadro 2.7 se presenta la lista de los sitios de los que se han tomado la mayor parte de términos que serán los valores de la propiedad `rdfs:label` de las clases de nuestra ontología. Cabe mencionar que la extracción de los términos se hizo de tal forma que la mayor parte de clases tuvieran al menos 5 valores de la propiedad `rdfs:label` y de esta forma la anotación de las publicaciones realizadas en nuestra red social: Moveek este más enriquecida. Estos sitios fueron tomados de [EBU12].

Es el momento de establecer las relaciones (también llamadas propiedades de objeto) entre las clases de *OntoScience*. Estas relaciones se establecen en función de las consultas que se van a realizar a futuro en la ontología y respetando la relación implícita que establece la definición

| Nombre del sitio | Descripción | Autor | País de origen |
|--|---|-----------------------------|----------------|
| Genciencia (Xatakaciencia) | Xatakaciencia es un Weblog colectivo dedicado a la divulgación científica, la ecología y el cambio climático. [Xat12] | Weblogs SL | España |
| SCIENTIA | Scientia es un blog científico en el que se pueden ver las novedades científicas y además donde también hay sitio para la historia de la ciencia, las curiosidades científicas y el humor. [NIC11] | José Manuel López Nicolás | España |
| Francis (th)E mule Science's News | Blog donde se publican artículos y novedades científicas de Física y Matemáticas principalmente. [WOR12a] | WordPress | España |
| Gaussianos | Gaussianos es un blog sobre matemáticas. Publica y aborda temas de interés para matemáticos y afines. [MED12] | Miguel Ángel Morales Medina | España |
| Informática Práctica | Informática Práctica pretende ser un sitio en el que puedas encontrar información interesante y útil sobre cualquier cosa relacionada con la Informática, Internet y la tecnología en general. [WOR12b] | WordPress | España |

Cuadro 2.7: Lista de sitios donde se extraerán los términos para realizar la anotación

de cada término. En la Figura 2.3 se muestra la relación que existe entre las clases que se van a utilizar para las consultas semánticas.

2.1.1.4. Implementación

Ahora que ya tenemos bien definidos los conceptos, las relaciones y las propiedades que se van a representar en OntoScience es momento de implementar la ontología. Para esta tarea se ha elegido el sistema Protégé por su facilidad de uso, portabilidad entre sistemas y por ser un software con licencia GNU. Este sistema nos permite utilizar como lenguaje de representación OWL, como lo habíamos planteado anteriormente.

2.1.1.4.1. Definición de las clases en Protégé

La versión del sistema Protégé utilizada para implementar nuestra ontología es la 4.1.0, compilación 239. Para crear una nueva ontología OWL, se selecciona la opción «Create new OWL Ontology» como se muestra en la Figura 2.4.

En este apartado mencionaremos, brevemente, la forma en la que se agrega una nueva clase a la Ontología. Se selecciona la pestaña de «Classes», se selecciona la clase «Thing» (a partir de la cual cualquier clase que agreguemos se deriva), se presiona el botón «Add subclass» y así comenzamos la definición de nuestras clases. En caso de desear agregar una clase en el mismo nivel que una clase en específico, se selecciona el botón «Add sibling class». En la Figura 2.5 se presenta una imagen que muestra las clases de nuestra Ontología capturadas.

2.1.1.4.2. Definición de las propiedades de anotación

Como se comentó en el apartado anterior, los vocablos que se tomaron en cuenta para establecer

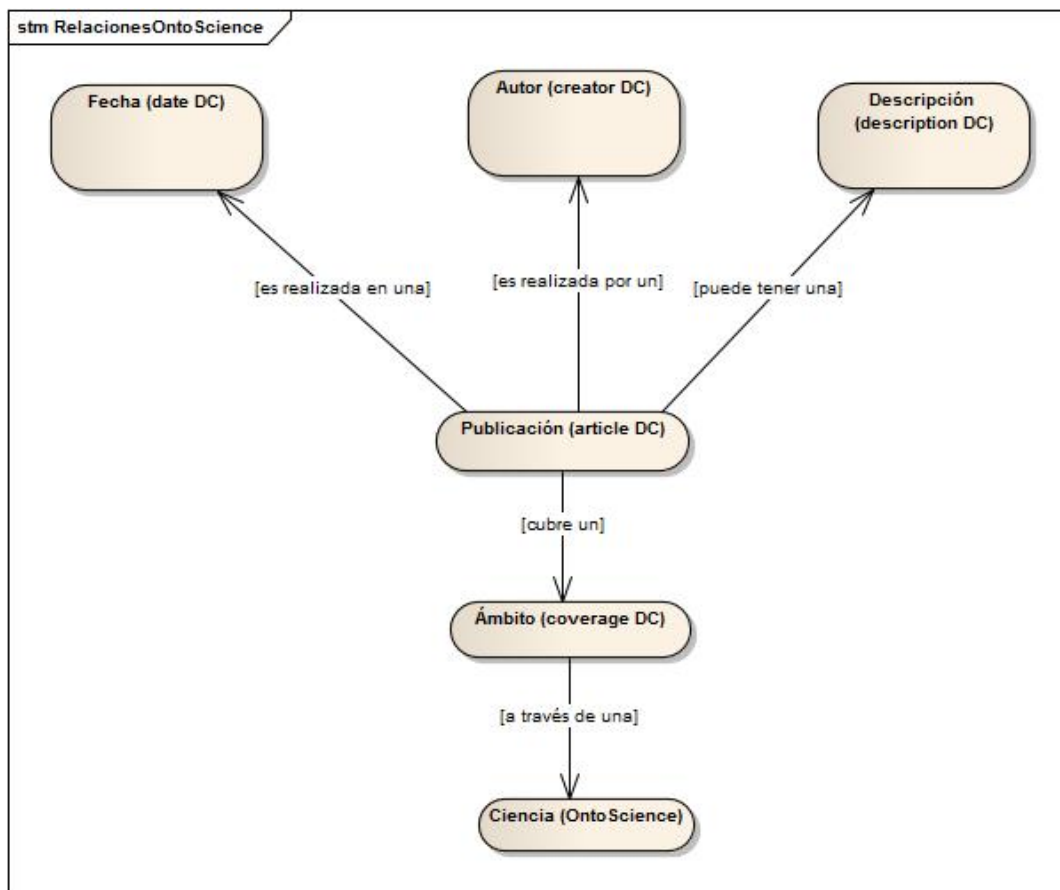


Figura 2.3: Relaciones entre las clases de OntoScience.

la propiedad `rdfs:label` de las clases científicas de nuestra ontología, fueron tomadas de diferentes sitios en la red dedicados a la divulgación científica informal. En la Figura 2.6 se muestra la forma en la que se anotó una clase. El término *tiranosaurio* se extrajo el día 19 de Mayo del 2012 del sitio [Xat12] y se configuro como la propiedad `rdfs:label` de una instancia con la clase *Historia*. De esta forma se ejemplifica la manera en la que se establecerán las propiedades de anotación de cada clase.

2.1.1.4.3. Definición de las relaciones Para establecer las relaciones entre las clases que hemos definido en el apartado anterior, utilizaremos el sistema Protégé. Estas relaciones vincularan las clases referentes a las ciencias jerarquizadas y las entidades referentes a las publicaciones (autor, descripción, fecha, etc.). En la Figura 2.7 se presenta la captura de pantalla de las relaciones (propiedades de objeto) entre clases definidas en el diagrama de relaciones (Ver Figura 2.3).

2.1.1.4.4. Definición de consultas

El objetivo que se persigue en este trabajo de tesis es realizar consultas semánticas sobre información publicada en redes sociales, específicamente en *Moveek*. Para ello haremos diferentes consultas a nuestra ontología utilizando la herramienta Protégé y su razonador *Pellet*. En la Figura 2.8 se muestra la consulta que se realizó con fines de ejemplificar la manera en la que

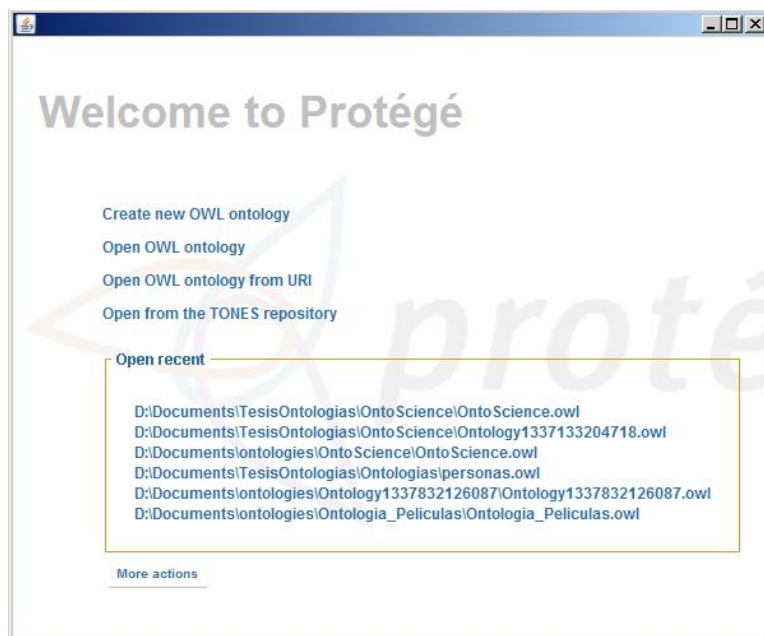


Figura 2.4: Pantalla inicial del sistema Protégé

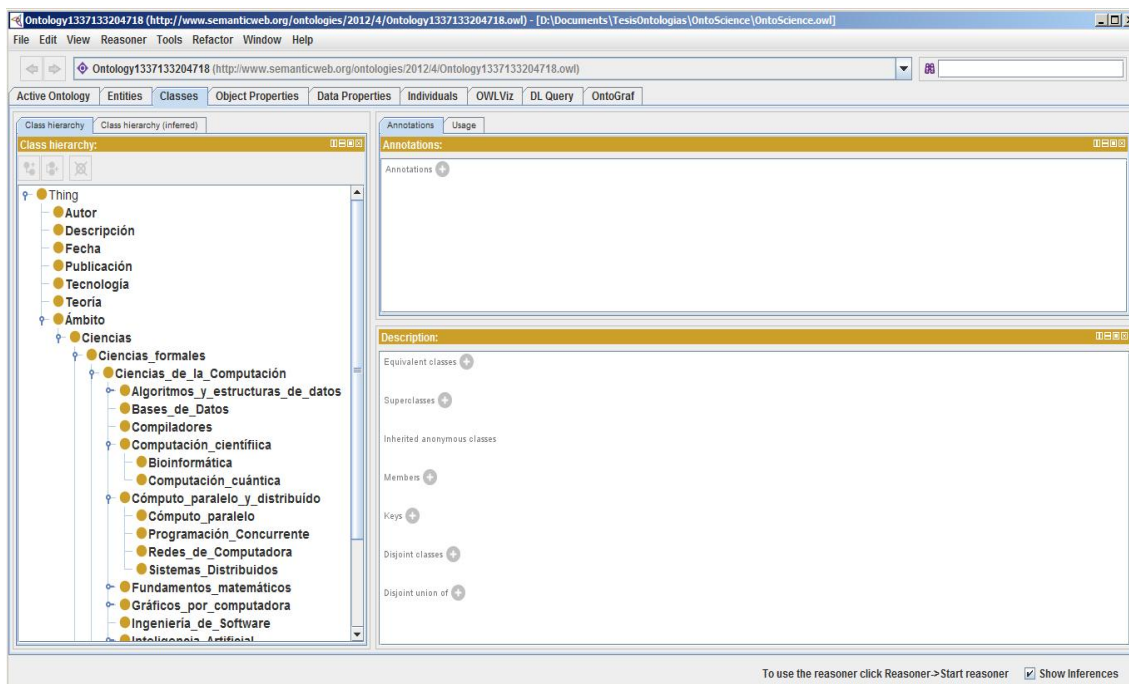


Figura 2.5: Clases de OntoScience

se va a descubrir el conocimiento en *OntoScience*. Se consultó el conjunto de publicaciones que estuvieran relacionadas con la instancia *tiranosaurios*.

Otra consulta que nos será muy útil al momento de realizar las consultas semánticas para la

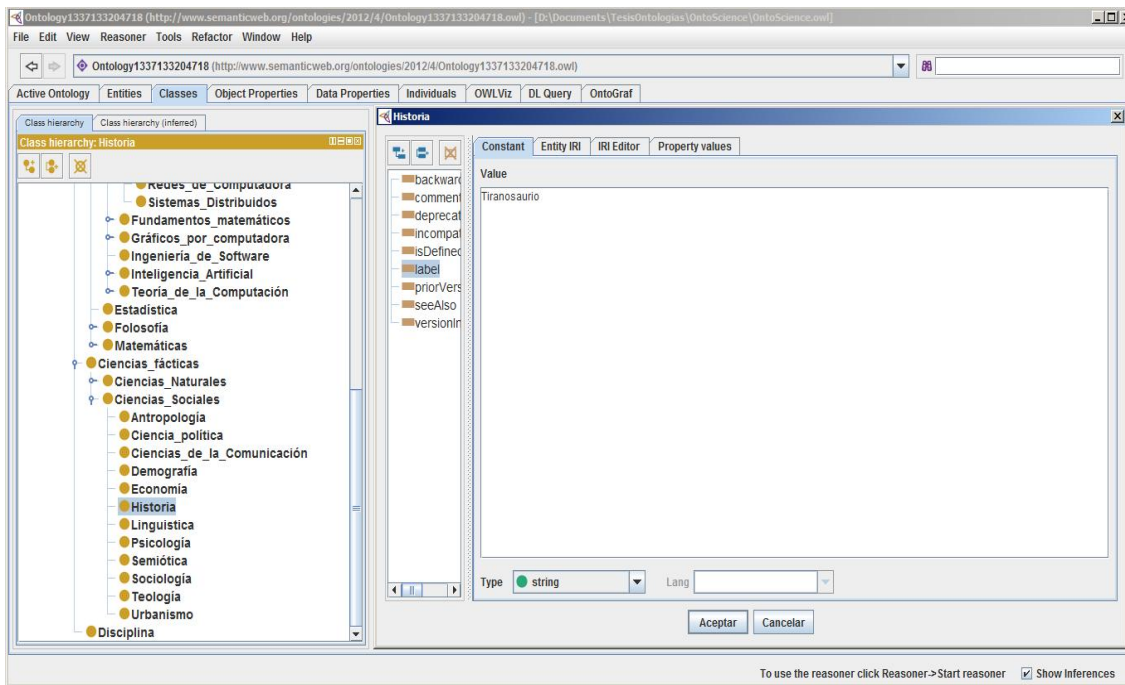


Figura 2.6: Ejemplo de configuración de la propiedad de anotación `rdfs:label`

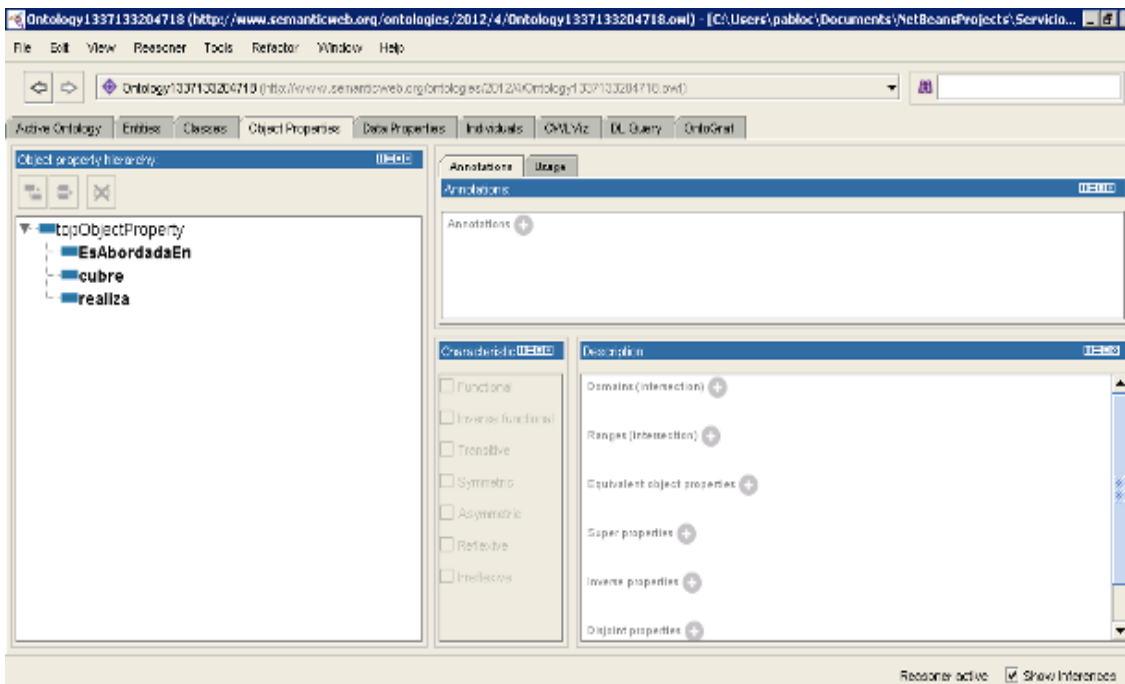


Figura 2.7: Propiedades de Objeto de la Ontología OntoScience.

red social es la relacionada con la relación entre los tipos de ámbitos que cubren las publicaciones realizadas por algún usuario (Autor). En la Figura 2.9 se muestra un ejemplo de una consulta de este tipo.

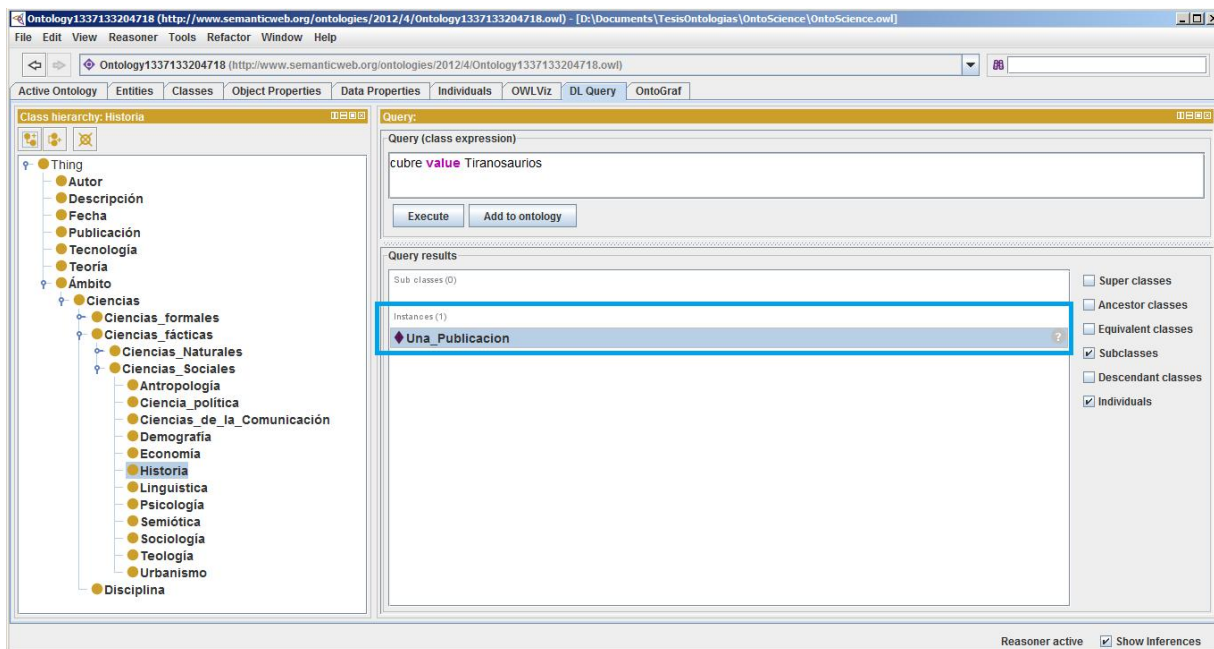


Figura 2.8: Ejemplo de consulta a OntoScience

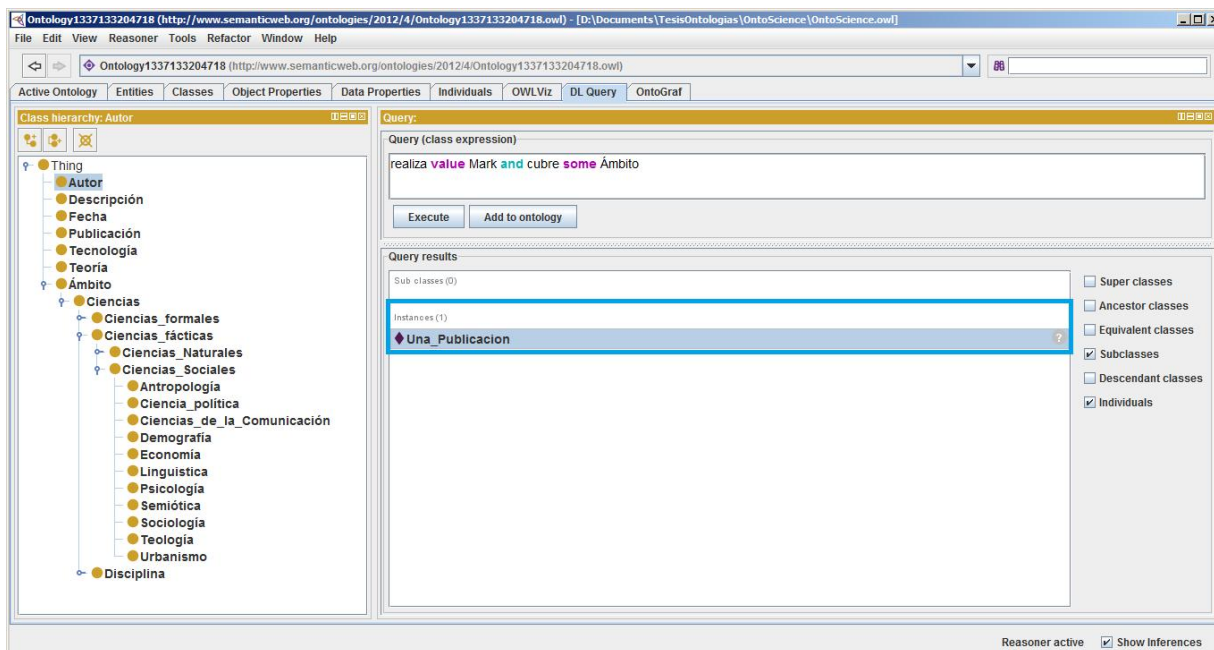


Figura 2.9: Ejemplo de consulta de publicaciones realizadas por un autor en un ámbito.

De esta forma se realizarán las consultas utilizando las relaciones que se han definido y las propiedades de anotación que cada instancia de cada clase tendrán.

2.1.1.5. Validación de la ontología

El razonador *Pellet* que previamente se usó para realizar las consultas, ahora lo presentamos como la herramienta que nos permitirá validar nuestra ontología.

Los aspectos que *Pellet* valida de una ontología son: la consistencia de la ontología y la obtención automática de la clasificación taxonómica. Aunque cualquier razonador se puede utilizar para realizar estas tareas de validación, en este caso usaremos *Pellet*. Analicemos los criterios anteriores considerados en *OntoScience*:

- **Consistencia de la ontología:** Permite validar que no existan contradicciones en los axiomas de la ontología. La semántica de OWL define una especificación formal para la definición de la consistencia en una ontología. Podemos validar la consistencia de la ontología programáticamente usando el API de desarrollo de Protégé.

```

1 URI uri = new
2 File("D:/Documents/TesisOntologias/OntoScience/OntoScience.owl").toURI();
3 try {
4     //Se carga la ontologa almacenada en memoria secundaria.
5     OWLOntology ont = manager.loadOntologyFromOntologyDocument(docIRI);
6     System.out.println("Cargada ! " + ont.getOntologyID());
7     //Se crea el razonador que validar la consistencia de la ontologa.
8     OWLReasonerFactory reasonerFactory = new Reasoner.ReasonerFactory();
9     ConsoleProgressMonitor progressMonitor = new ConsoleProgressMonitor
10        ();
11     OWLReasonerConfiguration config = new SimpleConfiguration(
12        progressMonitor);
13     OWLReasoner reasoner = reasonerFactory.createReasoner(ont, config);
14
15     reasoner.precomputeInferences();
16     //Se realiza la validacin.
17     boolean consistent = reasoner.isConsistent();
18     System.out.println("Consistent: " + consistent);
19     System.out.println("\n");
20 } catch (OWLOntologyCreationException e) {
21     e.printStackTrace();
22 }

```

Figura 2.10: Fragmento del código en Java que invoca a métodos del API de Protégé

Un objeto de la clase *OWLReasoner* razona sobre el conjunto de axiomas que está basado sobre las importaciones de una ontología particular (la ontología root). [HOR09]. En la Figura 2.10 se presenta el fragmento del código en Java que se empleó para validar la ontología programáticamente. En la Figura 2.11 se muestra el resultado de la validación de la consistencia de la ontología, invocando al método *isConsistent* de la clase *OWLReasoner* del API de desarrollo de protégé. El método antes mencionado, regresa el valor de true en caso de que la ontología sea consistente y false, en caso contrario.

- **Taxonomía:** Para validar la taxonomía de nuestra ontología presentamos el modelo jerárquico de clases inferido por Protégé. En la Figura 2.13 se presenta el modelo inferido por el razonador Pellet de Protégé de la ontología.

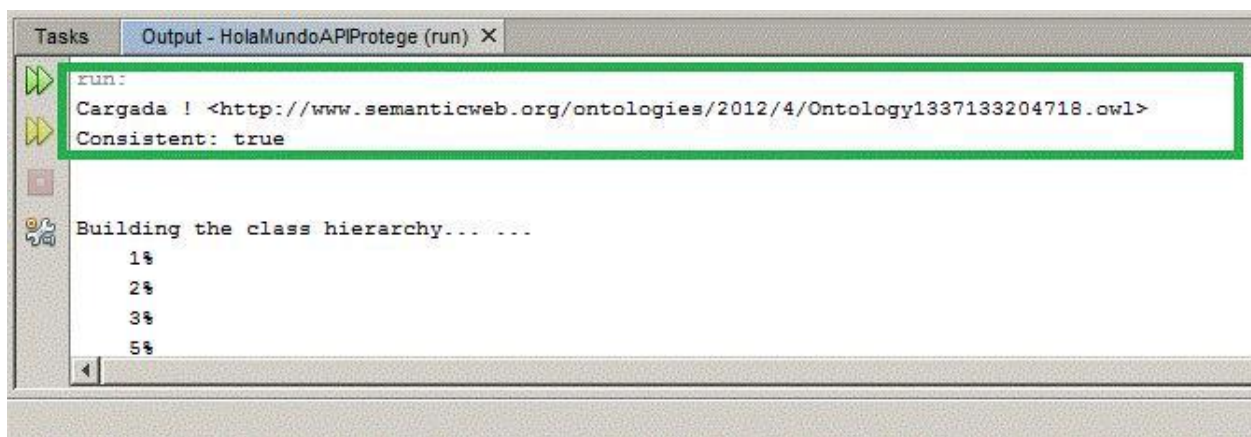


Figura 2.11: Resultado de la validación de consistencia a través del API de Protégé

Además, hemos incluido en la validación de la taxonomía de la ontología una de las consultas que planteamos en el desarrollo de la misma. Esta consulta se realizó utilizando una instancia del objeto *OWLReasoner* del API de desarrollo de Protégé. Para ello hemos creado tres instancias de prueba, una instancia de las clases: *Publicación*, *Autor* y *ámbito*, respectivamente. Para la instancia de la clase *Publicación*, se estableció la propiedad del objeto *cubre* como valor, la instancia de *ámbito*. De esta forma, la consulta que realizamos fue obtener la clase de la instancia que cubre la instancia de la clase *Publicación*. En la Figura 2.12 se muestra el resultado de esta consulta.

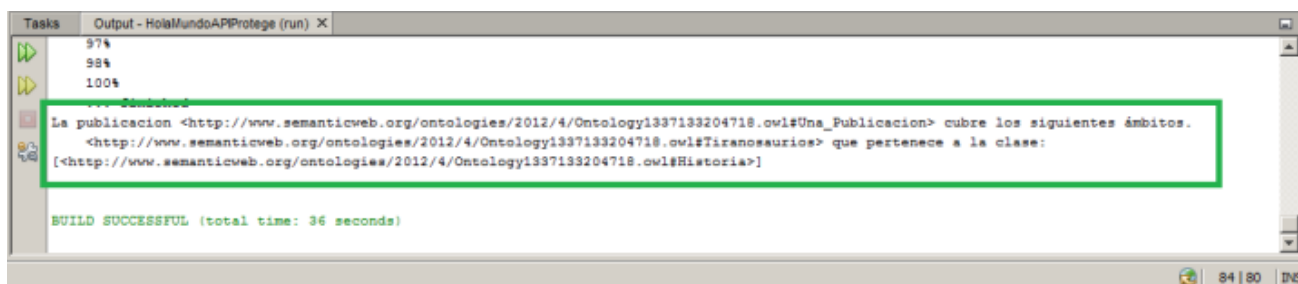


Figura 2.12: Resultado de la consulta realizada con el API de Protégé.

2.2. Conclusión

En este capítulo se ha abordado la construcción de la ontología, que es la forma de representación de conocimiento adoptada en nuestro trabajo de investigación y cuyo fin último es la realización de algunas tareas propias de la Web semántica que se detallaran posteriormente. También es importante mencionar que el uso de las ontologías tiene un propósito especial en este trabajo, ya que como forma de representación del conocimiento, ofrece una gama de posibilidades en la realización de consultas de la Web semántica a partir de las relaciones que se crean entre los términos de la ontología.

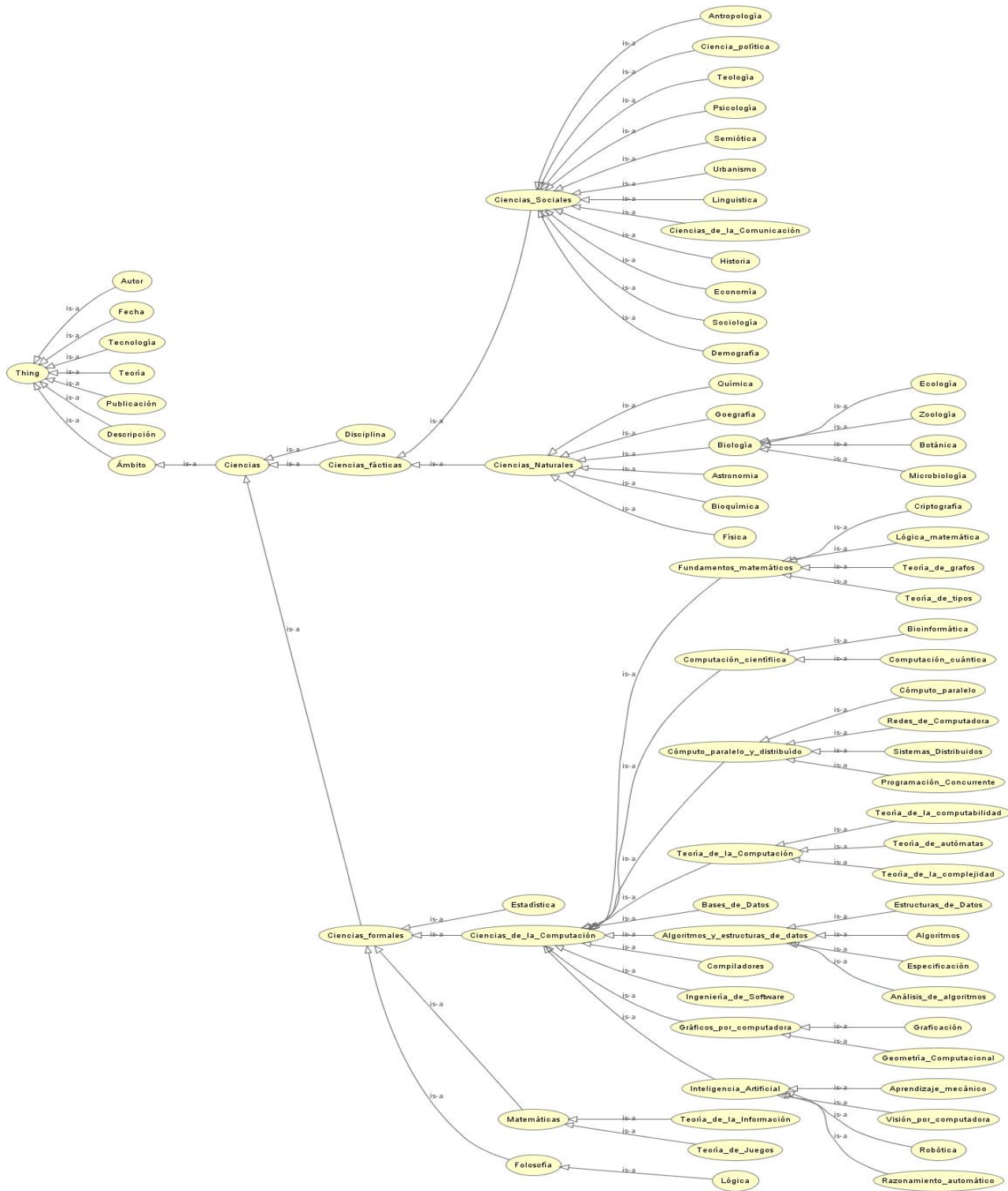


Figura 2.13: Modelo inferido por el razonador *Pellet* de la taxonomía de nuestra Ontología

Una vez que se ha implementado y validado la ontología, a continuación se presentara el marco teórico que sustenta la anotación semántica.

El proceso de asociar el contenido publicado en nuestra red social con los términos que se han modelado en la ontología, se conoce como **anotación semántica**.

Capítulo 3

La anotación en la Web semántica

En esta parte de la tesis se señalarán inicialmente las diferentes facetas de la anotación para después, estudiar el vínculo existente entre la anotación semántica y la ontología existente OntoSciencie. En seguida, se presentará una síntesis de herramienta de anotación semántica que se ha desarrollado para la tarea de anotación semántica en las publicaciones realizadas en Moveek.

3.1. La anotación y la Web semántica

En el estudio de la Web semántica, el objetivo es describir el contenido de los recursos anotándolos con información no ambigua con el fin de favorecer la explotación de estos recursos con agentes de software [GP04]. Ahora bien, los datos actuales de la web se escriben frecuentemente en lenguaje natural, ya que están destinados a los humanos. El lenguaje natural es por esencia muy ambiguo; es importante considerar alternativas formales y semánticamente explícitas para implementar mecanismos que disminuyan dichas ambigüedades, igualmente es importante considerar la relación entre el contenido de los recursos y sus anotaciones. La tarea de la anotación para la Web semántica consiste por lo tanto en tomar como entrada un recurso documental y proporcionar como salida el mismo contenido enriquecido con anotaciones semánticas basadas en la representación del conocimiento más o menos formal.

Nos interesamos más particularmente en los recursos terminológicos u ontológicos (RTO) [BAGC04] que permiten representar el conocimiento de un dominio. También ilustraremos cómo las anotaciones semánticas vinculan el documento fuente a los RTO [CB03]. Finalmente, se describirán algunos lenguajes que permiten expresar de una forma más o menos formal las anotaciones semánticas según los modelos formales utilizados.

3.1.1. Los recursos terminológicos u ontológicos (RTO)

Bourigault et al., definen la noción de recursos terminológicos u ontológicos en el cruce de las áreas de la terminología y de la inteligencia artificial, y más particularmente de la ingeniería de conocimientos [BAGC04]. Esta noción reagrupa varias clases de recursos desde los índices y glosarios hasta las ontologías pasando por las bases de datos léxicas y los tesauros. A continuación

se presentan los tres principales RTO que permiten representar y modelar el conocimiento de un dominio: Las taxonomías, los tesauros y las ontologías.

3.1.1.1. Las taxonomías

Los sistemas de representación de conocimientos y de razonamiento en inteligencia artificial necesitan representar la existencia de manera formal. Puede tratarse de objetos concretos como una Persona o un Automóvil o bien de nociones abstractas como un Estilo de Pintura o de una Idea. Ahora bien, la ciencia ha tenido como objetivo primario situar y clasificar los objetos del mundo para estudiarlos y comprenderlos [CHA02]. Esto es particularmente verdadero para las ciencias naturales que han construido clasificaciones, llamadas taxinomías o taxonomías, en nuestro ejemplo extraemos una parte de la taxonomía de la ontología, esta taxonomía es inferida por el razonador de Protégé. El objetivo de una taxonomía es el de conceptualizar los objetos del mundo y organizarlos jerárquicamente unos en relación a los otros.

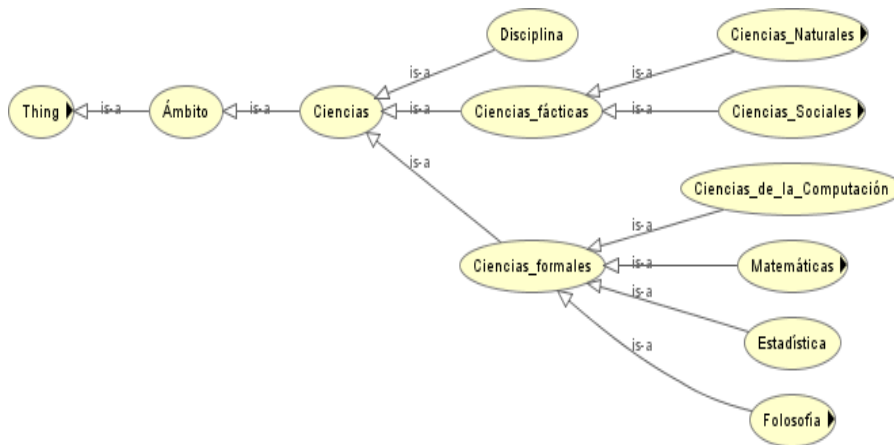


Figura 3.1: Extracto de la taxonomía de OntoScience.

La jerarquía de una taxonomía, modelada con una estructura arborescente representa más frecuentemente una relación de *subsumption*¹ o de hiperonomía², es decir “es un”. Esta relación es muy antigua y se remonta a la época de Aristóteles [CHA02]. Además los conceptos son próximos de la raíz, ya que su significado es general e inversamente, mientras estos son más próximos a las hojas, más su significado es específico al dominio considerado. Se llama también a esta relación “concepto/subconcepto” o “clase/subclase”. Esta relación compone generalmente la estructura básica de las ontologías como lo constataremos posteriormente, ya que es la base de los mecanismos de razonamiento en inteligencia artificial [LAU07]. En el ejemplo de la Figura 3.1, el concepto “Matemáticas” se considera como un subconcepto de “Ciencias_formales”, heredando por lo tanto todas las especificidades de este concepto.

¹La arquitectura subsumption es una arquitectura reactiva para los robots asociada fuertemente con la robótica basada en el comportamiento. Introducida por Rodney Brooks y sus colegas en 1986.

²En semántica lingüística, se denomina hiperónimo a aquel término que puede ser utilizado para referirse a la realidad nombrada por un término más particular. Por ejemplo, ser vivo es hiperónimo para los términos plantas y animales (hipónimos)

Es posible que la jerarquía represente otra relación, la de la agregación o de la meroninía³, es decir "parte/todo". Esta relación permite modelar el hecho de que un concepto se sitúa en el nivel inferior de la taxonomía al representar una subparte del concepto del nivel superior. La clasificación de los lugares geométricos es un ejemplo de utilización de una taxonomía de agregación: por ejemplo, la clase "México" puede modelarse como una subparte de la clase "América". Esta relación es igualmente frecuente en medicina para describir la organización atómica [CHA02]. En nuestro caso, la taxonomía que proponemos establece, por ejemplo, la clase "Robótica" como parte de la clase "Inteligencia Artificial".

3.1.1.2. Los tesauros

Las áreas de la computación documental y de la terminología se han interesado hace mucho tiempo en la modelación de los recursos terminológicos como los tesauros [CBT04]. Bourigault et al. definen un tesoro como "un lenguaje documental basado en una estructuración jerarquizada", si sabemos que un lenguaje documental es un "conjunto organizado de términos normalizados, utilizado para representar el contenido de los documentos para fines de memorización en una búsqueda posterior" [BAGC04]. Un tesoro es por lo tanto considerado como un vocabulario controlado y estructurado en el cual las relaciones entre los términos del dominio considerado se especifican claramente formando por lo tanto una red terminológica.

En el Cuadro 3.1 se muestran los términos que forman parte del tesoro de nuestra ontología. Estos términos son una muestra de los términos observados en los sitios mencionados en el Capítulo 2.

3.1.1.3. Las ontologías

El término ontología se ha definido inicialmente en filosofía como una rama de la metafísica que se interesa en la existencia, en el ser como tal y en las categorías fundamentales de la existencia [CHA02]. Es decir, este término se construye a partir de las raíces griegas *ontos*, es decir lo que existe, el ser, el existente, y *logos*, es decir el estudio, el discurso, de ahí su traducción del "estudio del ser" y por extensión "de la existencia". A principios de los 90's, los investigadores de inteligencia artificial se interesaron en esta noción para la formalización de los conocimientos. En esta disciplina, lo que "existe" puede ser "representado". En este contexto, ellos definieron una ontología como un artefacto que permite representar la existencia por medio de un vocabulario formal y consensual. Una de las primeras definiciones de la ontología comúnmente utilizada en inteligencia artificial se propuso por Gruber [GRU93a] como la "especificación explícita de una conceptualización". Esta definición de la ontología se detalló por R. Studer et al. [SBF98] como "especificación formal y explícita de una conceptualización compartida":

- Formal → la ontología debe ser legible por una máquina, lo que excluye el lenguaje natural. Es decir debe ser computable.
- Explícita → la definición explícita de conceptos utilizados y de restricciones de su utilización.

³La **meroninía** es una relación semántica entre los significados de dos palabras dentro del mismo campo semántico. Se denomina merónimo a la palabra cuyo significado constituye una parte del significado total de otra palabra, denominada holónimo. Ejemplo, dedo es un merónimo de mano y mano es merónimo de brazo; a su vez, brazo el holónimo de mano, y mano es holónimo de dedo.

| Ciencias Formales (Ciencias de la Computación) | | | |
|---|---------------------------------|------------------------------------|----------------------------|
| Arquitectura de computadoras | Computación cuántica | Algoritmos | Paralelismo |
| microprocesadores | cubit | algoritmo | paralelismo |
| ALU | | ordenacion | monitores |
| CPU | Computación evolutiva | arboles | hilos |
| Von Neumann | algoritmo genetico | arreglos | |
| hardware | | backtracking | Sistemas Operativos |
| memoria | Gráficos por computadora | busqueda | kernel |
| memoria principal | 3D | colas | procesos |
| perifericos | 2D | conjuntos | |
| procesador | animacion | grafos | Bases de Datos |
| procesador central | color | listas | SQL |
| puertos | escalar | pilas | registro |
| | iluminacion | | relacional |
| Inteligencia Artificial | luz | | |
| ontologia | mapeo | | |
| razonamiento | reflejo | | |
| red bayesiana | renderizado | | |
| red neuronal | rotar | | |
| semantica | textura | | |
| sistemas expertos | trasladar | | |
| | realidad virtual | | |
| | | | |
| Ciencias Fáticas | | | |
| Ciencias Naturales | | Ciencias Sociales | |
| Astronomia | Bioquimica | Antropología | Ciencia Política |
| galaxia | ADN | autoridad | consenso |
| sol | virus | | disenso |
| plantas | | Ciencias de la Comunicación | guerra |
| materia | oscura | Quimica | audiovisual justicia |
| observatorio | carbono | periodismo | negociacion |
| | etanol | publicidad | obediencia |
| Biología | plomo | | orden |
| aerobio | | Economía | paz |
| ala | Física | moneda | politica |
| alga | fuerza | | justicia |
| animales | gluones | Historia | |
| boca | luminosidad | Imperio Romano | |
| citoplasma | | | |
| digestion | Geografía | Psicología | |
| duodeno | rural | conducta | |
| ojos | territorio | | |
| organismos | territorios | | |
| poblacion | transporte | | |
| vida | turismo | | |
| adenina | | | |

Cuadro 3.1: Términos del tesoro de OntoScience

- Conceptualización → el modelo abstracto de un fenómeno del mundo real por identificación de los conceptos claves de este fenómeno.
- Compartida → la ontología no es propiedad de una persona, pero si representa un consenso aceptado por una comunidad de usuarios.

Claramente, una ontología proporciona los medios para expresar los conceptos de un dominio y organizarlos jerárquicamente definiendo sus propiedades semánticas en un lenguaje de representación de conocimientos formales, lo cual favorece el compartimiento de una vista consensual sobre este dominio entre las aplicaciones computacionales que la utilizan [BAGC04].

3.1.2. Los RTO y la anotación semántica

Vamos a revisar como los recursos terminológicos u ontológicos (RTO) pueden utilizarse para anotar semánticamente un documento. Hemos constatado que los RTO no ofrecen por supuesto el mismo nivel de representación del conocimiento: las taxonomías son menos expresivas que los tesauros, son de hecho menos formales y precisas que una ontología. La elección de un RTO por lo tanto va a influenciar la modelización de las posibles anotaciones semánticas.

Sabemos igualmente que los tesauros no constituyen una ontología por sí mismos, su capacidad de expresar un vocabulario controlado así como un conjunto de relaciones léxicas entre los términos de este vocabulario permite constituir un primer nivel de anotación semántica. En este caso, las anotaciones consisten en simples apuntadores hacia los términos de los tesauros, igualmente llamados descriptores. Estos funcionan entonces como entradas de índice que mejoran los resultados de los motores de búsqueda que utilizan esta clase de anotación.

Las ontologías de dominio son cada vez más utilizadas para anotar los recursos documentales. De hecho, un mismo recurso puede anotarse por medio de diferentes ontologías de dominio al mismo tiempo, ofreciendo así diferentes puntos de vista sobre un mismo contenido. No solamente, las anotaciones semánticas creadas a partir de una ontología mejoran la búsqueda de información, ya que estas pueden contribuir o evolucionar los mecanismos de inferencia y de razonamiento, e igualmente pueden combinarse con el enriquecimiento de una base de conocimiento del dominio.

Es decir, el conocimiento del dominio contenido en el documento puede extraerse a partir de las anotaciones semánticas con el fin de ser almacenado y explotado en una base de conocimiento restringida por la ontología de dominio.

De hecho, los tesauros y las ontologías ofrecen acceso a los contenidos desde diferentes ángulos: el del vocabulario para los tesauros y el de la conceptualización de un dominio para la ontología. Estos tienen por lo tanto, roles complementarios en el proceso de la anotación semántica como se podrá apreciar en el siguiente ejemplo.

Se trata de un resumen de un artículo titulado "El clan Coppola", publicado el 30/02/2003 en la revista ELLE, ver la Figura 3.2. Anotamos inicialmente el extracto con un tesauro geográfico como el TNG. En la Figura 3.2, estas anotaciones corresponden a las palabras en verde:

- Detroit es una referencia al descriptor que tiene como ID 7013547 en el TGN y que hace referencia a la ciudad de Detroit en el estado de Michigan.
- Michigan es una referencia al descriptor que tiene como ID 7007520 y que hace referencia al estado de Michigan en Estados Unidos.

- New York es una referencia al descriptor que tiene como ID 7007567 en el TGN y que hace referencia a la ciudad de New York en los Estados Unidos.

En CLAN COPPOLA Francis Coppola nació el 7 de Abril de 1939 en [Detroit](#) en [Michigan](#). Es el segundo de tres hijos de Carmine e Italia Coppola. Su papá, originario de [New York](#), es director de orquesta. Francis le pedirá que componga la música del "Padrino" en 1972. Su mamá, es hija del célebre compositor napolitano Francesco Pennino, autor de la opera "Senza Mamma" un pequeño extracto se utilizó en "El Padrino II". Como comediente, ella acutúa en varias películas de Vittorio De Sica, antes de ser mamá.

Figura 3.2: Extracto del artículo "El clan Coppola".

Después, suponemos que existe una ontología en el dominio de la "prensa de gente famosa" que podríamos utilizar para anotar el extracto antes mencionado. Esta ontología contendrá los conceptos "Personalidad" y "Obra Artística", de los atributos como "lugar de nacimiento", "fecha de nacimiento", "fecha de creación", "Profesión", etc., y las relaciones tales como "EsCreadoPor(Obra Artística, Personalidad)" o "TieneVínculoFamiliarCon(Personalidad, Personalidad)". Los documentos serían entonces anotados en función de estos elementos, es decir:

- Por las instancias de conceptos. Por ejemplo, una anotación será creada con "Francis Ford Coppola" que representa una instancia del concepto "Personalidad" y con "El Padrino", instancia del concepto "Obra Artística".
- Por los valores de atributos. Por ejemplo, "7 de abril de 1939" puede ser el valor del atributo "fecha de nacimiento" que será vinculada a la instancia "Francis Ford Coppola" de nuestra anotación.
- Por las instancias de relaciones. Por ejemplo, la instancia "El Padrino" del concepto Obra Artística y la instancia "Francis Ford Coppola" del concepto Personalidad pueden ser conectadas por la relación "es-creada-por". El documento será anotado con esta instancia de relación "EsCreadaPor(El Padrino, Francis Ford Coppola)".

Para este ejemplo, vimos como los tesauros y otros vocabularios controlados pueden utilizarse para proporcionar un conjunto de términos agregados en los dominios particulares como los lugares geográficos. Este primer nivel puede a continuación combinarse con las anotaciones semánticas basadas en la ontología del dominio que describen el contenido del documento en función de los conceptos, atributos y relaciones modeladas en esta ontología. Mientras más precisa sea una ontología en su modelización del dominio, más las anotaciones estarán restringidas en función de las restricciones impuestas en sus valores autorizados, sobre sus relaciones, etc. Por consecuencia, ya que son de naturaleza diferente, estos enfoques son complementarios unos con otros. Es por lo general interesante anotar un recurso documental por estos diferentes enfoques con el fin de ofrecer ángulos diferentes a los servicios y a otros agentes de software que exploten las anotaciones creadas. Esto permite notablemente proporcionar diferentes accesos hacia los mismos recursos documentales en búsqueda de información (búsqueda por palabras clave, por extensión semántica, búsqueda semántica, etc.).

Presentamos la noción de RTO, después las taxonomías hasta las ontologías pasando por los tesauros, y mostramos como estos pueden emplearse para la anotación semántica. Presentaremos cuales lenguajes computacionales pueden utilizarse para representar el conocimiento de un

dominio, instrumentar en un agente de software y así crear las anotaciones semánticas que se derivan.

3.2. Herramienta de anotación

Una herramienta de anotación semántica es una herramienta de software que permite insertar y administrar anotaciones semánticas vinculadas al menos a un recurso documental dado. En el marco de la Web Semántica, las herramientas de anotación semántica utilizan una ontología, o por lo menos un modelo formal, que formaliza y estructura las anotaciones producidas en función de los conceptos y restricciones definidos en esta ontología.

Nosotros presentamos en esta sección el componente de software que nos permite insertar la anotación semántica en las publicaciones de nuestra red social. Como tal la tarea de realizar la anotación semántica consiste en dos fases: la primera es la extracción de la información y la segunda es la anotación semántica a partir de la información extraída.

En nuestra aplicación, la fase de la extracción de la información se realiza a través de un servicio Web que se detallará en el Capítulo 4. En esta sección analizaremos el resultado de este servicio para introducir la tarea de la anotación semántica. En la Figura 3.3 se muestra el resultado de la invocación al servicio Web que realiza la extracción de la información contenida en las publicaciones realizadas en la red social Moveek.

```

1 <GateDocument>
2 <!-- The document's features -->
3
4 <GateDocumentFeatures>
5 <Feature>
6   <Name className="java.lang.String">MimeType</Name>
7   <Value className="java.lang.String">text/plain</Value>
8 </Feature>
9 <Feature>
10  <Name className="java.lang.String">gate.SourceURL</Name>
11  <Value className="java.lang.String">file:/C:/Program Files/glassfish-3.1.1/
    glassfish/domains/domain1/publicacion.tmp.txt</Value>
12 </Feature>
13 <Feature>
14  <Name className="java.lang.String">docNewLineType</Name>
15  <Value className="java.lang.String"></Value>
16 </Feature>
17 </GateDocumentFeatures>
18 <!-- The document content area with serialized nodes -->
19
20 <TextWithNodes><Node id="0"/>el <Node id="3"/>ADN<Node id="6"/> es la clave de
    la <Node id="25"/>evolucion<Node id="34"/></TextWithNodes>
21 <!-- The default annotation set -->
22
23 <AnnotationSet>
24 <Annotation Id="16" Type="Split" StartNode="34" EndNode="34">
25 <Feature>

```

```

26     <Name className=" java.lang.String">kind</Name>
27     <Value className=" java.lang.String">external</Value>
28 </Feature>
29 </Annotation>
30 <Annotation Id="18" Type="Lookup" StartNode="3" EndNode="6">
31 <Feature>
32     <Name className=" java.lang.String">majorType</Name>
33     <Value className=" java.lang.String"></Value>
34 </Feature>
35 <Feature>
36     <Name className=" java.lang.String">classURIList</Name>
37     <Value className=" java.util.HashSet" itemClassName=" java.lang.String">http://
        www.semanticweb.org/ontologies/2012/4/Ontology1337133204718.owl#Bioquimica
        </Value>
38 </Feature>
39 <Feature>
40     <Name className=" java.lang.String">heuristic_level</Name>
41     <Value className=" java.lang.Integer">0</Value>
42 </Feature>
43 <Feature>
44     <Name className=" java.lang.String">type</Name>
45     <Value className=" java.lang.String">instance</Value>
46 </Feature>
47 <Feature>
48     <Name className=" java.lang.String">classURI</Name>
49     <Value className=" java.lang.String">http://www.semanticweb.org/ontologies
        /2012/4/Ontology1337133204718.owl#Bioquimica</Value>
50 </Feature>
51 <Feature>
52     <Name className=" java.lang.String">URI</Name>
53     <Value className=" java.lang.String">http://www.semanticweb.org/ontologies
        /2012/4/Ontology1337133204718.owl#ADN</Value>
54 </Feature>
55 </Annotation>
56 <Annotation Id="20" Type="Lookup" StartNode="25" EndNode="34">
57 <Feature>
58     <Name className=" java.lang.String">majorType</Name>
59     <Value className=" java.lang.String"></Value>
60 </Feature>
61 <Feature>
62     <Name className=" java.lang.String">classURIList</Name>
63     <Value className=" java.util.HashSet" itemClassName=" java.lang.String">http://
        www.semanticweb.org/ontologies/2012/4/Ontology1337133204718.owl#Biologa</
        Value>
64 </Feature>
65 <Feature>
66     <Name className=" java.lang.String">heuristic_level</Name>
67     <Value className=" java.lang.Integer">0</Value>
68 </Feature>
69 <Feature>
70     <Name className=" java.lang.String">type</Name>
71     <Value className=" java.lang.String">instance</Value>
72 </Feature>

```

```
73 <Feature>
74   <Name className="java.lang.String">classURI</Name>
75   <Value className="java.lang.String">http://www.semanticweb.org/ontologies
      /2012/4/Ontology1337133204718.owl#Biologa</Value>
76 </Feature>
77 <Feature>
78   <Name className="java.lang.String">URI</Name>
79   <Value className="java.lang.String">http://www.semanticweb.org/ontologies
      /2012/4/Ontology1337133204718.owl#evolucion</Value>
80 </Feature>
81 </Annotation>
82 </AnnotationSet>
83
84 <!-- Named annotation set -->
85
86 <AnnotationSet Name="Original markups">
87 <Annotation Id="0" Type="paragraph" StartNode="0" EndNode="34">
88 </Annotation>
89 </AnnotationSet>
90
91 </GateDocument>
```

Figura 3.3: Resultado del servicio de extracción de la información

Una vez obtenida el documento XML mostrado en la Figura 3.3 se procede a realizar la anotación. La implementación del servicio Web que realiza la extracción de la información se realiza en Java por la naturaleza de la herramienta GATE utilizada para realizar esta tarea, sin embargo la herramienta que realiza la tarea de la anotación semántica será implementado en el lenguaje que consume el servicio Web de la extracción de la información, es decir se desarrollará en C#.

En la Figura 3.4 se muestra el diagrama de actividad que describe la forma en la que se realiza la tarea de anotar la publicación entrante a partir del documento XML obtenido del servicio Web de extracción de la información.

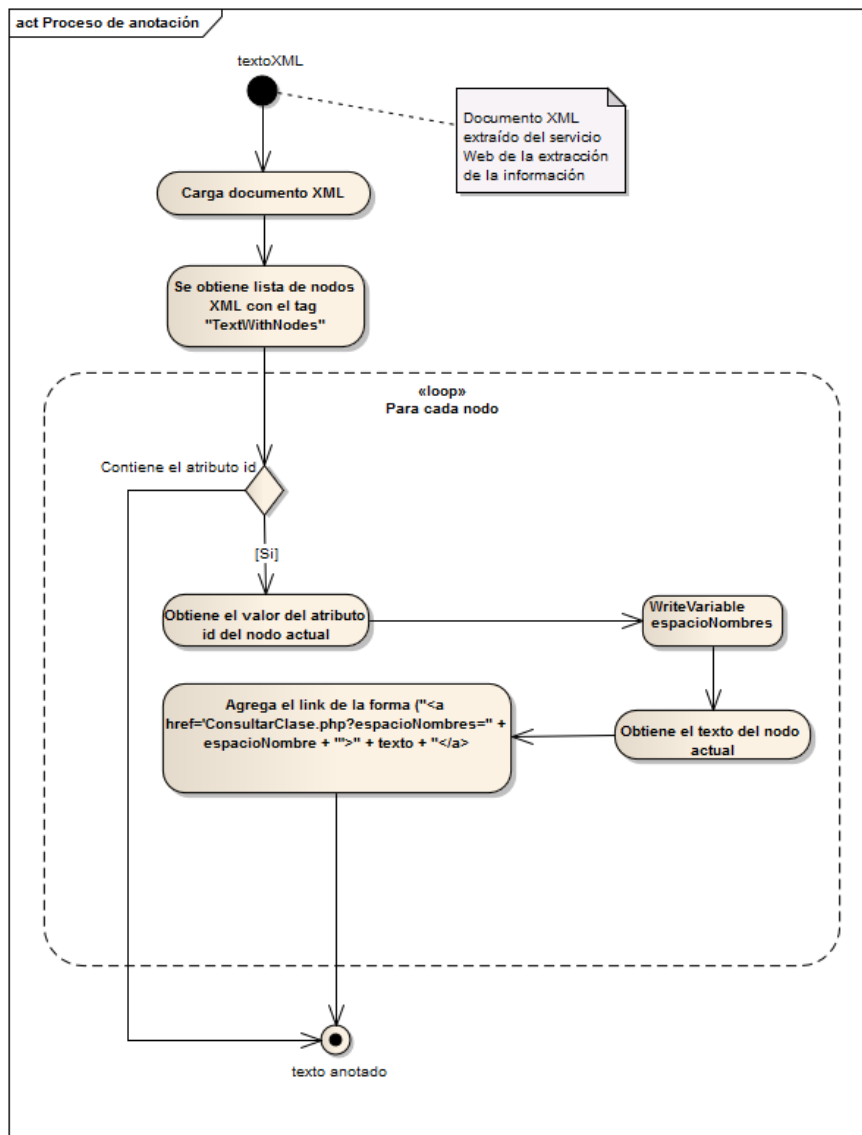


Figura 3.4: Diagrama de actividad del proceso de anotación semántica

El texto que se obtiene como resultado de la acción de hacer la anotación semántica descrita en la Figura 3.4 es similar al mostrado en la Figura 3.5.

```

1 el <a href='ConsultarClase.php?espacioNombres=http://www.semanticweb.org/ontologies/2012/4/Ontology1337133204718.owl#Bioquimica '>ADN</a> es la clave de la <a href='ConsultarClase.php?espacioNombres=http://www.semanticweb.org/ontologies/2012/4/Ontology1337133204718.owl#Biologa '>evolucion</a>

```

Figura 3.5: Ejemplo de publicación anotada en HTML

3.3. Discusión

3.3.1. Síntesis con respecto a la anotación semántica

Acabamos de presentar en este capítulo el concepto de anotación semántica y constatamos que en el estudio de la Web Semántica, está intrínsecamente vinculado a la modelización de una ontología. Es decir, esta ontología va a representar los conceptos, atributos y relaciones de un ámbito con ayuda de un lenguaje de representación de los conocimientos orientados a la Web como OWL. Las anotaciones semánticas se estructuran con ayuda de esta ontología y sus valores apuntan hacia las instancias de la ontología de referencia o, en algunos casos, directamente hacia los propios conceptos. Los otros Recursos Terminológicos u Ontológicos, como el tesoro, pueden también utilizarse como valor de la anotación semántica con el fin de proporcionar una perspectiva diferente al usuario final en el tema de un mismo recurso documental.

Vimos que dado que existe una gama bastante importante de herramientas de anotación semántica, que tienen cada una características propias pero cuyo objetivo cada vez es más orientado hacia la asistencia de los anotadores humanos en la creación de las anotaciones. El nivel de automatización depende del motor de extracción de información integrado en la herramienta. Las anotaciones pueden almacenarse en un servidor de anotación, en modo embarcado o descargado frente al documento de origen, o en una base de conocimiento. Pueden, entre otras cosas, servir para mejorar los sistemas de búsqueda de información, para poblar (enriquecer) una ontología existente, o incluso ayudar a la construcción de ontologías. Nosotros enfocamos nuestro estudio en las herramientas presentadas que son más o menos resultantes de la investigación. A pesar de una evolución rápida de los lenguajes y reglas, gracias, en particular, al desarrollo de la Web Semántica, estas herramientas siempre no se adaptan a las utilidades concretas del mundo real. Y para que las empresas, que tienen deseos de beneficiarse del nuevo modelo especificado por el marco fundador de la web Semántica, puedan integrarlos en sus aplicaciones, algunos límites aún existentes deben rechazarse.

1. Las herramientas de anotación están intrínsecamente vinculadas al motor de extracción utilizado para extraer la información de los textos y así anotarlos. En nuestro caso la herramienta de anotación será proporcionada a la red social mediante los servicios Web que realizarán las tareas de anotación. Pensamos que es necesario a toda costa disociarlos para que la herramienta de anotación pueda utilizar tal o cual motor de extracción en función de las necesidades de la aplicación en la cual se inscribe.
2. La mayoría de los motores de extracción se basan en los procesos de aprendizaje supervisado. Estos sistemas son quizás potentes para el contenido estructurado, o incluso semi-estructurado. Pero sobre la Web y en las empresas, el contenido no estructurado es el que contiene potencialmente nueva información estratégica, en particular, las relaciones semánticas entre entidades, y cuya semántica es la más difícil de extraer [AS00]. Consideramos que en este caso, es necesario darle prioridad a los sistemas de extracción basados en distintos análisis lingüísticos finos que permitan definir esta semántica en el contenido de los documentos tratados.

3. La anotación semántica debe seguir favoreciendo los enfoques basados en ontologías de ámbito, y no genéricas. Esto es sobre todo en el caso de las aplicaciones destinadas al mundo de las empresas. Estas últimas están interesadas por hechos inherentes a su propio ámbito de aplicación y no por generalidades.
4. Es necesario poder proporcionar una ayuda máxima al usuario y, en particular, cuando el objetivo consiste también en poblar (enriquecer) la ontología. Los procesos y las interfaces deben no solamente presentarle sugerencias, propuestas sino que deben también estar en condiciones de guiarlo tomando en cuenta las dificultades y restricciones modeladas en la ontología de referencia. Estamos convencidos de que la cuestión de la integridad de las anotaciones, y más concretamente de la base de conocimiento que acoge a las nuevas instancias resultantes de estas anotaciones, es primordial para la explotación de los resultados proporcionados por tal aplicación.

3.4. Conclusión

Una vez abordado el marco teórico referente a la anotación semántica es momento de concluir la importancia de la misma en este trabajo de tesis.

La anotación semántica es el mecanismo encargado de establecer un vínculo entre el contenido publicado en la Web (en nuestro caso las publicaciones textuales realizadas en nuestra red social) y alguna representación de conocimiento (tesauros, taxonomías u ontologías). En nuestro caso la ontología brinda un marco para clasificar el contenido de la web.

También, la anotación semántica permite hacer que la red social *comprenda*, por decirlo de alguna forma, el contenido que se publica en la red social a partir de la clasificación que le otorga la ontología, es decir es el núcleo que hace funcionar a la web semántica.

Ahora que ya se ha mencionado la forma de representación de conocimiento (Capítulo 2) y la forma en la que se vinculará el contenido de la red social a la ontología (Capítulo 3) es momento de abordar el desarrollo de la red social para conocer la manera en la que se implementarán en la red social las tareas de la anotación semántica y la manera en la que realiza las consultas semánticas a la ontología.

Capítulo 4

La red social Moveek

En este capítulo se presentará la red social que se ha usado en este trabajo de tesis, si bien se ha mencionado brevemente, en este capítulo se detallará su implementación. Primero, se presentan los diagramas de análisis y diseño elaborados durante el desarrollo de la red social, después se presenta la composición de servicios Web empleados para hacer interoperables los servicios desarrollados en las plataformas de .NET y Java que realizan las tareas propias de la gestión de la información de la red social y la anotación semántica u operaciones con la ontología respectivamente, es decir realizan el descubrimiento de conocimiento contenido en las publicaciones de la red social.

4.1. Análisis y diseño de la red social Moveek

Como parte del desarrollo de cualquier aplicación Web, se realizaron las tareas básicas de análisis y diseño, a continuación se presentan los diagramas más importantes de estas etapas y en el apéndice C se muestran todos los diagramas desarrollados.

4.1.1. Análisis

Como primer fase del análisis de nuestra red social, se realizó la ingeniería de requerimientos, la cual ha dado como resultado el siguiente diagrama de casos de uso (Ver Figura 4.1).

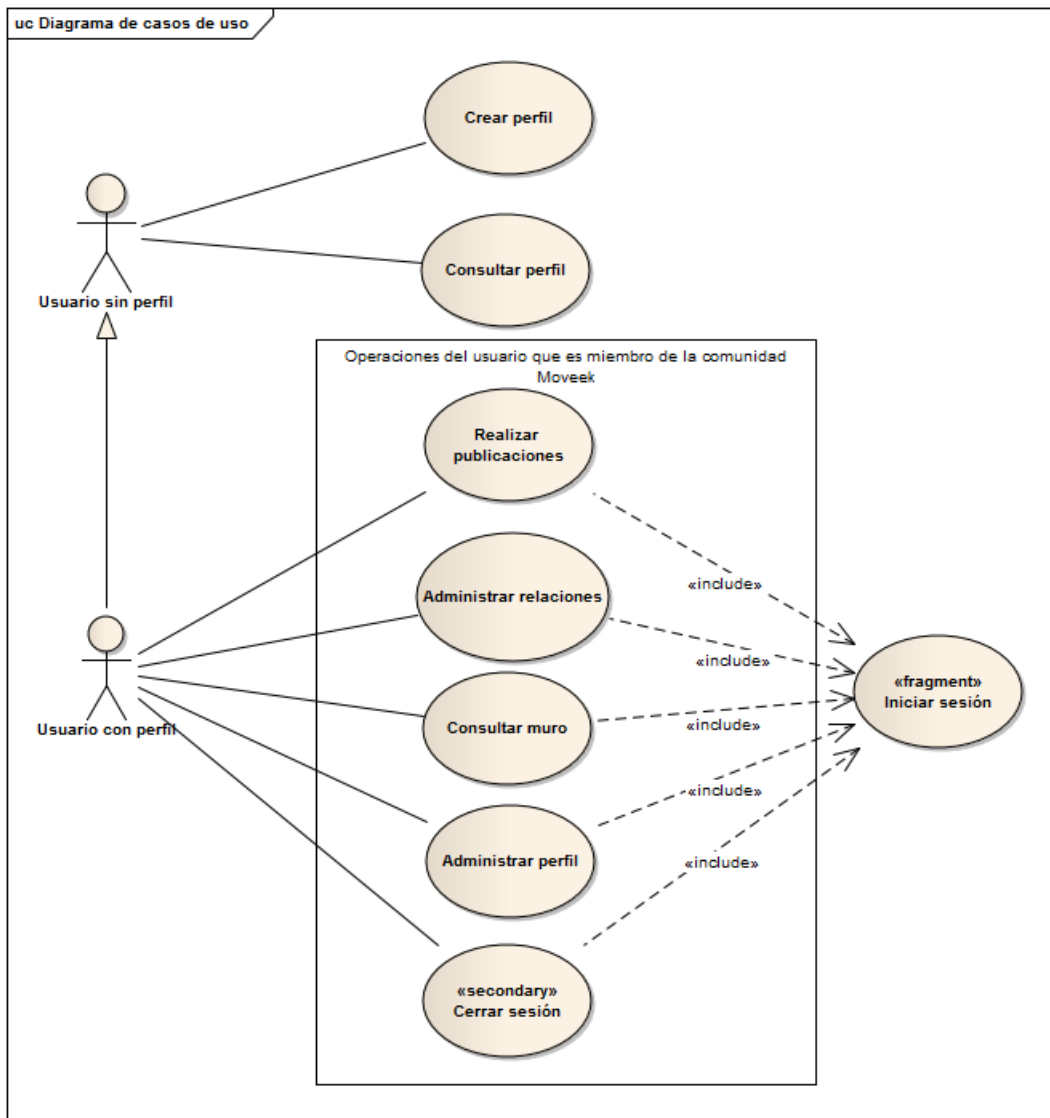


Figura 4.1: Diagrama de Casos de Uso de Moveek

Se puede observar que en total son 2 tipos de actores los que se han identificado que podrán interactuar con la red social. Existen 6 casos de uso principales, 1 fragmento y 1 caso de uso secundario, en total 8 casos de uso. Para no abrumar al lector con diagramas y descripciones, se omiten detalles finos del proceso empleado en la concepción de la red social, pero esto no quiere decir que no se ha seguido un estricto proceso de ingeniería Web.

Una vez identificados los casos de uso de nuestra red social, es momento de realizar el diagrama de clases de análisis, el cual modela las necesidades del usuario en términos de los conceptos inherentes a cada tarea modelada en el diagrama de casos de uso. En la Figura 4.2 se presenta este diagrama. Como se puede observar en la Figura 4.2, cada caso de uso ha quedado modelado en este diagrama de clases de análisis. Como última parte de la etapa de análisis de nuestra red social, se presenta el modelo de navegación de la misma. Este modelo de navegación sirve para

darnos una idea del número de páginas, frames, mensajes de error, etc. Que se deben considerar al momento de implementar el sitio Web.

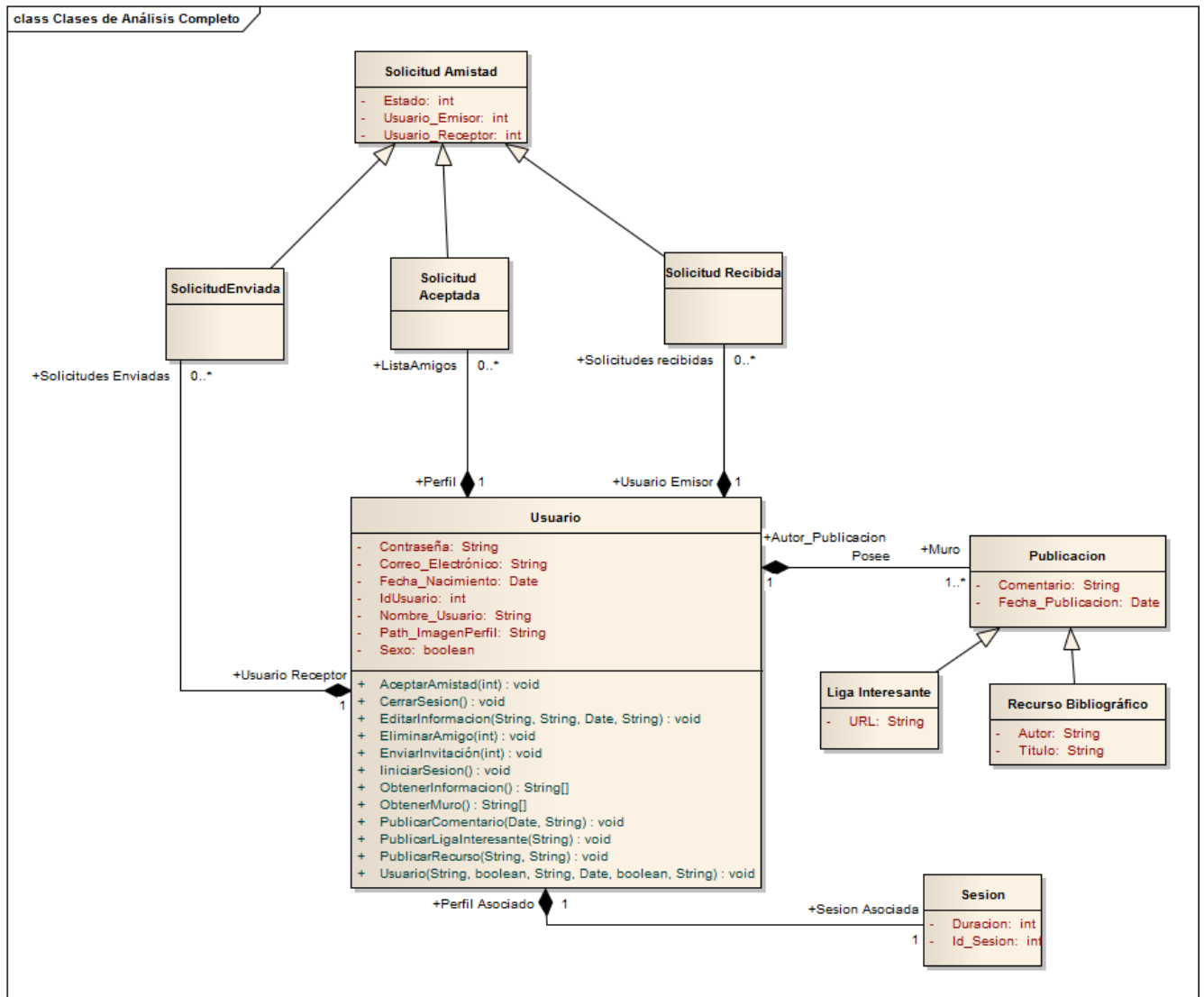


Figura 4.2: Diagrama de Clases de Análisis de Moveek.

La Figura 4.3 presenta el modelo de navegación general de Moveek. Debemos recordar que este modelo es también resultado de la elaboración de la maqueta de interacción humano-computadora. (Ver Apéndice C).

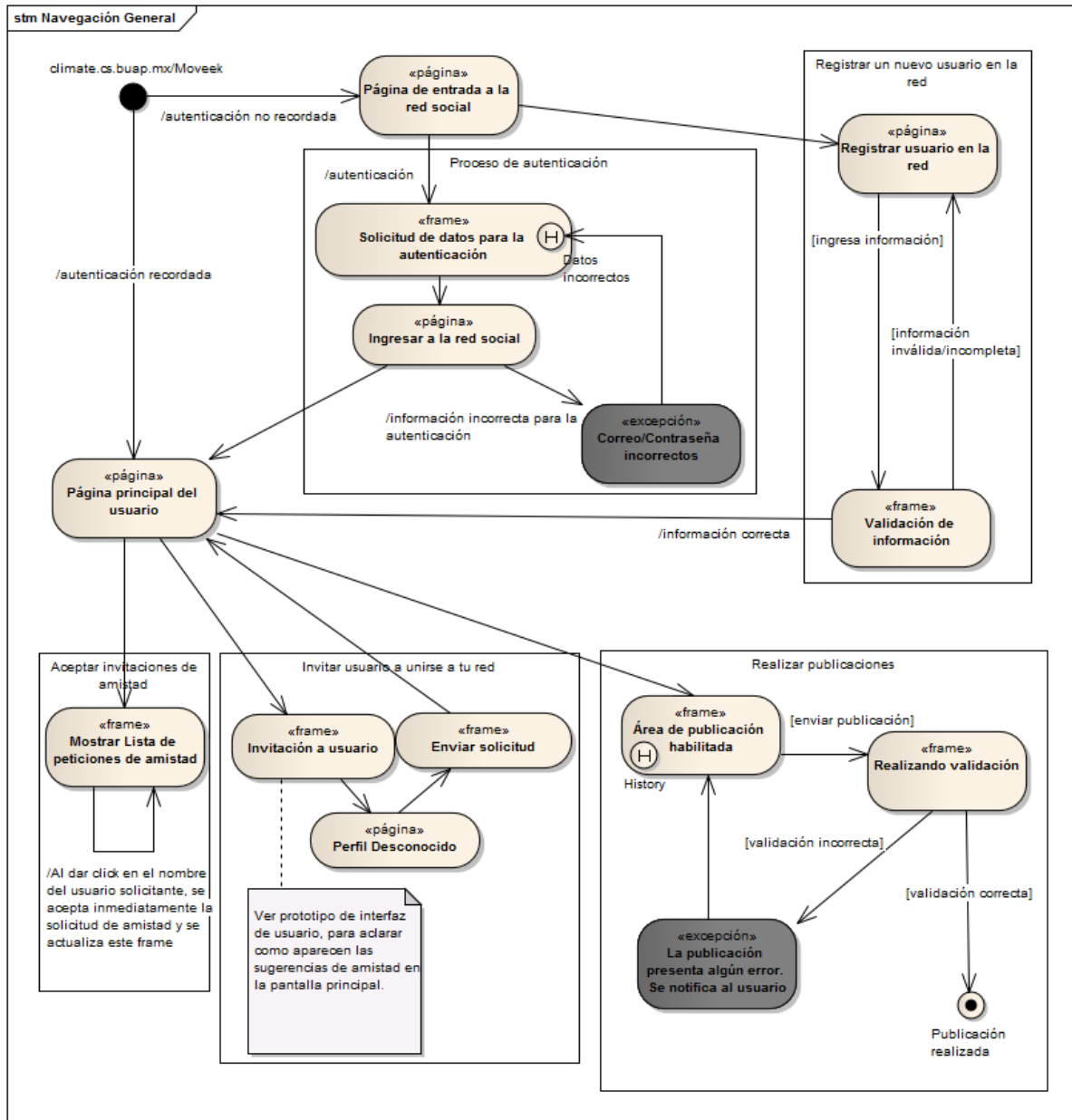


Figura 4.3: Modelo de Navegación general de Moveek

4.1.2. Diseño

Una vez que se han presentado los resultados de la etapa de análisis, es momento de presentar los resultados de la etapa de diseño, esto con el fin tener lo necesario para poder codificar la red social.

Primero presentamos los diagramas de secuencia detallados de los principales casos de uso para después presentar el diagrama de clases de diseño. En la Figura 4.4 se presenta el diagrama de secuencia detallado del caso de uso *Crear Perfil*. Además, se presenta el diagrama de secuencia

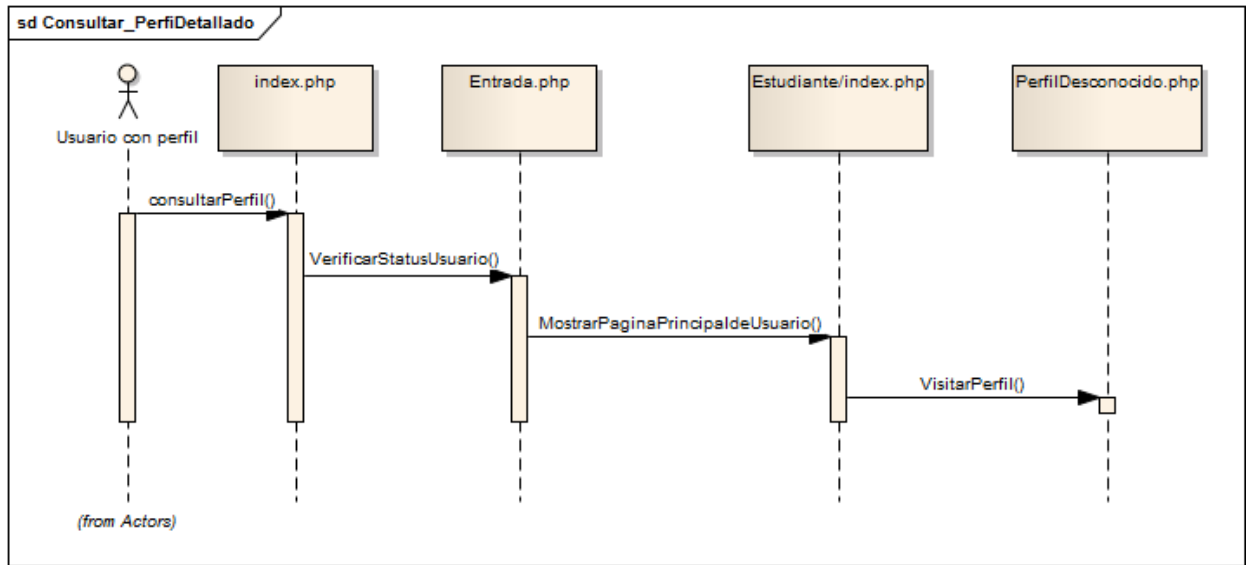


Figura 4.4: Diagrama de Secuencia del Caso de Uso Crear Perfil

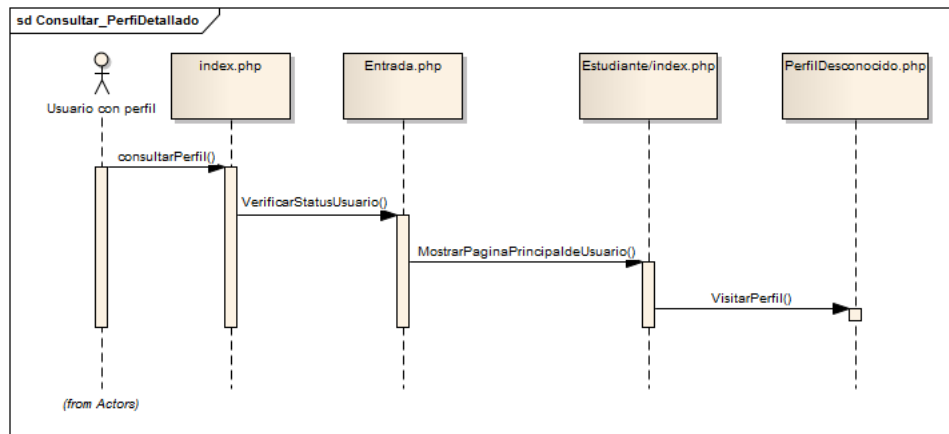


Figura 4.5: Diagrama de Secuencia del Caso de Uso Consultar Perfil

que modela la forma en la que se visita un perfil de la red. Ver Figura 4.5.

Se puede observar que en esta clase de diagramas se presentan todos los objetos necesarios para llevar a cabo la tarea del caso de uso.

En la Figura 4.6 se presenta el diagrama de secuencia del caso de uso realizar publicaciones el cual muestra los objetos y los mensajes que son necesarios para realizar una publicación.

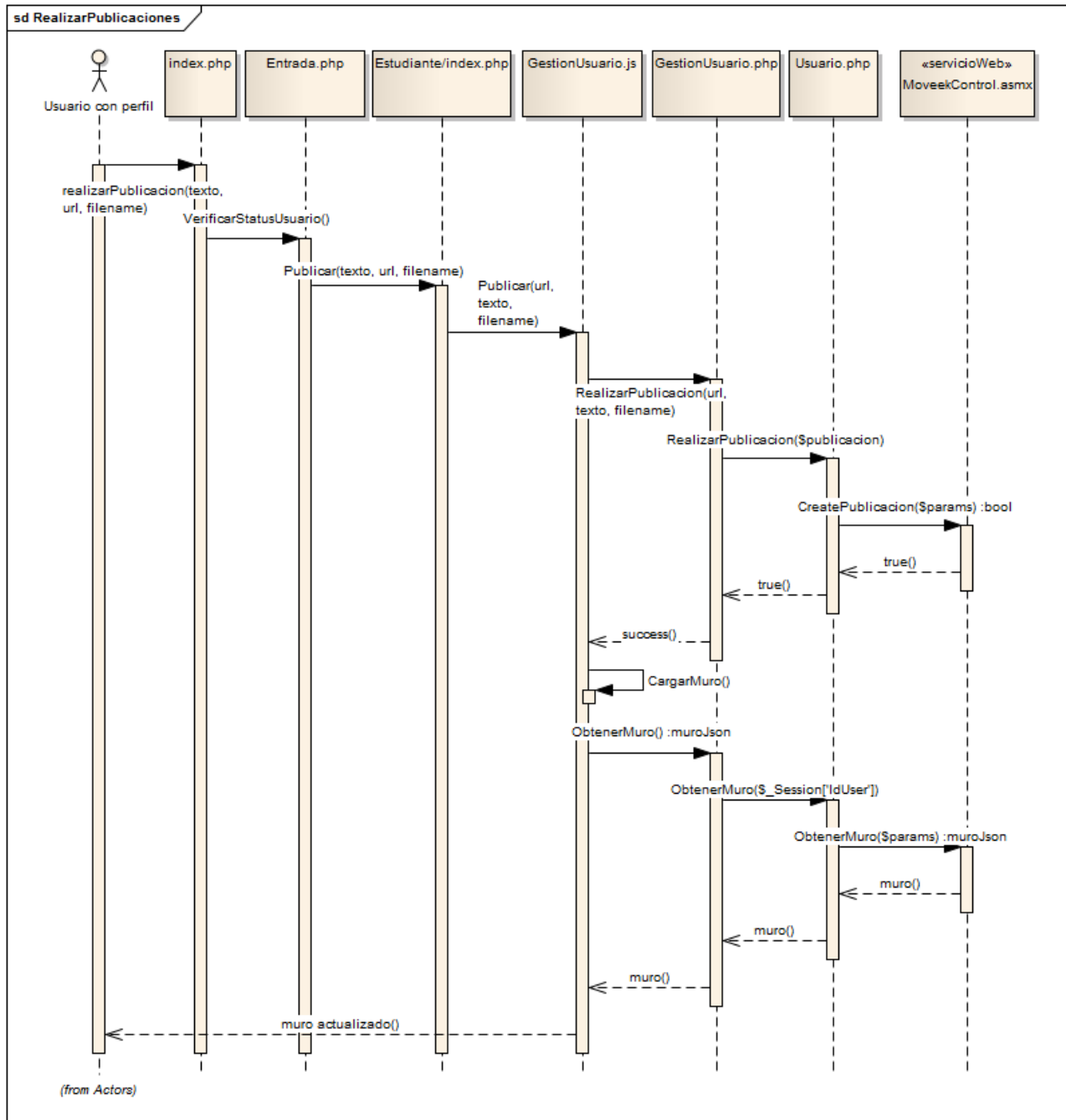


Figura 4.6: Diagrama de secuencia del caso de uso *Realizar Publicaciones*

Las tareas referentes a la *Administración de Relaciones* son las acciones de *Aceptar Amistad* y *Enviar la Solicitud de Amistad*.

En la Figura 4.7 se presenta el diagrama de secuencia de la acción de *Aceptar Amistad* y en la Figura 4.8 se presenta el diagrama de secuencia de la acción de *Enviar la Solicitud de Amistad*.

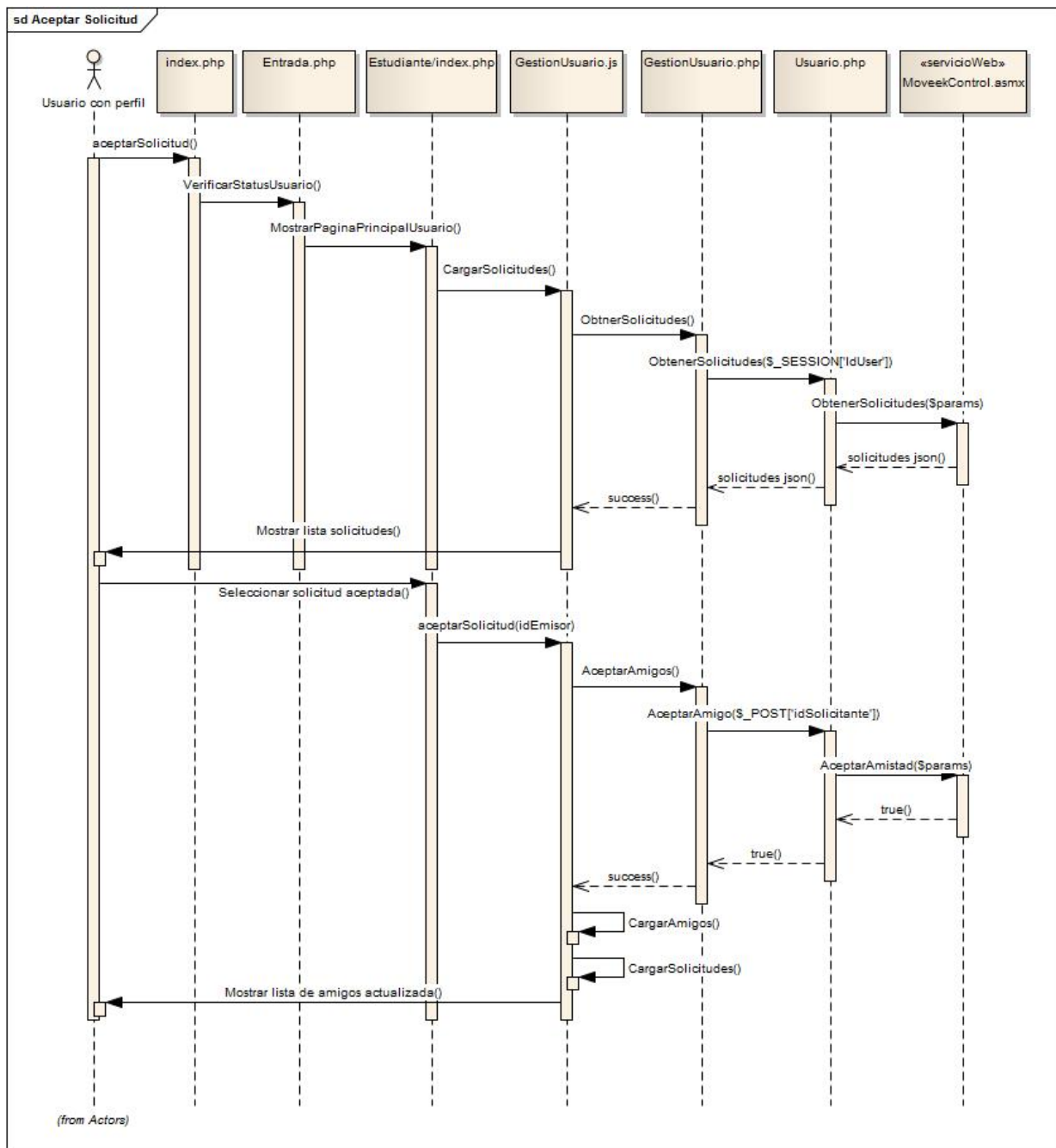


Figura 4.7: Diagrama de secuencia del caso de uso de aceptar amistad

Ahora que se han identificado los actores que participan en cada caso de uso, es momento de presentar el diagrama de clases de diseño que se obtuvo para la codificación de la red. En la Figura 4.9 se presenta el diagrama de clases referente a las clases que se necesitan codificar tanto para los servicios Web, como para los programas del lado del servidor y scripts del lado del cliente.

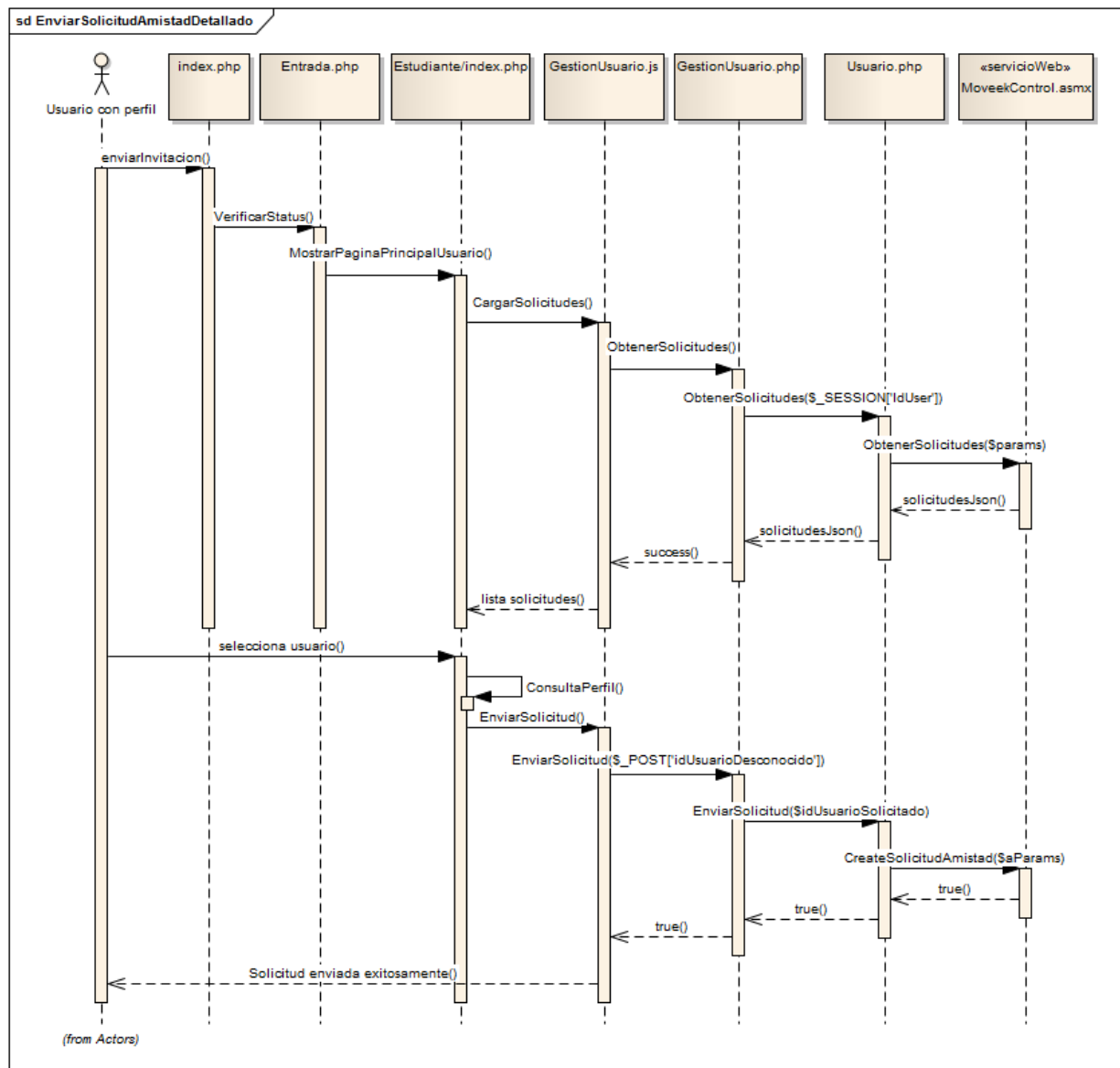


Figura 4.8: Diagrama de Secuencia del Caso de Uso de Enviar Solicitud

4.2. Servicios Web

Hemos mencionado que nuestra red social está usando el paradigma orientado a servicios, es por ello que se ha decidido retomar en este momento el marco teórico necesario para describir la colaboración presente entre los servicios Web que hemos propuesto en el apartado anterior (MoveekControl.asmx.cs, OperacionesProtege.java, ServiciosAnotacion.java).

4.2.1. ¿Por qué usar composición de Servicios Web?

Por la naturaleza de nuestro desarrollo, los scripts PHP del lado del servidor solo interactúan con el servicio Web MoveekControl.asmx.cs, sin embargo, los APIs de Protégé (para las consultas semánticas a la ontología) y Gate (para las tareas de anotación semántica) hacen uso de Java, por

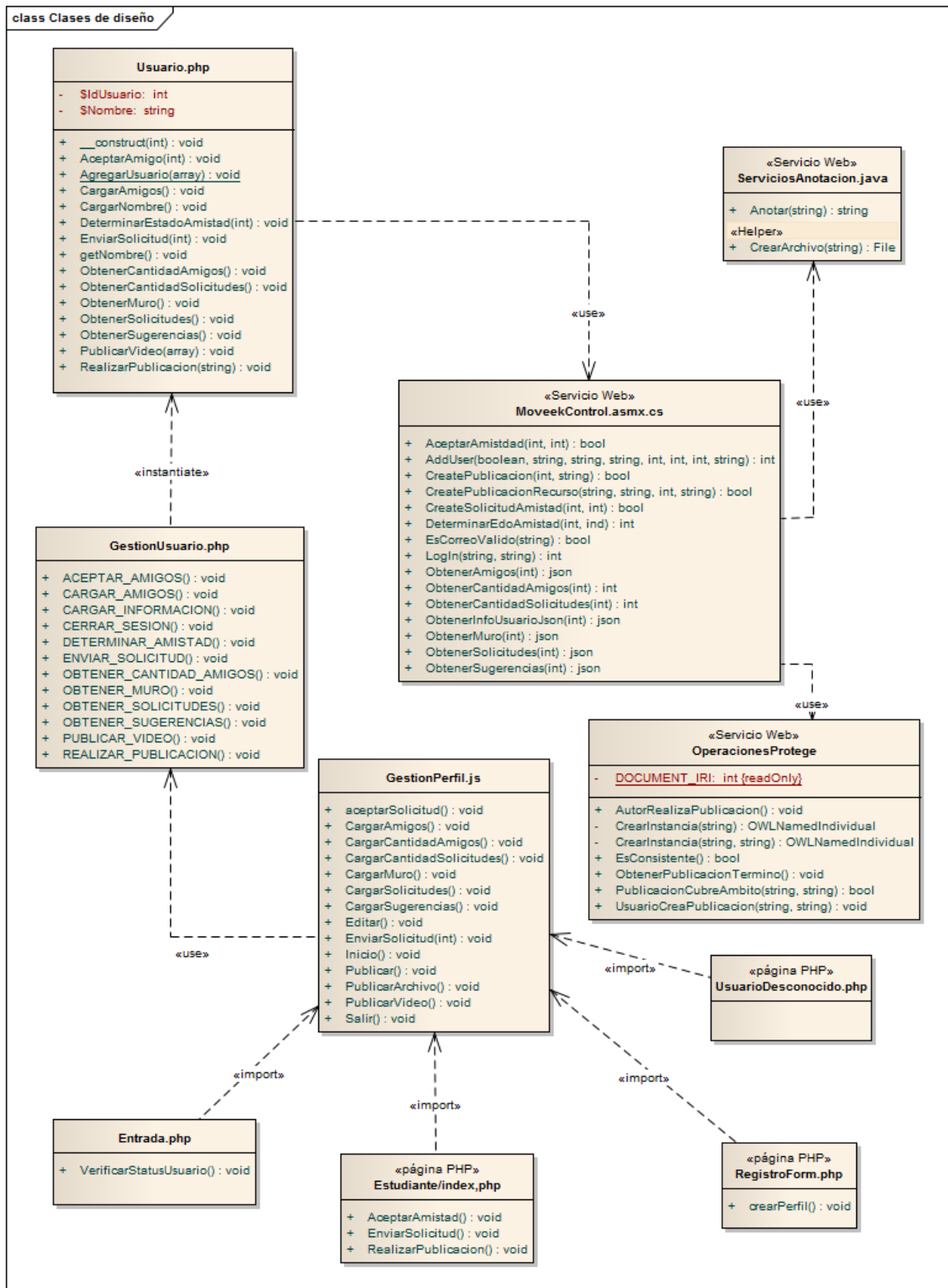


Figura 4.9: Diagrama de Clases de Diseño de Moveek.

lo cual es necesario hacer interoperables los servicios Web. La forma que hemos seleccionado para conseguir esta interoperabilidad es la combinación de servicios Web. Por un lado se tienen los servicios Web programados en C# y por otro lado se deben ofrecer como servicio las operaciones de la anotación semántica y de las consultas semánticas. Es por ello que a continuación se presenta el marco teórico que respalda nuestra selección de combinación.

4.2.2. Descubriendo el conocimiento: la combinación de servicios Web.

La combinación de servicios Web para la implementación de procesos de alto nivel, requiere de diversos estándares que nos permitan modelar las posibles interacciones entre los servicios. Los métodos actuales de creación de procesos de negocio no están diseñados para trabajar con procesos que involucren componentes de distintas organizaciones, ya que esto conllevaría a una coordinación extra a la que ordinalmente ofrecen. En esta sección introduciremos la combinación de servicios como la forma en la que realizamos el descubrimiento del conocimiento inicialmente planteado. Así pues, introducimos los términos de orquestación y coreografía, que tratan de describir aspectos relacionados con la creación de procesos de negocio que involucren varios servicios Web. La combinación de servicios Web para la implementación de procesos de alto nivel, requiere de diversos estándares que nos permitan modelar las posibles interacciones entre los servicios. Los métodos actuales de creación de procesos de negocio no están diseñados para trabajar con procesos que involucren componentes de distintas organizaciones, ya que esto conllevaría a una coordinación extra a la que ordinalmente ofrecen. Así pues, introducimos los términos de orquestación y coreografía, que tratan de describir aspectos relacionados con la creación de procesos de negocio que involucren varios servicios Web. Ver Figura 4.10

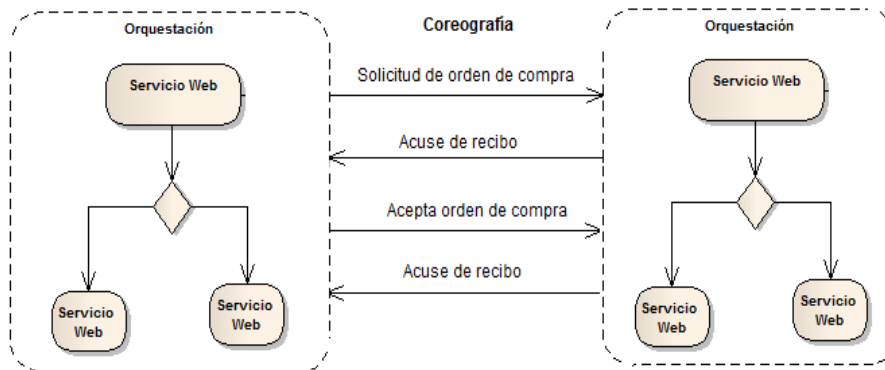


Figura 4.10: Relación entre orquestación y coreografía de Servicios Web

La orquestación, permite diseñar procesos de negocio ejecutables que puede interactuar (a nivel de mensaje) tanto con servicios Web internos como externos. Con la orquestación, siempre representamos el control del proceso de negocio desde el punto de vista de una de las partes participantes que intervienen en el mismo. Sin embargo, la coreografía es mucho más colaborativa, ya que permite trazar las secuencias de mensajes que se suceden entre todas las partes participantes del proceso de negocio en lugar de centrarse en los mensajes que intercambian los servicios Web que implementan los procesos de negocio ejecutados únicamente por una parte. Con lo anterior el tipo de combinación será la coreografía, ya que se necesita un alto nivel de

colaboración entre nuestros servicios Web [PEL03].

A continuación presentamos un diagrama que presenta los tres servicios Web que se combinarán para realizar las tareas de la red social, la anotación semántica y la consulta semántica. La Figura 4.11 muestra la relación de los servicios Web.

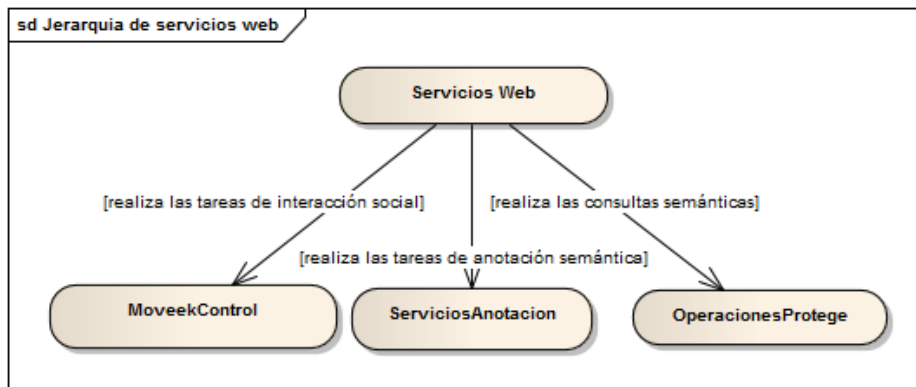


Figura 4.11: Relación de los servicios Web.

Ahora se detallarán gráficamente las operaciones de estos servicios que involucran la invocación a alguno de los otros. El servicio que interactúa con la red social es MoveekControl ya que este proporciona las operaciones que hacen posible la interacción social entre usuarios, es por ello que se presentan las operaciones que van a invocar a alguno de los otros servicios.

La Figura 4.12 muestra la representación gráfica de la operación CreatePublicacion del servicio MoveekControl que invoca a la operación Anotar del servicio ServiciosAnotacion.

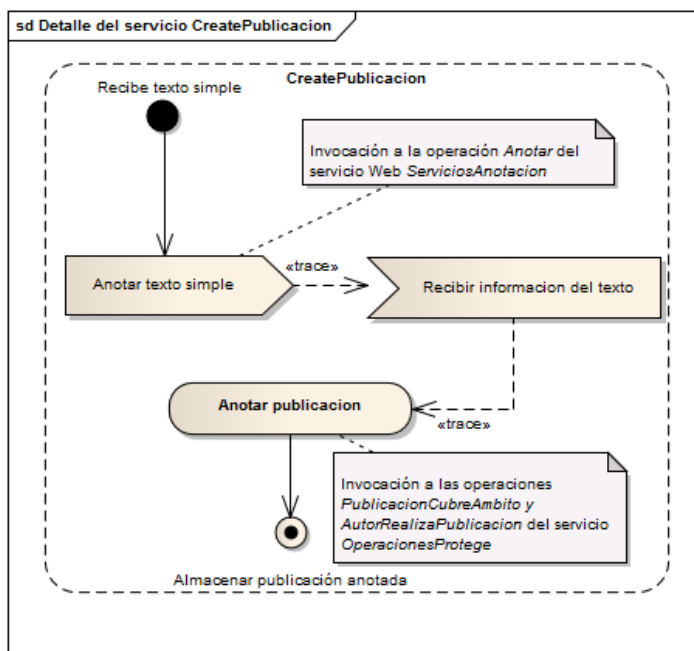


Figura 4.12: Diagrama de la operación CreatePublicacion

En la Figura 4.13 se muestra el detalle de la operación encargada de realizar las tareas de anotación semántica. Se debe mencionar que la información extraída del texto simple es invocando una aplicación previamente desarrollada en el entorno GATE ⁵ que devuelve un documento con formato XML que contiene la información semántica del texto. Esta información contiene, principalmente, las menciones a términos contenidos en la ontología *OntoScience*, lo cual le permite comprender a la computadora los términos que contiene cada publicación y de esta forma se pueden realizar, posteriormente, las consultas semánticas a la ontología.

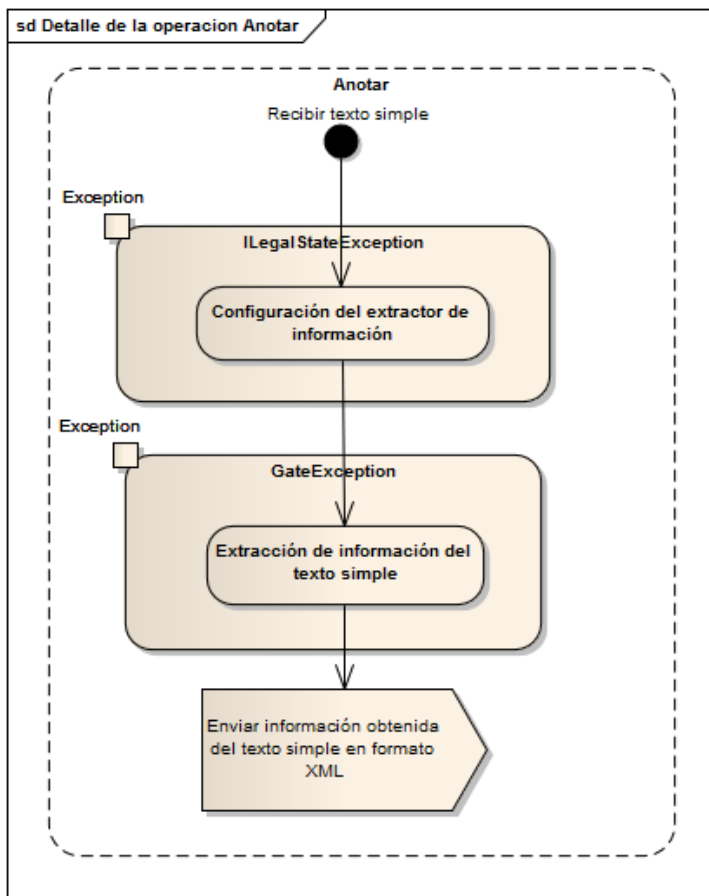


Figura 4.13: Diagrama de la operación Anotar del servicio Web ServiciosAnotacion

La operación *PublicacionCubreAmbito* del servicio *OperacionesProtege* es detallada en la Figura 4.14. Esta operación permite “decirle” explícitamente a la ontología que una publicación cubre o se trata de un ámbito en específico.

La Figura 4.15 muestra el diagrama de la operación *AutorRealizaPublicacion* que le “dice” a la ontología que determinado autor realiza cierta publicación. Esto da lugar a que las consultas semánticas tengan sentido desde el momento en que se pueda inferir qué otros temas se le pueden sugerir al usuario a través del razonador de *Protégé*.

⁵GATE: General Architecture for Text Engineering, se detalla en el Apéndice D

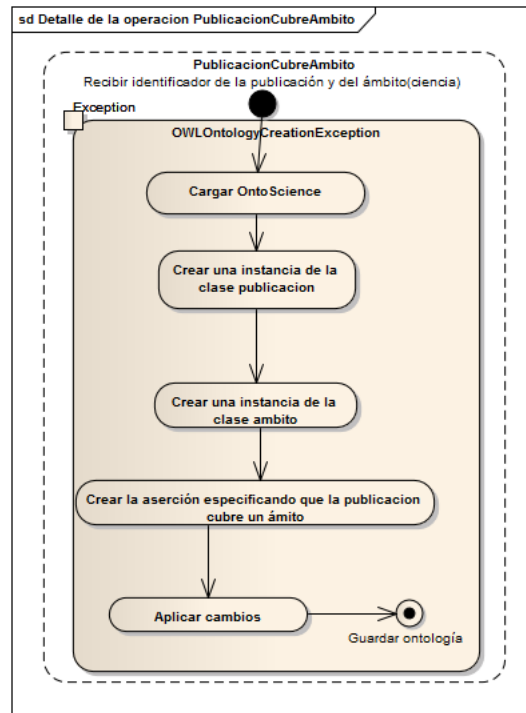


Figura 4.14: Diagrama de la operación *PublicacionCubreAmbito*

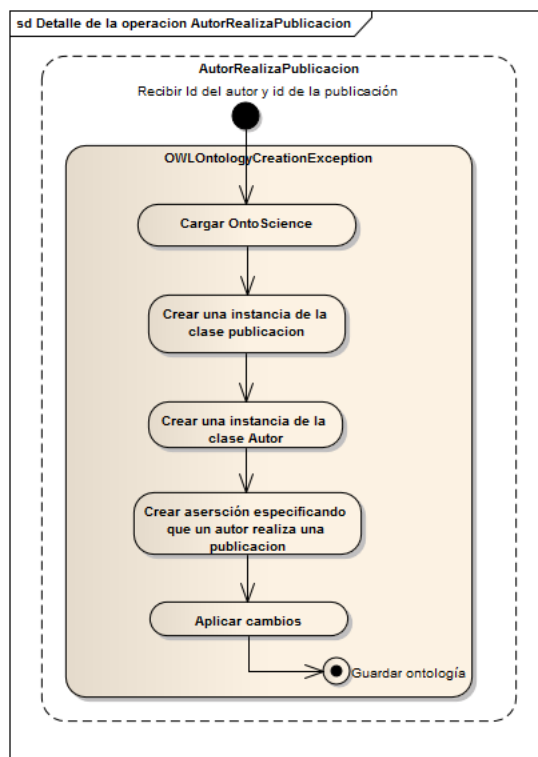


Figura 4.15: Diagrama de la operación *AutorRealizaPublicacion*

Otra operación crítica del servicio MoveekControl es la encargada de obtener la lista de publicaciones que se desplegará al usuario de la red. Esta operación consiste en la invocación de una operación del servicio Web OperacionesProtege. En la Figura 4.16 se muestra el diagrama de este servicio.

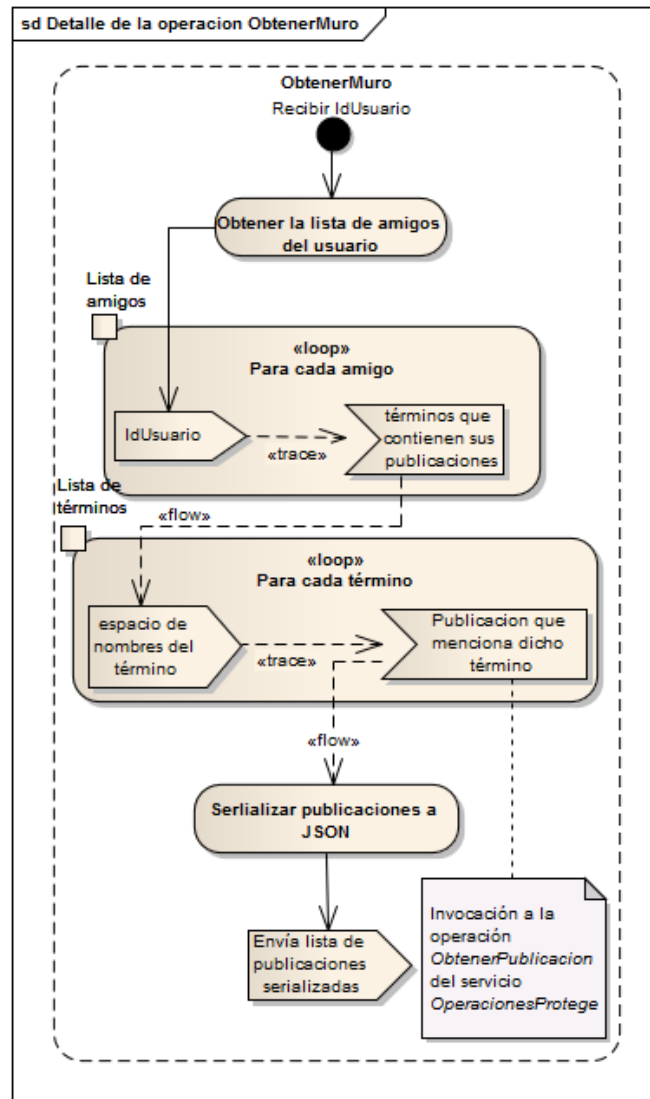


Figura 4.16: Diagrama de la operación *ObtenerMuro*

En la Figura 4.17 se muestra el detalle de la operación que realiza directamente la consulta semántica a la ontología a través del API de Protégé.

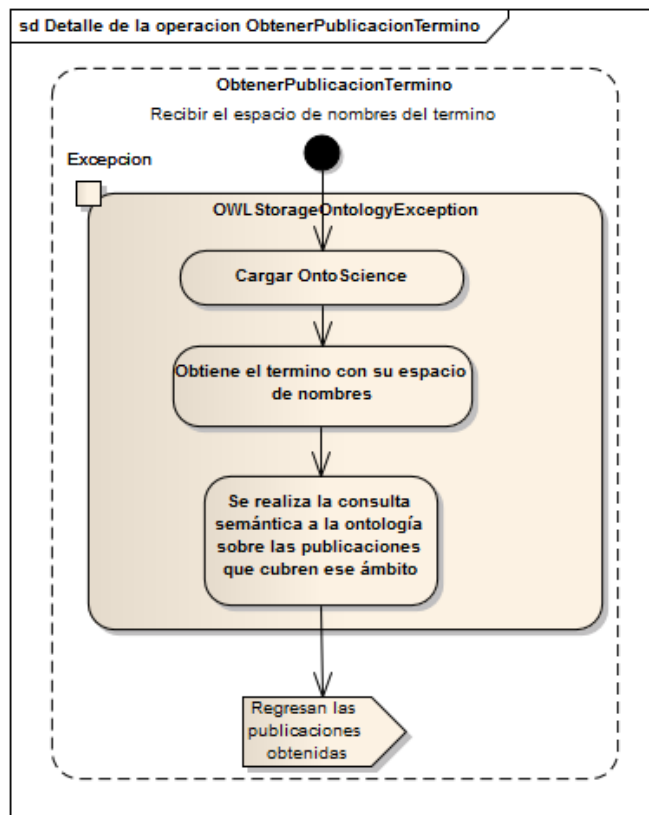


Figura 4.17: Diagrama de la operación *ObtenerPublicacionTermino*

Ahora que ya se ha presentado el detalle de cada servicio es momento de presentar la combinación que proponemos para realizar la colaboración entre los tres servicios y otorgarle la funcionalidad deseada a Moveek. La Figura 4.18 presenta la relación entre cada uno de los servicios y su colaboración en las operaciones *CreatePublicación* y *ObtenerMuro* del servicio *MoveekControl*, el cual es la interfaz que se presenta al cliente PHP que realiza las tareas de la red social.

La tecnología de los servicios Web permite a las aplicaciones dialogar a distancia vía Internet, independientemente de las plataformas y de los lenguajes utilizados en su desarrollo, éstos servicios Web se apoyan en los protocolos estándar (WSDL, SOAP, UDDI). Los servicios Web, son da vez ms utilizados en el mundo industrial, y por obvias razones presentan algunos problemas, como por ejemplo, ¿Cómo componer estos servicios?.

La mayor parte de las técnicas de composición de servicios Web, necesitan “razonar” sobre estos servicios. Como la descripción sintáctica de un servicio con WSDL no es suficiente, se debe agregar una descripción semántica OWL-S.

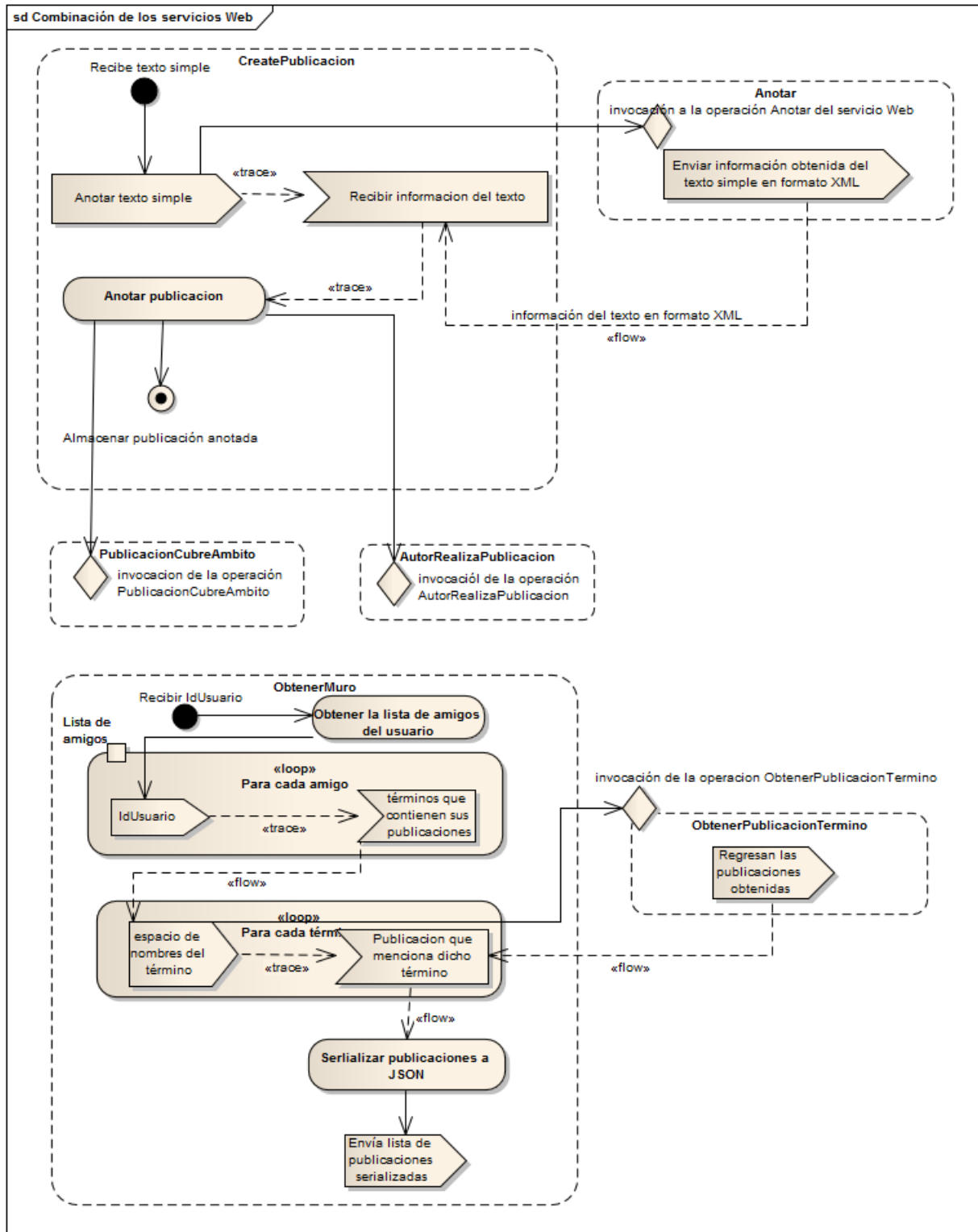


Figura 4.18: Colaboración de los tres servicios Web

Sabemos que existen dos grandes áreas de trabajo. Algunas de ellas proponen la composición de servicios con técnicas de ingeniería de software, es decir la orquestación y la coreografía, técnicas que adoptamos en este trabajo. Otras técnicas utilizan un enfoque más dinámico: la composición de servicios en este caso, utiliza técnicas de inteligencia artificial (cálculo de situaciones, sistemas multiagente o planificación).

4.3. Conclusión

En este último capítulo hemos mencionado las técnicas utilizadas para implementar los mecanismos de la anotación semántica y las operaciones con la ontología. Podemos concluir que el uso de servicios Web hace más fácil la implementación de las funciones de la red social, además que, debido a la interoperabilidad que proporcionan los servicios Web, se pudo hacer uso de la combinación de los servicios Web programados en diferentes tecnologías como se mencionó en el capítulo.

A continuación se presentan algunos detalles de la red social que se desarrolló para el propósito más importante de trabajo de tesis, la obtención de conocimientos.

Por lo tanto, se presentan algunos resultados obtenidos en la red social Moveek, como se podrá apreciar la red social es funcionalmente estable y en esta parte, el lector entenderá mejor la forma en la que se realizan los procesos de la anotación semántica y respectivamente las consultas a la ontología, todo ello a través de funcionalidades propias de la red social.

Capítulo 5

Resultados

En ese capítulo se presentan los resultados de las funcionalidades más importantes de nuestra red social. En primer término se presentan imágenes de la red social y después presentaremos los resultados de dos consultas semánticas realizadas de forma implícita a través de la red social. La red social está alojada en la siguiente dirección: <http://climate.cs.buap.mx/Moveek>. Está disponible para que cualquier persona cree su perfil en nuestra red social y comenzar a compartir su conocimiento.

5.1. Funciones de la red social

En la Figura 5.1 se presenta la imagen de la pantalla principal del usuario que no se ha autenticado o no tiene cuenta de usuario en Moveek.

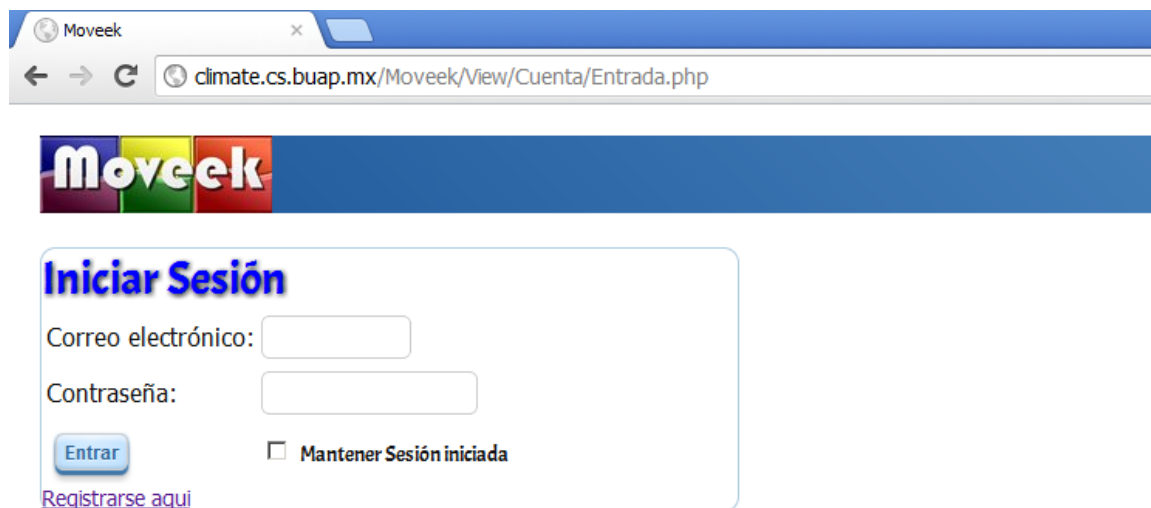


Figura 5.1: Pantalla de entrada a Moveek

A partir de este punto el usuario puede autenticarse, y en caso de no tener perfil, puede crear su cuenta. En la Figura 5.2 se presenta el formulario de registro para nuestra red. Se puede observar que no son muchos los valores necesarios para la crear el perfil de usuario.

Formulario de Registro

climate.cs.buap.mx/Moveek/View/Cuenta/RegistroForm.php

Moveek

Datos del usuario

Nombre:

Correo electrónico:

Repetir correo electrónico:

Fecha de nacimiento (dd/mm/aaaa): - -

Sexo Hombre Mujer

Contraseña:

Repetir la contraseña:

Selecciona la imagen de su perfil:

[Iniciar Sesión](#)

Figura 5.2: Formulario de registro de Moveek.

Moveek

Pablo Camarillo

Jsted tiene 2 amigos

¿Qué deseas compartir?

Publicaciones **Ligas** **Documentos**

Compartir

Usuarios que desean unirse a tu red

Los efectos del [plomo](#), son la presencia de retraso en el desarrollo, problemas de aprendizaje, trastornos en la [conducta](#), alteraciones del lenguaje y de la capacidad auditiva, anemia, vómito y dolor abdominal recurrente
Publicado el 17/07/2012 09:34:47 a.m.

se entiende por [consenso](#) de Washington un listado de políticas económicas
Publicado el 17/07/2012 09:32:33 a.m.

El método de fechamiento o determinación de edades de objetos antiguos, de origen humano, animal o vegetal es una técnica llamada prueba de [Carbono 14](#)
Publicado el 17/07/2012 09:29:51 a.m.

La [negociacion](#) es un esfuerzo de interacción que se realiza a fin de generar beneficios entre los miembros de esta
Publicado el 17/07/2012 09:16:54 a.m.

Lo maravilloso de la [guerra](#) es que cada jefe de asesinos hace bendecir sus banderas e invocar solemnemente a Dios antes de lanzarse a exterminar a su enemigo
Publicado el 17/07/2012 09:09:52 a.m.

Usuarios que podrian unirse a tu red

Figura 5.3: Pantalla principal del perfil de Moveek

Cuando el usuario se ha autenticado correctamente se presenta directamente la página principal del perfil. (Ver Figura 5.3)

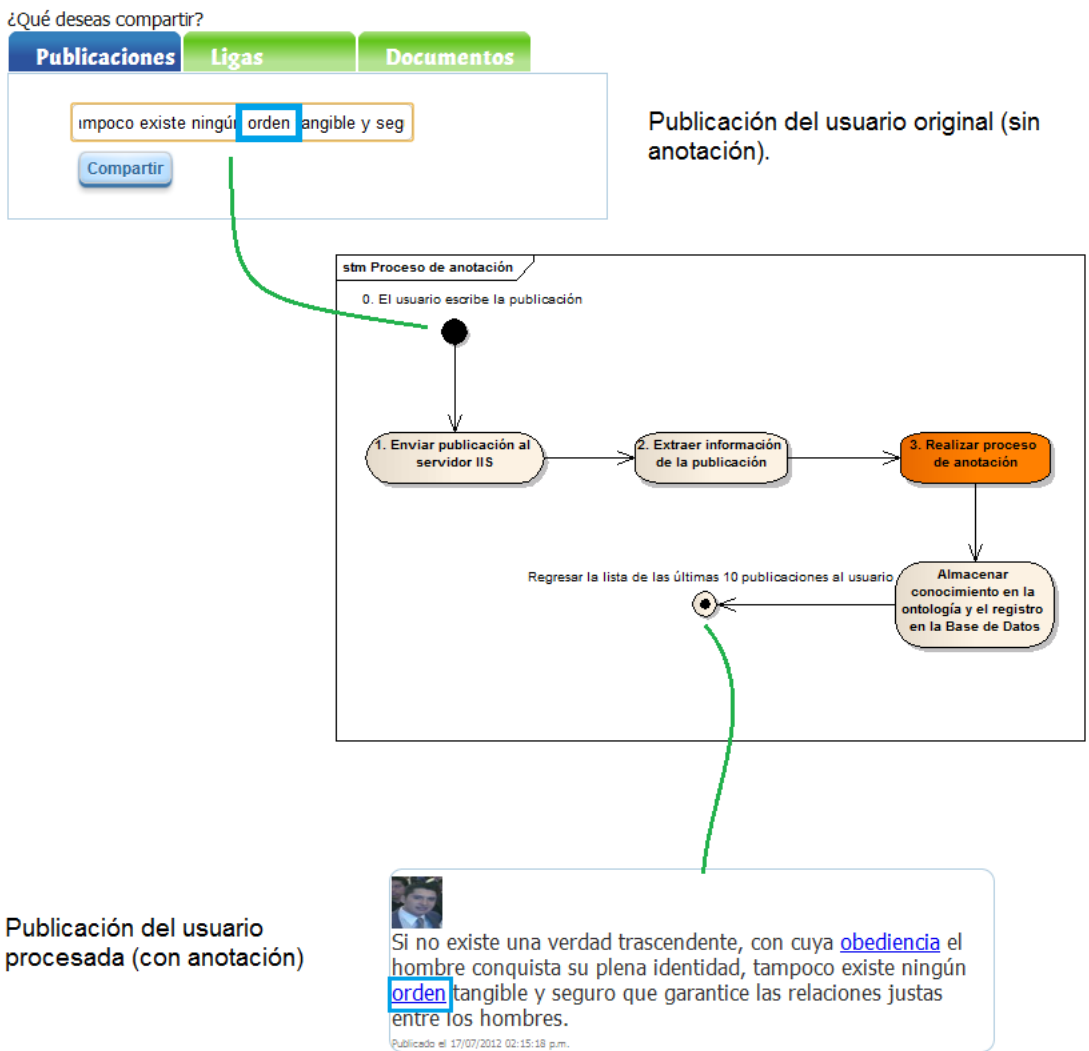


Figura 5.4: Proceso de publicación en Moveek.

Para ilustrar la función de la publicación, la cual conlleva la función de la anotación semántica, presentamos un diagrama de estados que presenta el momento en que una publicación realiza el proceso de anotación semántica sobre cada publicación. En la Figura 5.4 se presenta dicho esquema.

5.2. Consultas semánticas

Una vez presentadas las imágenes de las funciones principales de la red social ahora se presentan dos consultas semánticas a las publicaciones mostradas en la Figura 5.3. Primero mostramos un esquema donde muestra los estados que toma el proceso de consulta. La Figura 5.5 muestra las etapas de este proceso.

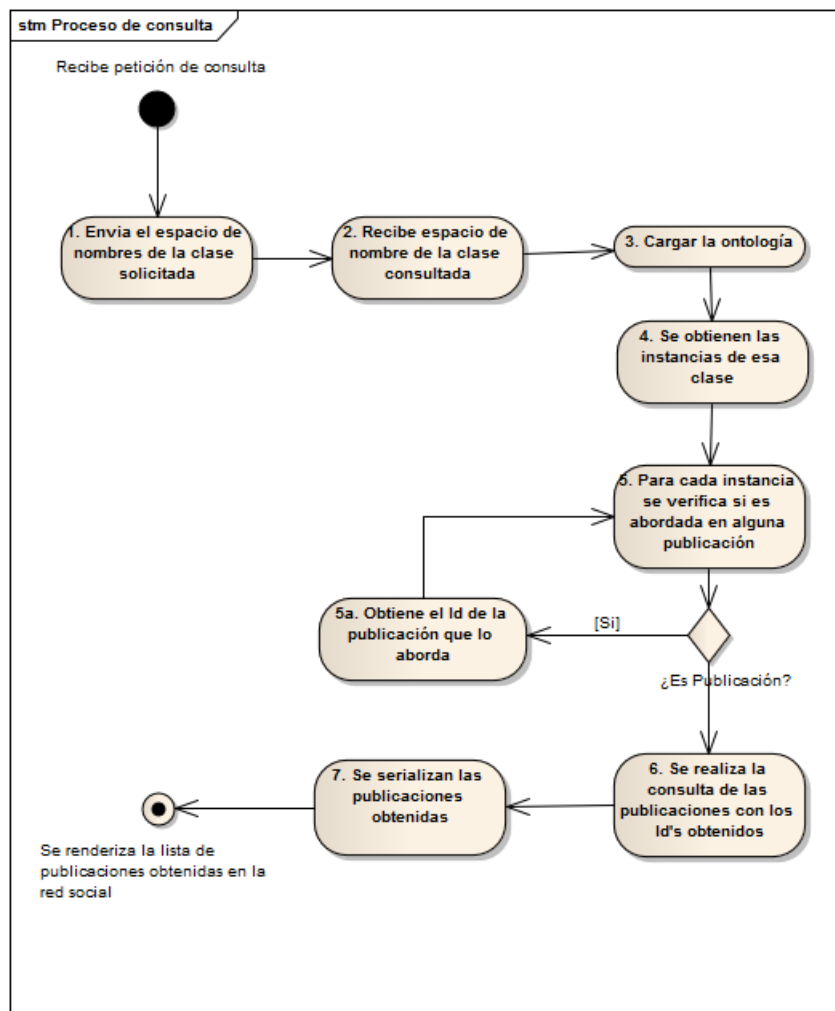


Figura 5.5: Proceso de la consulta semántica.

Una vez que hemos mencionado las etapas del proceso de la consulta a la ontología, es momento de detallar cada paso. Por un lado detallaremos algunas de las etapas ejemplificando una consulta a través del editor de la ontología (Protégé) para después mostrar el resultado de dos consultas a la ontología en la red social.

1. **Envío del espacio de nombre o identificador en la ontología de la clase que se desea consultar.** El envío del espacio es a través de un script PHP mediante el método GET hacia el servicio Web OperacionesProtege.

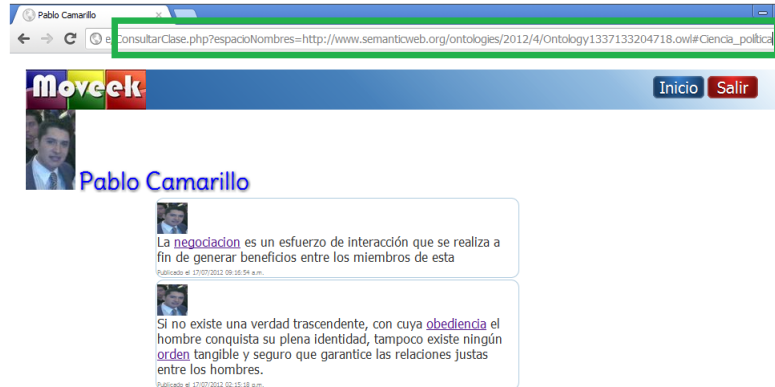


Figura 5.6: Ejemplo de la petición por el método GET de la clase consultada.

2. **Recepción del espacio de nombres o identificador de la clase.** El Servicio Web OperacionesProtege recibe el identificador de la clase a través de una recepción SOAP.

SOAP Request

```
<?xml version="1.0" encoding="UTF-8"?>
<S:Envelope xmlns:S="http://schemas.xmlsoap.org/soap/envelope/">
  <S:Header/>
  <S:Body>
    <ns2:ObtenerPublicacionesTermino xmlns:ns2="http://gestionOntologia/">
      <IriCiencia>http://www.semanticweb.org/ontologies/2012/4/Ontology1337133204718.owl#Ciencia_política</IriCiencia>
    </ns2:ObtenerPublicacionesTermino>
  </S:Body>
</S:Envelope>
```

Figura 5.7: Ejemplo de la petición SOAP Request del servicio Web de la consulta semántica.

3. **Se Carga la ontología.** Se ilustra esta etapa mostrando la ontología cargada en el editor de ontologías *Protégé*.
4. **Se obtienen las instancias de esa clase.** Una vez cargada la ontología se hace la consulta a la ontología utilizando las propiedades de objeto. En este caso se usa la propiedad *EsAbordadaEn*, la cual nos devuelve la lista de instancias (y las publicaciones) que abordan la clase consultada. En la Figura 5.9 se muestra el resultado de la consulta realizada a la ontología. Esta consulta se podría traducir como “*Dame todas las instancias de la clase **Ciencia_política** donde cada instancia **EsAbordadaEn** alguna **Publicación***”. Dicha consulta regresa la lista de instancias de la clase *Ciencia_política* que contienen la propiedad *EsAbordadaEn* y las relaciona con alguna *Publicación*.

Una vez obtenidas las instancias de la clase *Ciencia_política* se navega a través de cada instancia obtenida y se explora la publicación que aborda dicha clase.

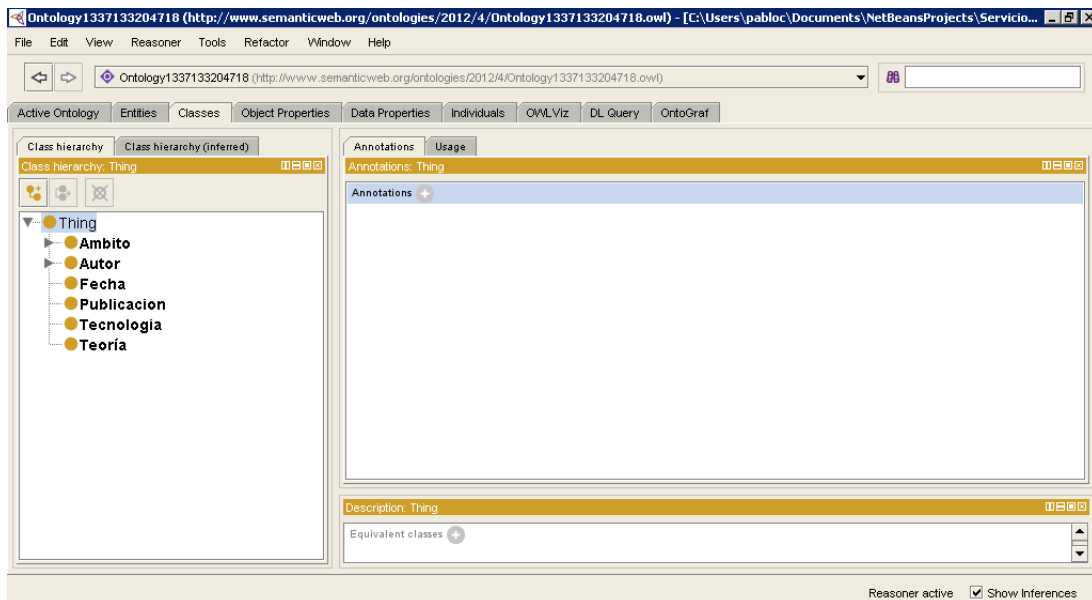


Figura 5.8: Ontología cargada en el editor Protégé.

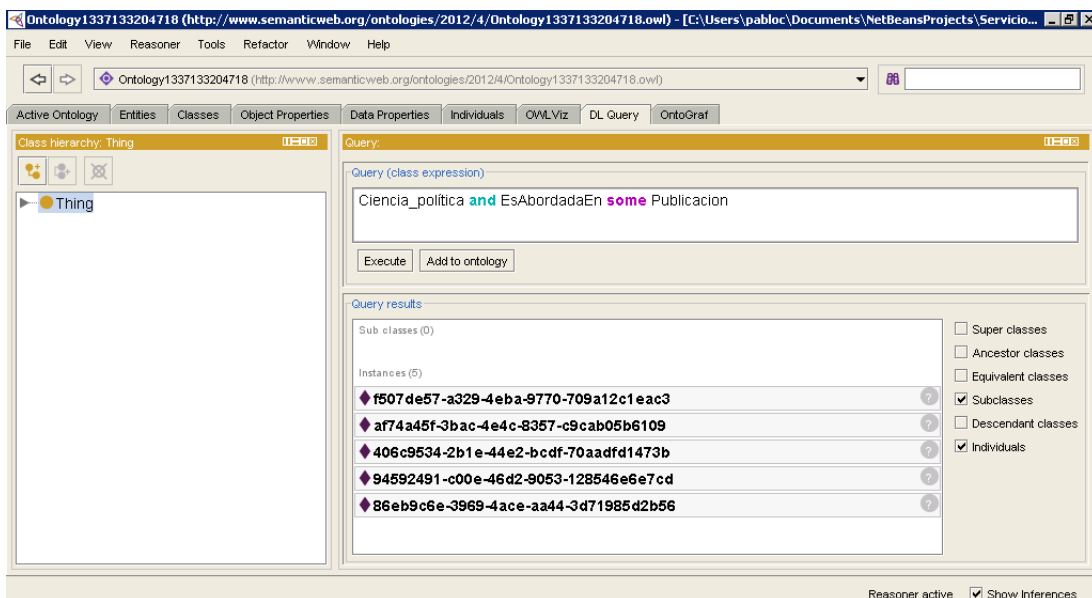


Figura 5.9: Consulta semántica en Protégé.

5. Se verifica que cada instancia si es abordada en alguna publicación. En caso de ser abordada en alguna publicación se obtiene el ID de dicha publicación. En la Figura 5.10 se ejemplifica este proceso en una instancia en particular, se señala el Id (identificador) de la publicación que aborda esta clase en el recuadro verde.
6. Se serializa y se envía la lista de publicaciones que abordan a la clase consultada. Al estar realizando esta consulta a través de servicios Web que usan el estándar SOAP, es obvio que la respuesta como tal debe ser un sobre SOAP Response.

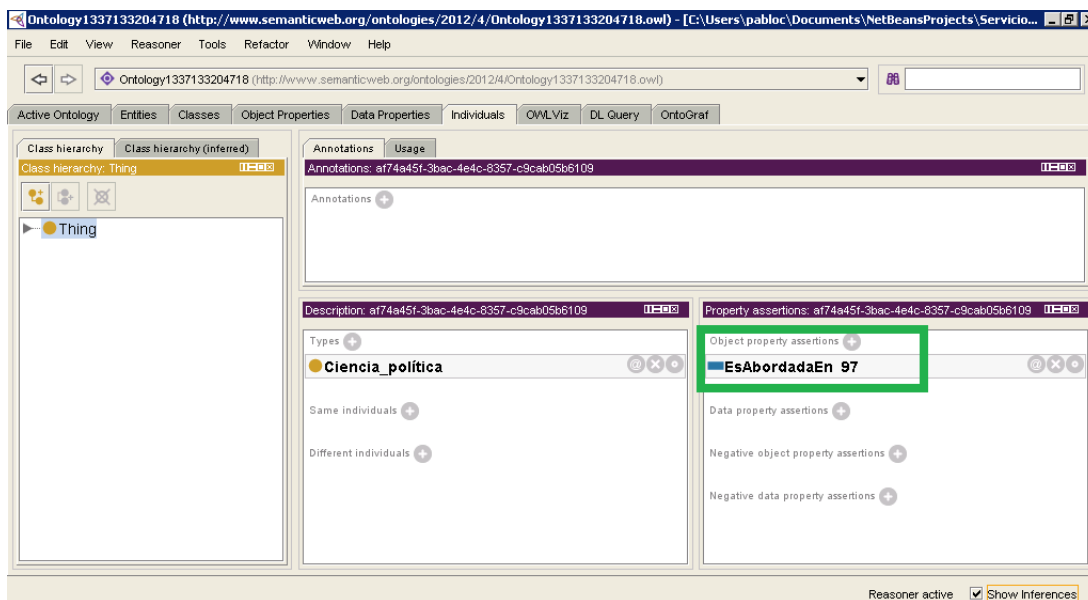


Figura 5.10: Navegación por una instancia devuelta por la consulta semántica.

Una vez descritas las etapas de la consulta semántica, su resultado visualizado en Protégé y su respuesta en formato SOAP estamos en condiciones de presentar dos consultas, pero ahora observando su resultado ya en la red social. En la Figura 5.11 se observan 6 publicaciones, donde cada una contiene uno o más términos anotados. Cuando el usuario presiona la liga, la red social muestra el resultado de la consulta semántica del término presionado. En la Figura 5.11 se muestra el resultado de la consulta semántica a la clase Química.

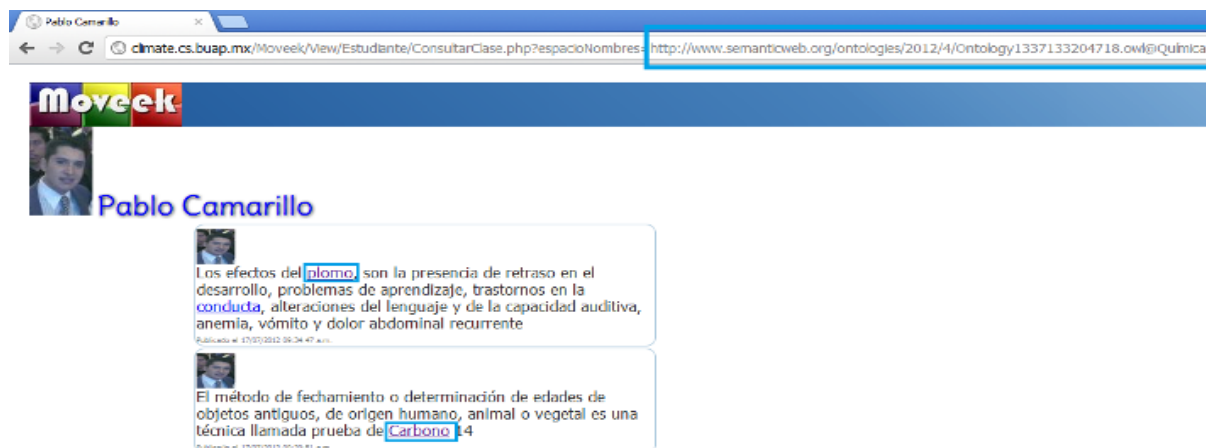


Figura 5.11: Resultado de la consulta semántica a la clase Química de la ontología.

Finalmente se presenta el resultado de la consulta semántica a la clase *Ciencia_política* de la ontología, resultado de haber presionado un término referente a esta clase. La Figura 5.12 presenta el resultado de esta consulta.



Figura 5.12: Resultado de la consulta semántica a la clase *Ciencia_política* de la ontología.

5.3. Conclusión

Los resultados obtenidos en la red social muestran que las funciones principales de Moveek funcionan correctamente. Se entiende, en este contexto, como funciones principales a las acciones de realizar publicaciones y consultar términos.

Si bien en este apartado mostramos que las funciones principales de la red social funcionan conforme lo planeado en los objetivos principales de este trabajo de tesis, aún se deben evaluar los efectos que van a tener en la experiencia con el usuario de la red social: ¿Va a ser útil la consulta semántica proporcionada por Moveek al usuario?, ¿Cuál será la ciencia o disciplina más mencionada por los usuarios de Moveek?. Todas estas cuestionamientos y otros serán resueltos en los trabajos futuros de este proyecto.

Capítulo 6

Conclusiones y trabajo futuro

El estudio y desarrollo de la Web Semántica incluye diversos ámbitos que van desde el modelado correcto de una aplicación Web hasta la elaboración de una buena representación del conocimiento y la relación que deben tener ambas para hacer la Web más humana. En este trabajo de tesis involucramos todos estos aspectos referentes a la Web Semántica con el objetivo de entender, pero sobre todo de desarrollar una aplicación que reflejará las ventajas (y desventajas) de la Web Semántica. A lo largo del desarrollo de los trabajos de investigación y desarrollo de nuestra red social nos hemos dado cuenta que, si bien la Web Semántica cumple su cometido de hacer más eficientes las consultas de información sobre el contenido de la Web, también conlleva ciertas desventajas, sobre todo las referentes a la infraestructura necesaria para no afectar el rendimiento de las aplicaciones Web. Otro aspecto interesante que podemos resaltar de este trabajo es que la creciente presencia de recursos multimedia en la Web hacen más desafiante el desarrollo de la Web Semántica.

Con lo último podemos introducir el trabajo futuro de este proyecto. En primer lugar debemos de perfeccionar el modelado de la red social para poder escalar la aplicación y pensar en más funcionalidades tales como la gestión de cursos para crear redes de colaboración científica en nuestra red social. Otro punto crucial de nuestro trabajo futuro es nuestro interés de automatizar el proceso de aprendizaje de nuestra ontología y de esta forma enriquecer el dominio científico que ésta actualmente cubre.

Con lo anterior podemos decir que tenemos muchas áreas de oportunidad para mejorar y evolucionar nuestro trabajo pero también sabemos que hemos avanzado mucho en el estudio de la Web Semántica. Deseamos que nuestro trabajo también sea una inspiración para las futuras generaciones que deseen involucrarse en el desarrollo de la Web Semántica, es por ello que en este trabajo se presentan los contratos de servicio en los cuales esta basada nuestra red social en caso de que el lector se interese en participar en la evolución de esta aplicación a apartir de los servicios que ofrecemos.

Bibliografía

- [AB06] BEN ADIDA and MARK BIRBECK. Rdfa primer 1.0., 2006.
- [ACKM04] GUSTAVO ALONSO, FABIO CASATI, HARUMI KUNO, and VIJAY MACHIRAJU. *Web Services: Concepts, Architectures and Applications*, volume 1. Springer-Verlag, 2004.
- [ADR06] SOREN AUER, SEBASTIAN DIETZOLD, and THOMAS RIECHERT. Ontowiki a tool for social. In Springer-Verlag, editor, *Proceedings of the Fifth International Semantic*, pages 736–749, 2006.
- [Alv10] Rubén Darío Alvarado. *Metodología para el desarrollo de ontologías*, 2010.
- [AMA07] FLORENCE AMARDEILH. *Web Sémantique et Informatique Linguistique : propositions méthodologiques et réalisation d'une plateforme logicielle*. PhD thesis, UNIVERSITE PARIS X NANTERRE, Paris, Francia, Mayo 2007.
- [AS00] N. AUSSENAC and P. SEGUELA. Les relations sémantiques : du linguistique au formel. *Cahiers de grammaire*, 25(Numéro spécial «Sémantique et Corpus »):175–198, 2000.
- [BAGC04] D. BOURIGAULT, N. AUSSENAC-GILLES, and J. CHARLET. *Techniques Informatiques et Structuration de Terminologies*, volume 18 of *Numéro Spécial de la Revue d'Intelligence Artificielle (RIA)*, chapter Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas, pages 87–110. PIERREL J.-M. ET SLODZIAN M., Paris, 2004.
- [BAT97] W.N. BORST, J.M. AKKERMANS, and J.L. TOP. Engineering ontologies. *International Journal of Human-Computer Studies*, pages 365–406, 1997.
- [BCEH04] J-F. BAGET, E. CANAUD, J. EUZENAT, and M.-S. HACID. *Le Web sémantique*, volume 4 of *Hors série de la Revue Information - Interaction - Intelligence*, chapter Les langages du Web Sémantique, pages 21–43. Cépaduès, Toulouse, 2004.
- [BER03] AMANN BERND. Le web sémantique:rdf et owl, 2003.
- [BHWR06] JEFFREY BOASE, JOHN B. HORRIGAN, BARRY WELLMAN, and LEE RAINIE. The strength of internet ties. Technical report, 2006.
- [BUN12] MARIO BUNGE. *La ciencia. Su método y su filosofía*. 2012.
- [CB03] J. CONTRERAS and R. BENJAMINS. Annotation tools and services. Délivrable 3.1, Esperonto Services Project, 2003.
- [CBT04] J. CHARLET, B. BACHIMONT, and R. TRONCY. *Le Web sémantique*, volume 4 of *Hors série de la Revue Information - Interaction - Intelligence*, chapter Ontologies pour le Web Sémantique, pages 69–100. Cépaduès, Toulouse, 2004.
- [CHA02] J. CHARLET. L'ingénierie des connaissances : développements, résultats et perspectives pour la gestion des connaissances médicales, habilitation à diriger des recherches,. Master's thesis, Université Paris VI, Paris, 2002.
- [CON09] WORLD WIDE WEB CONSORTIUM, 2009. <http://www.w3.org/TR/owl-ref/>.

- [COR06] O. CORCHO. Ontology based document annotation: trends and open research problems. *International Journal of Metadata, Semantics and Ontologies*, 1(1):47–57, 2006.
- [CSH02] F. CIRAVEGNA, S. STAAB, and S. HANDSCHUH. S-cream - semi-automatic creation of metadata. In LNCS 2473, editor, *Proceedings of the 13th International Conference on Knowledge Engineering and Management (EKAW'02)*, Madrid, España, 2002. Springer-Verlag.
- [EBU12] EBZZING, Mayo 2012. <http://labs.ebuzzing.es/top-blogs/ciencia>.
- [EUZ05] J. EUZENAT. *Ingénierie des connaissances*, chapter L'annotation formelle de documents en 8 questions, pages 251–271. TEULIER R. And CHARLET J And TCHOOUNIKINE P., Paris, 2005.
- [FEJ⁺02] CIRAVEGNA F., MOTTA E., DOMINGUE J., LANZONI M., STUTT A., and M. VARGAS-VERA. Mnm: Ontology driven semi-automatic and automatic support for semantic markup. In *Proceedings of the 13th International Conference on Knowledge Engineering and Management (EKAW02)*, pages 379–391, Madrid, España, 2002. Springer Verlag,.
- [FOX95] GRUNINGER M. FOX. Methodology for the design and evaluation of ontologies. In *AI Workshop on Basic Ontological Issues in Knowledge Sharing*, 1995.
- [FUC93] C. FUCHS. *Linguistique et Traitements automatiques des langues*. Hachette, Paris, 1993.
- [GAR04] SERGE GARLATTI. Le web sémantique, semantic web: The new internet of meanings. Reporte final, École National Supérieure des Télécommunications (ENST) de Bretagne, 2004.
- [GH06] JENNIFER GOLBECK and JAMES HENDLER. Filmtrust: Movie recommendations using trust in web-based social networks. In *Proceedings of the IEEE Consumer Communications and Networking Conference.*, 2006.
- [GP04] S. GARLATTI and Y. PRIE. *Le Web sémantique*, chapter Méta-données et annotations dans le Web sémantique, pages 45–68. CHARLET J., LAUBLET P. & REYNAUD C., Cépaduès, Toulouse, 2004.
- [GRU93a] T. GRUBER. A translation approach to portable ontology specifications. *Knowledge Acquisition Journal.*, 5(2):199–200, 1993.
- [GRU93b] TOM R. GRUBER. *Formal Ontology*, chapter Towards Principles for the Design of Ontologies Used for. Kluwer Academic Publishers, Deventer, Holanda, 1993.
- [GRU23] TOM GRUBER. <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>., Mayo 23. Citado el: 4 de Julio de 2012.
- [HAB05] B. HABERT. *Instruments et ressources électroniques pour le français.*. Collection "L'essentiel Français", Paris, 2005.
- [HAN05] S. HANDSCHUH. *Creating Ontology-based Metadata by Annotation for Semantic Web*. PhD thesis, Université de Karlsruhe, Karlsruhe, 2005.
- [HEN00] JAMES HENDLER. Agents and the semantic web, 2000.
- [HH01] JEFF HELFIN and JAMES HENDLER. A portrait of the semantic web in action. In *IEEE Intelligent Systems*, pages 54–59, 2001.
- [HOR09] MATTHEW HORRIDGE. Interface owlreasoner, Junio 2009.
- [LAR02] Alberto CHA LARRIEU. *Elementos de epistemología*. Trilce, Montevideo, 2002.
- [LAU07] P. LAUBLET. Web sémantique et ontologies. *Humanités Numériques*, 1(Nouvelles technologies cognitives et concepts des sciences humaines et sociales), 2007.
- [LEA03] KAHNEY LEANDER. *Making Friendsters in High Places*. Wired, 2003.
- [LHL01] TIM BERNERS LEE, JAMES HENDLER, and ORA LASILLA. The semantic web. *Scientific American*, 2001.

- [MAD⁺03] GORANOV M., KIRYAKOV A., MANOV D., KIRILOV A., OGNJANOFF D., and B. POPOV. Towards semantic web information extraction. In Sanibel, editor, *Proceedings of the Human Language Technologies Workshop (ISWC'03)*, pages 1–22, Florida, 2003.
- [MED12] MIGUEL ÁNGEL MORALES MEDINA, Mayo 2012. <http://gaussianos.com/>.
- [MIC10] PEACOCK MICHEL. *Introduction to social networks. PHP 5 Social Networking*. 2010.
- [MIK07] PETER MIKA. *Social Networks and the Semantic Web*. Springer, Madrid, España, 2007.
- [NIC11] JOSE MANUEL LÓPEZ NICOLÁS, Mayo 2011. <http://scientia1.wordpress.com/>.
- [PEL03] CHRIS PELTZ. Web services orchestration and coreography. *IEEE Computer Society*, pages 2–3, 2003.
- [PHA05] Quang T. T. PHAN. Ontologies et web services. Reporte Final 1, Institut de la Francophonie pour l'Informatique, Julio 2005.
- [PLC09] ASUNCIÓN GOMEZ PÉREZ, MARIANO FERNÁNDEZ LOPEZ, and OSCAR CORCHO. Ontological engineering. Springer, 2009.
- [PRI00] F. LE PRIOL. *Extraction et capitalisation automatique de connaissances à partir de documents textuels. SEEK-JAVA : identification et interprétation de relations entre concepts*,. PhD thesis, Université Paris IV-Sorbonne, Paris, 2000.
- [RAE12] Real academia española, 2012. www.rae.es.
- [RDH⁺03] F. RINALDI, J. DOWDALL, M. HESS, ELLMAN J., ZARRI G.-P., PERSIDIS A., BERNARD L., and KARANIKAS H. Multilayer annotations in parmenides. In Sanibel, editor, *Proceedings of the Knowledge Markup and Semantic Annotation Workshop*, pages 33–40, Florida, USA, 2003.
- [RMEK01] SWICK R., KOIVUNEN M.R., PRUD'HOMMEAUX E., and K. KAHAN. Annotea: An open rdf infrastructure for shared web annotations. In ACM Press, editor, *Proceedings of the 10th International World Wide Web Conference (WWW'01)*, pages 623–632, Hong-Kong, 2001.
- [SBF98] R. STUDER, V. R. BENJAMINS, and D. FENSEL. Knowledge engineering: principles and methods. *IEEE Transactions on Data and Knowledge Engineering*, 25(1 And 2):161–197, 1998.
- [SSC04] S. STAAB, HANDSCHUH S., and P. CIMIANO. Towards the self-annotating web. In ACM Press, editor, *Proceedings of the 13th International World Wide Web Conference (WWW'04)*, pages 462–471, New-York, USA, 2004.
- [Ste83] Ilya Prigogine e Isabelle Stengers. *La nueva alianza: metamorfosis de la ciencia*. Alianza Universidad, Madrid, 1983. ISBN 84-206-2368-7.
- [Uni12] Stanford University, Agosto 2012. <http://protege.stanford.edu/ontologies/ontologyOfScience>.
- [VKV⁺06] MAZ VOLKEL, MARKUS KROTZSCH, DENNY VRANDECIC, HEIKO ALLER, and RUDI STUDER. Semantic wikipedia. In *Proceedings of the 15th International World*, pages 585–594, Edinburgo, 2006.
- [W3C02] W3C. Web services architecture requirements. Technical report, 2002. <http://www.w3c.org/2002/-ws/arch/2/wd-wsawg-reqs-04232002.html>.
- [W3C04] W3C. Web services architecture (wsa). Technical report, Febrero 2004. <http://www.w3c.org/TR/2004/NOTE-ws-arch-20040211/>.
- [WCD03] Y. WILKS, F. CIRAVEGNA, and A. DINGLI. Automatic semantic annotation using unsupervised information extraction and integration. In Sanibel, editor, *Proceedings of the Workshop on Knowledge Markup and Semantic Annotation (KCAP'03)*, Florida, 2003.
- [WF23] INC. WIKIMEDIA FOUNDATION. Wikipedia. <http://fr.wikipedia.org/wiki/Accueil>, Mayo 23. Citado el: 4 de Julio de 2012.

- [WIK12] WIKIPEDIA, 2012. <http://es.wikipedia.org/>.
- [WOR12a] WORDPRESS, 2012. <http://francisthemulenews.wordpress.com/>.
- [WOR12b] WORDPRESS, Mayo 2012. <http://informatica-practica.net/wordpress/>.
- [Xat12] Xatakaciencia, Mayo 2012. <http://www.xatakaciencia.com/>.

Apéndice A

Glosario de términos

| Nombre | Descripción |
|------------------------------------|--|
| Términos referentes a las ciencias | |
| Ciencia | El conjunto de conocimientos sistemáticamente estructurados, y susceptibles de ser articulados unos con otros. La ciencia surge de la obtención del conocimiento mediante la observación de patrones regulares, de razonamientos y experimentación de ámbitos específicos, a partir de los cuales se generan preguntas, se construyen hipótesis, se deducen principios y se elaboran leyes generales y sistemas metódicamente organizados. [RAE12] |
| Disciplina | Arte, facultad o ciencia. [RAE12] |
| Teoría | 1. f. Conocimiento especulativo considerado con independencia de toda aplicación. 2. f. Serie de las leyes que sirven para relacionar determinado orden de fenómenos. 3. f. Hipótesis cuyas consecuencias se aplican a toda una ciencia o a parte muy importante de ella. 4. f. Entre los antiguos griegos, procesión religiosa. [RAE12] |
| Tecnología | 1. f. Conjunto de teorías y de técnicas que permiten el aprovechamiento práctico del conocimiento científico. 2. f. Tratado de los términos técnicos. 3. f. Lenguaje propio de una ciencia o de un arte. 4. f. Conjunto de los instrumentos y procedimientos industriales de un determinado sector o producto. [RAE12] |
| Ciencias formales | Se llaman a menudo ciencias formales, precisamente porque sus objetos no son cosas ni procesos, sino, para emplear el lenguaje pictórico, formas en las que se puede verter un surtido ilimitado de contenidos, tanto fácticos como empíricos. [BUN12] |
| Ciencias facticias | Las ciencias fácticas o ciencias factuales (física, química, fisiología, biología, entre otras) están basadas en buscar la coherencia entre los hechos y la representación mental de los mismos. Esta coherencia es necesaria pero no suficiente, porque además exige la observación y la experimentación. [WIK12] |
| Continúa en la siguiente página | |

Cuadro A.1 – continuación de la página anterior

| Nombre | Descripción |
|-----------------------------|---|
| Matemáticas | Es una ciencia formal que, partiendo de axiomas y siguiendo el razonamiento lógico, estudia las propiedades y relaciones entre entes abstractos (números, figuras geométricas, símbolos). Las matemáticas se emplean para estudiar relaciones cuantitativas, estructuras, relaciones geométricas y las magnitudes variables. [WIK12] |
| Lógica | La lógica es una ciencia formal y una rama de la filosofía que estudia los principios de la demostración e inferencia válida. [WIK12] |
| Teoría de la Información | Esta teoría está relacionada con las leyes matemáticas que rigen la transmisión y el procesamiento de la información y se ocupa de la medición de la información y de la representación de la misma así como también de la capacidad de los sistemas de comunicación para transmitir y procesar información. [WIK12] |
| Teoría de juegos | La teoría de juegos es un área de la matemática aplicada que utiliza modelos para estudiar interacciones en estructuras formalizadas de incentivos (los llamados «juegos») y llevar a cabo procesos de decisión. [WIK12] |
| Cibernética | La cibernética es el estudio interdisciplinario de la estructura de los sistemas reguladores. La cibernética está estrechamente vinculada a la teoría de control y a la teoría de sistemas. [WIK12] |
| Estadística | La estadística es una ciencia formal que estudia la recolección, análisis e interpretación de datos, ya sea para ayudar en la toma de decisiones o para explicar condiciones regulares o irregulares de algún fenómeno o estudio aplicado, de ocurrencia en forma aleatoria o condicional. Sin embargo estadística es más que eso, en otras palabras es el vehículo que permite llevar a cabo el proceso relacionado con la investigación científica. [WIK12] |
| Teoría de la computación | La teoría de la computación es una rama de la matemática y la computación que centra su interés en las limitaciones y capacidades fundamentales de las computadoras. Específicamente esta teoría busca modelos matemáticos que formalizan el concepto de hacer un cómputo (cuenta o cálculo) y la clasificación de problemas. [WIK12] |
| Criptografía | Consta de algoritmos para proteger datos privados, incluyendo el cifrado. [WIK12] |
| Teoría de grafos | Recursos elementales para las estructuras de almacenamiento de datos y para los algoritmos de búsqueda. [WIK12] |
| Teoría de tipos | Análisis formal de los tipos de datos, y el uso de estos para entender las propiedades de los programas, en particular la seguridad de estos. [WIK12] |
| Teoría de autómatas | La teoría de autómatas es una rama de las ciencias de la computación que estudia las máquinas abstractas y los problemas que éstas son capaces de resolver. La teoría de autómatas está estrechamente relacionada con la teoría del lenguaje formal ya que los autómatas son clasificados a menudo por la clase de lenguajes formales que son capaces de reconocer. [WIK12] |
| Teoría de la computabilidad | La Teoría de la computabilidad es la parte de la computación que estudia los problemas de decisión que pueden ser resueltos con un algoritmo o equivalentemente con una máquina de Turing. [WIK12] |

Continúa en la siguiente página

Cuadro A.1 – continuación de la página anterior

| Nombre | Descripción |
|--|--|
| Teoría de la complejidad computacional | Límites fundamentales (en especial de espacio en memoria y tiempo) de los cómputos. [WIK12] |
| Análisis de algoritmos | El análisis de algoritmos es una parte importante de la Teoría de complejidad computacional más amplia, que provee estimaciones teóricas para los recursos que necesita cualquier algoritmo que resuelva un problema computacional dado. [WIK12] |
| Algoritmos | Son los procesos formales usados para los cómputos, y eficiencia de estos procesos. [WIK12] |
| Estructuras de datos | Mecanismos que sirven para organizar y manipular datos. [WIK12] |
| Especificación | Desarrollo de métodos para definir formalmente (matemática y lógicamente), el comportamiento esperado de un algoritmo, para luego probar terminación y correctitud. [WIK12] |
| Compiladores | Formas de traducir programas computacionales, usualmente a partir de lenguajes de alto nivel a lenguajes de bajo nivel. [WIK12] |
| Teoría de lenguajes de programación | Lenguajes formales para expresar algoritmos y las propiedades de estos lenguajes. [WIK12] |
| Bases de Datos | Es un conjunto de datos pertenecientes a un mismo contexto y almacenados sistemáticamente para su posterior uso. [WIK12] |
| Minería de Datos | Estudio de algoritmos para buscar y procesar información en documentos y bases de datos; muy relacionada con la adquisición de información. [WIK12] |
| Programación concurrente | Teoría y práctica de cómputos simultáneos y computación interactiva. [WIK12] |
| Redes de computadora | Algoritmos y protocolos para comunicar eficientemente datos a través de largas distancias, incluye también la corrección de errores. [WIK12] |
| Cómputo paralelo | Computación usando múltiples computadoras y múltiples procesadores en paralelo. [WIK12] |
| Cómputo distribuido | Sistemas utilizando múltiples procesadores repartidos en una gran área geográfica. [WIK12] |
| Inteligencia Artificial | La implementación y estudio de sistemas que exhiben (ya sea por su comportamiento o aparentemente) una inteligencia autónoma o comportamiento propio, a veces inspirado por las características de los seres vivos. Las ciencias de la computación están relacionadas con la IA, ya que el software y las computadoras son herramientas básicas para el desarrollo y progreso de la inteligencia artificial. [WIK12] |
| Continúa en la siguiente página | |

Cuadro A.1 – continuación de la página anterior

| Nombre | Descripción |
|---------------------------|---|
| Razonamiento automatizado | Es un área de las ciencias de la computación dedicada a comprender diferentes aspectos del razonamiento de forma que permita la creación de programas con el objeto de permitir a los ordenadores razonar de forma autónoma o casi autónoma. Es considerado un subcampo de la Inteligencia Artificial. [WIK12] |
| Robótica | Algoritmos para controlar el comportamiento de los robots. [WIK12] |
| Visión por computadora | Algoritmos para extraer objetos tridimensionales de una imagen bidimensional. [WIK12] |
| Computación gráfica | Algoritmos tanto para generar sintéticamente imágenes visuales como para integrar o alterar la información visual y espacial tomada del mundo real. [WIK12] |
| Geometría computacional | Es una rama de las ciencias de la computación dedicada al estudio de los algoritmos que pueden ser expresados en términos de geometría. Algunos de los problemas puramente geométricos surgen del estudio de los algoritmos geométricos computacionales, y este tipo de problemas también son considerados para formar parte de la geometría computacional. [WIK12] |
| Bioinformática | Es la aplicación de tecnología de computadores a la gestión y análisis de datos biológicos. Hace referencia a campos de estudios interdisciplinarios muy vinculados, que requieren el uso o el desarrollo de diferentes técnicas que incluyen las ciencias de la computación y otras ciencias para solucionar problemas, analizar datos, o simular sistemas o mecanismos, todos ellos de índole biológica, y usualmente (pero no de forma exclusiva) en el nivel molecular. El núcleo principal de estas técnicas se encuentra en la utilización de recursos computacionales para solucionar o investigar problemas sobre escalas de tal magnitud que sobrepasan el discernimiento humano. La investigación en biología computacional se solapa a menudo con la biología de sistemas. [WIK12] |
| Computación cuántica | Paradigma de computación basado en la Mecánica Cuántica. [WIK12] |
| Ingeniería de software | Ingeniería de software es la aplicación práctica del conocimiento científico al diseño y construcción de programas de computadora y a la documentación asociada requerida para desarrollar, operar y mantenerlos. Se conoce también como desarrollo de software o producción de software [WIK12] |
| . Ciencia computacional | Las ciencias de la computación son aquellas que abarcan las bases teóricas de la información y la computación, así como su aplicación en sistemas computacionales. [WIK12] |
| . Química | Ciencia que estudia tanto la composición, estructura y propiedades de la materia como los cambios que ésta experimenta durante las reacciones químicas y su relación con la energía. [WIK12] |

Continúa en la siguiente página

Cuadro A.1 – continuación de la página anterior

| Nombre | Descripción |
|---------------------------------|---|
| . Biología | la ciencia que tiene como objeto de estudio a los seres vivos y, más específicamente, su origen, su evolución y sus propiedades: génesis, nutrición, morfogénesis, reproducción, patogenia, etc. [WIK12] |
| Sociología | Es la ciencia social que estudia los fenómenos colectivos producidos por la actividad social de los seres humanos dentro del contexto histórico-cultural en el que se encuentran inmersos. [WIK12] |
| Antropología | Es una ciencia social que estudia al ser humano de una forma integral. Para abarcar la materia de su estudio, la Antropología recurre a herramientas y conocimientos producidos por las ciencias naturales y las ciencias sociales. [WIK12] |
| Economía | Es la ciencia social que estudia el comportamiento económico de agentes individuales: producción, intercambio, distribución, consumo de bienes y servicios, entendidos estos como medios de satisfacer necesidades humanas y resultado individual o colectivo de la sociedad. [WIK12] |
| Astronomía | Es la ciencia que se compone del estudio de los cuerpos celestes del Universo, incluidos los planetas y sus satélites, los cometas y meteoroides, las estrellas y la materia interestelar, los sistemas de estrellas, gas y polvo llamados galaxias y los cúmulos de galaxias; por lo que estudia sus movimientos y los fenómenos ligados a ellos. [WIK12] |
| Botánica | Es una rama de la biología y es la ciencia que se ocupa del estudio de los vegetales, bajo todos sus aspectos, lo cual incluye su descripción, clasificación, distribución, identificación y el estudio de su reproducción, fisiología, morfología, relaciones recíprocas, relaciones con los otros seres vivos y efectos provocados sobre el medio en el que se encuentran [WIK12] |
| Ecología | Es la ciencia que estudia a los seres vivos, su ambiente, la distribución, abundancia y cómo esas propiedades son afectadas por la interacción entre los organismos y su ambiente. [WIK12] |
| Zoología | Es la disciplina biológica que se encarga del estudio de los animales. [WIK12] |
| Microbiología | La microbiología es la rama de la biología encargada del estudio de los microorganismos, seres vivos también conocidos como microbios. [WIK12] |
| Geografía | La geografía es la ciencia que estudia la superficie terrestre, las sociedades que la habitan y los territorios, paisajes, lugares o regiones, que forman al relacionarse entre sí. [WIK12] |
| Física | Es la ciencia natural de las propiedades del espacio, el movimiento , el tiempo, la materia , la energía y la Masa así como sus interacciones. [WIK12] |
| Bioquímica | Es una ciencia que estudia la composición química de los seres vivos, especialmente las proteínas, carbohidratos, lípidos y ácidos nucleicos, además de otras pequeñas moléculas presentes en las células y las reacciones químicas que sufren estos compuestos (metabolismo) que les permiten obtener energía (catabolismo) y generar biomoléculas propias (anabolismo). [WIK12] |
| Continua en la siguiente página | |

Cuadro A.1 – continuación de la página anterior

| Nombre | Descripción |
|--|--|
| Lingüística | La lingüística es el estudio científico tanto de la estructura de las lenguas naturales como del conocimiento que los hablantes poseen de ellas. [WIK12] |
| Psicología | Es la ciencia que estudia la conducta o comportamiento humano y los procesos mentales. [WIK12] |
| Ciencia política | Ciencia política, teoría política, doctrinas políticas o politología son denominaciones de una ciencia social que estudia la teoría y práctica de la política, los sistemas y comportamientos políticos. Su objetivo es establecer, a partir de la observación de hechos de la realidad política, principios generales acerca de su funcionamiento. Interactúa con otras muchas ciencias sociales, como la economía, la sociología, las relaciones internacionales, etc. [WIK12] |
| Demografía | Es la ciencia que tiene como objetivo el estudio de las poblaciones humanas y que trata de su dimensión, estructura, evolución y características generales. [WIK12] |
| Urbanismo | Es la ciencia y técnica de la ordenación de las ciudades y del territorio. El objeto principal de la disciplina urbanística es la planificación de las ciudades y del territorio, los estudios sobre el fenómeno urbano, la acción de la urbanización y la organización de la ciudad y el territorio. [WIK12] |
| Ciencias de la comunicación | Las ciencias de la comunicación son aquellas disciplinas de las ciencias sociales que se encargan de estudiar la esencia de los procesos de comunicación como fenómenos en sí mismos, los medios que se emplean y el conjunto semiótico que construyen, generando sus propios métodos de estudio y herramientas analíticas. [WIK12] |
| Filosofía | Es el estudio de una variedad de problemas fundamentales acerca de cuestiones como la existencia, el conocimiento, la verdad, la moral, la belleza, la mente y el lenguaje. [WIK12] |
| Historia | La historia es la ciencia que tiene como objeto de estudio el pasado de la humanidad y como método el propio de las ciencias sociales. [WIK12] |
| Teología | Es el estudio y conjunto de conocimientos acerca de la divinidad. [WIK12] |
| Semiótica | Se define como el estudio de los signos, su estructura y la relación entre el signifiante y el concepto de significado. [WIK12] |
| Términos referentes a las redes sociales | |
| Artículo (Article WikiOnt) | En nuestro contexto, así llamaremos en la ontología a la publicación realizada por algún usuario de Moveek. |
| Autor (Creator DC) | Autor de una publicación. |
| Descripción (Description DC) | Comentario textual de algún recurso multimedia publicado en la red social. |
| Continua en la siguiente página | |

Cuadro A.1 – continuación de la página anterior

| Nombre | Descripción |
|----------------------|--|
| Fecha (date DC) | Fecha de publicación. |
| ámbito (Coverage DC) | Término de OntoScience al que se relaciona la publicación realizada. |

Cuadro A.1: Glosario de términos

Apéndice B

Clases de la Ontología

| Términos de redes sociales |
|-----------------------------------|
| Artículo (Article WikiOnt) |
| Autor (Creator DC) |
| Descripción (Description DC) |
| Fecha (date DC) |
| Ámbito (Coverage DC) |

Cuadro B.1: Términos referentes a las redes sociales

| Términos científicos | |
|--|-----------------------------|
| Ciencia | Química |
| Disciplina | Biología |
| Teoría | Sociología |
| Tecnología | Antropología |
| Ciencias formales | Economía |
| Ciencias facticias | Astronomía |
| Matemáticas | Botánica |
| Lógica | Ecología |
| Teoría de la Información | Zoología |
| Teoría de juegos | Microbiología |
| Cibernética | Geografía |
| Estadística | Física |
| Teoría de la computación | Bioquímica |
| Criptografía | Lingüística |
| Teoría de grafos | Psicología |
| Teoría de tipos | Ciencia política |
| Teoría de autómatas | Demografía |
| Teoría de la computabilidad | Urbanismo |
| Teoría de la complejidad computacional | Ciencias de la Comunicación |
| Análisis de algoritmos | Filosofía |
| Algoritmos | Historia |
| Estructuras de datos | Teología |
| Especificación | Semiótica |
| Compiladores | Bases de Datos |
| Teoría de lenguajes de programación | Minería de Datos |
| Programación concurrente | Redes de computadora |
| Cómputo paralelo | Cómputo distribuido |
| Inteligencia Artificial | Razonamiento automatizado |
| Robótica | Visión por computadora |
| Computación gráfica | Geometría computacional |
| Bioinformática | Computación cuántica |
| Ingeniería de software | Ciencia computacional |

Cuadro B.2: Términos científicos de la ontología.

Apéndice C

Diagramas

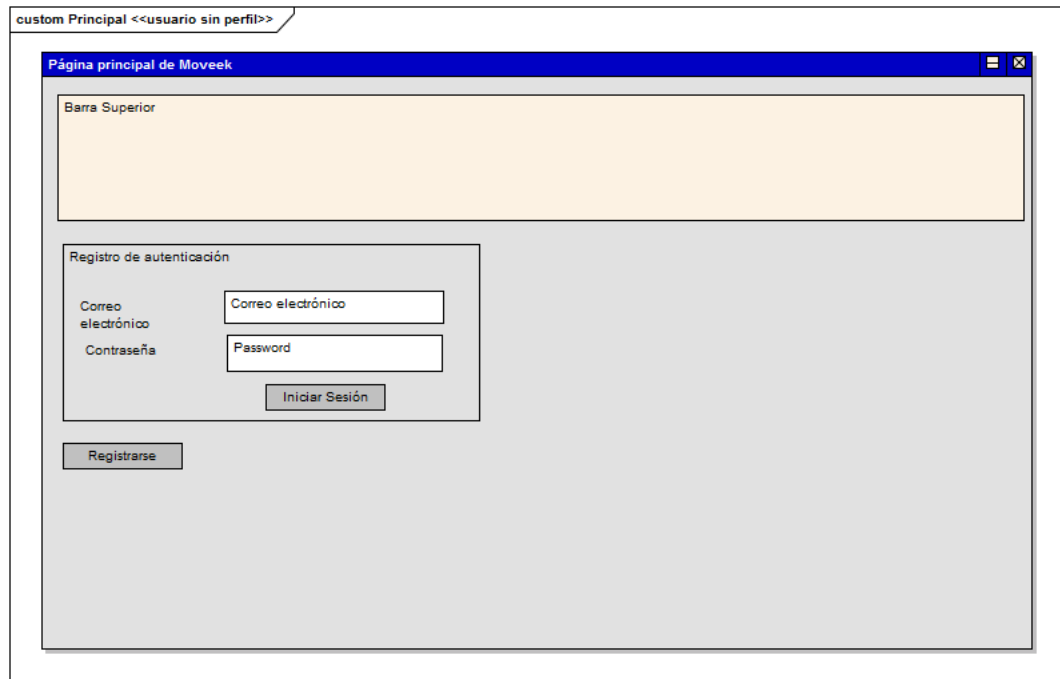


Figura C.1: Prototipo de Interfaz para el ingreso a la red social.

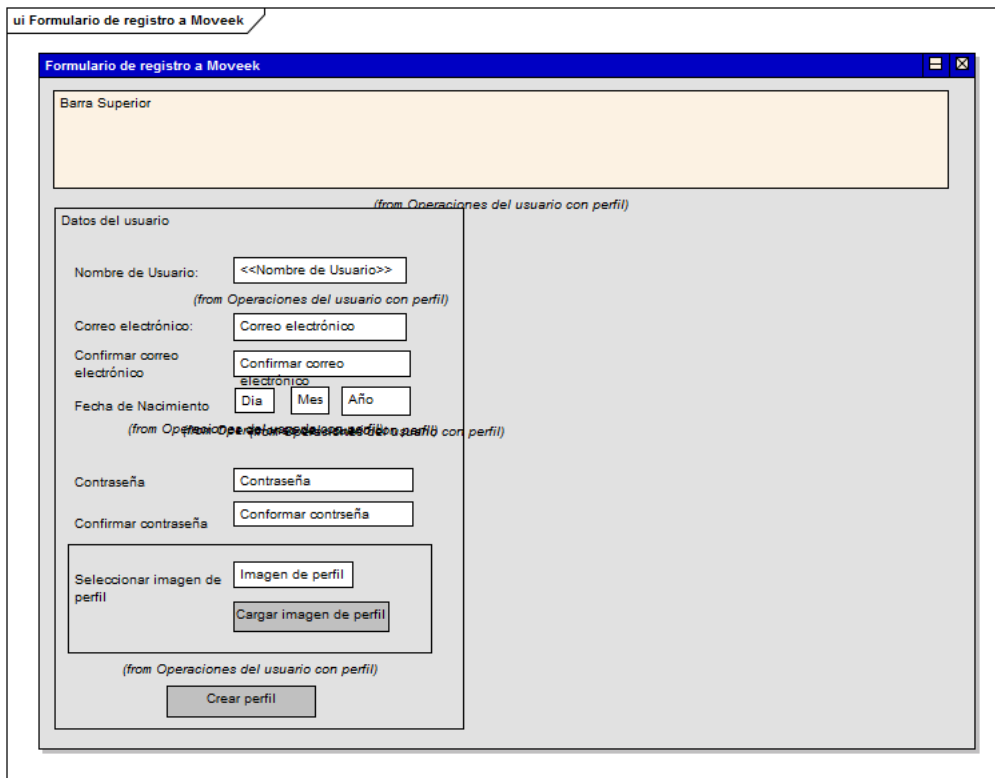


Figura C.2: Prototipo de Interfaz para el formulario de registro a Moveek.

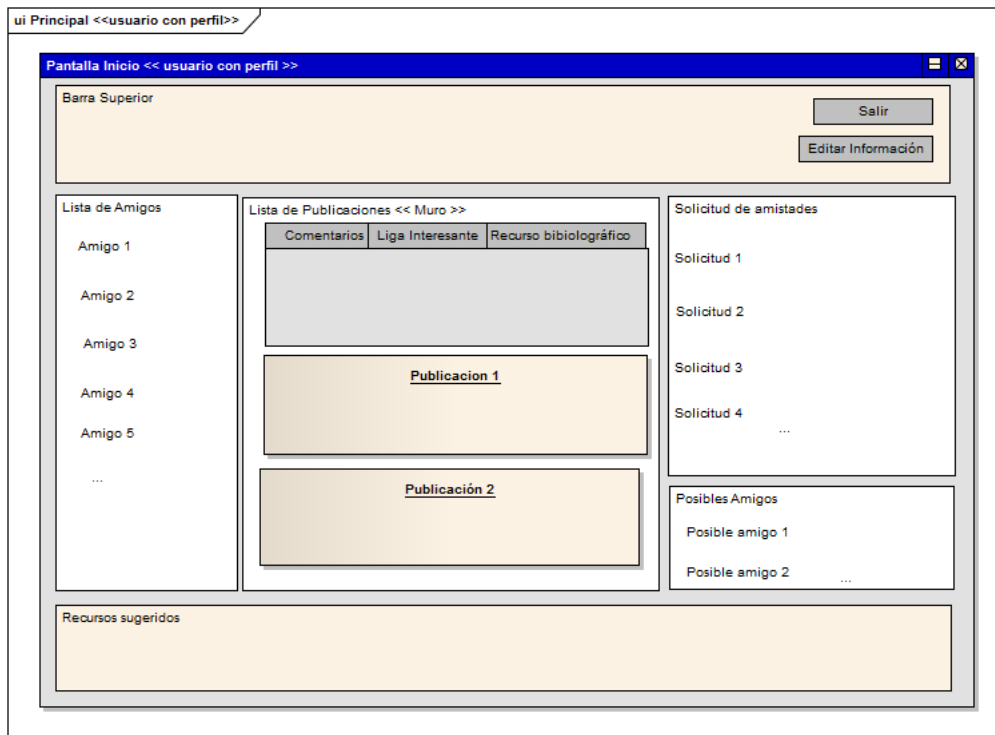


Figura C.3: Prototipo de Interfaz de la pantalla principal del usuario.

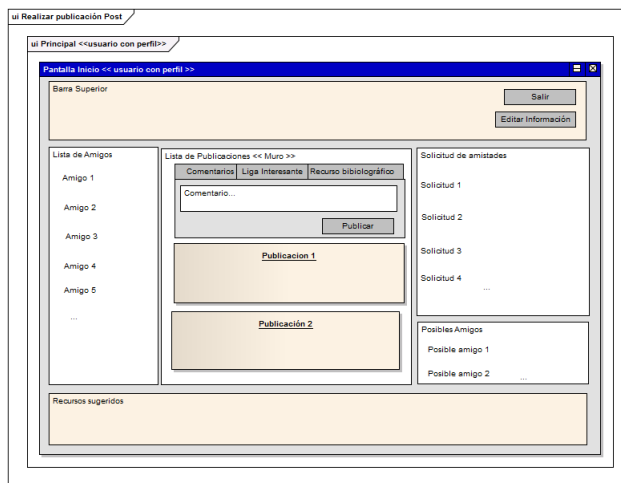


Figura C.4: Prototipo de Interfaz para la acción de realizar publicación.

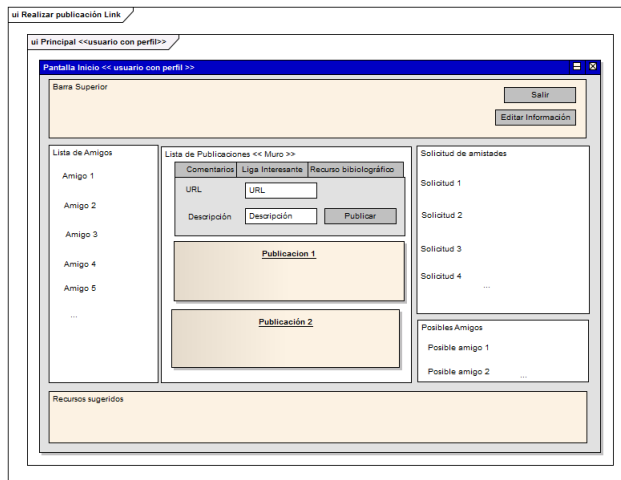


Figura C.5: Prototipo de Interfaz para la acción de realizar publicación de tipo liga.

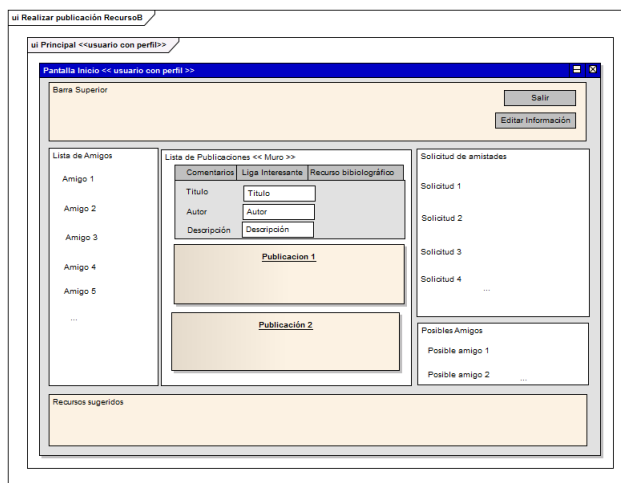


Figura C.6: Prototipo de Interfaz para la acción de realizar publicación de tipo documental.

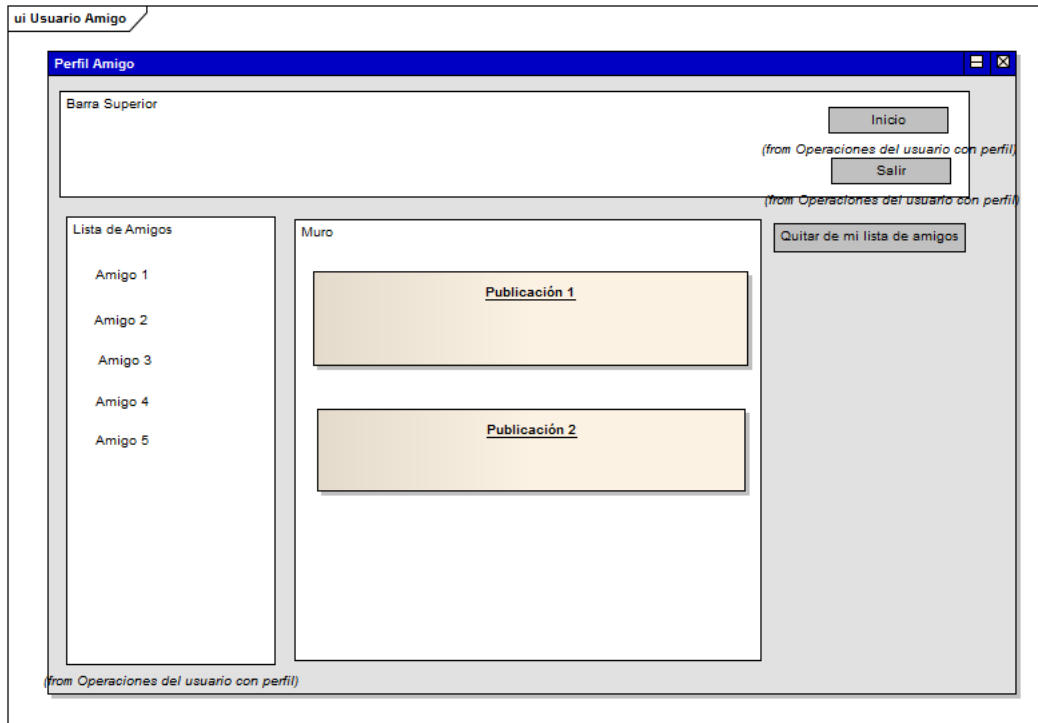


Figura C.7: Prototipo de Interfaz de la visita a un usuario Amigo.

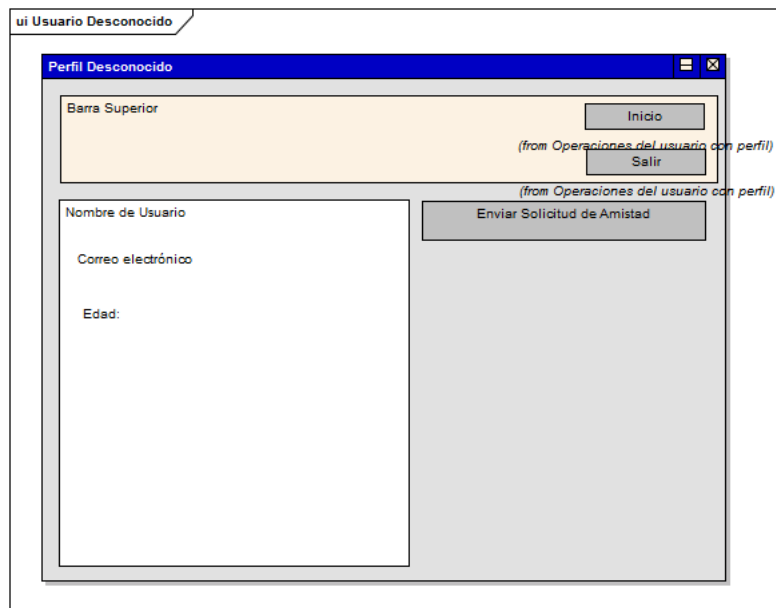


Figura C.8: Prototipo de Interfaz de la visita a un usuario desconocido.

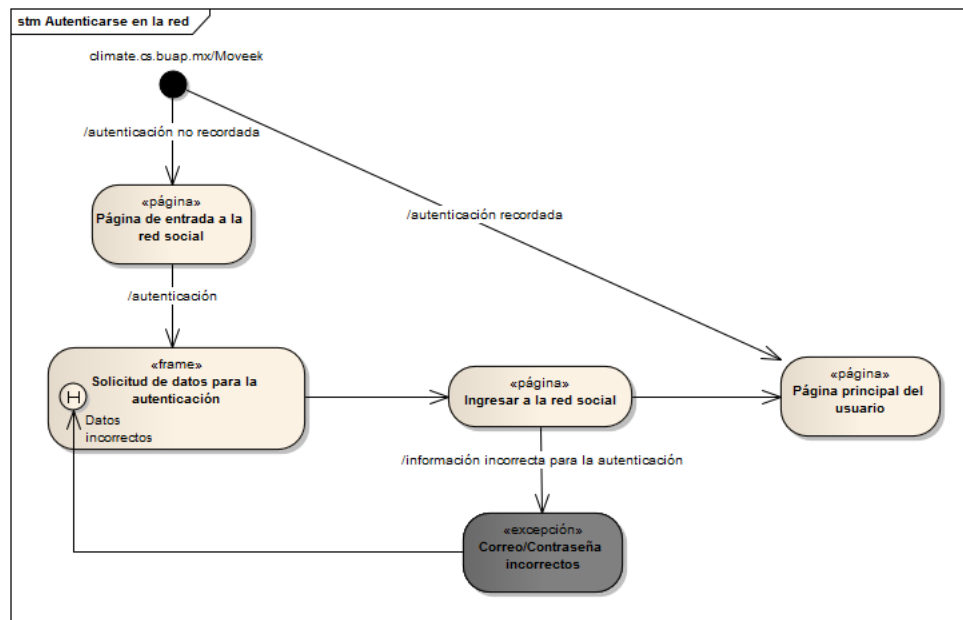


Figura C.9: Modelo de navegación para autenticarse en la red social.

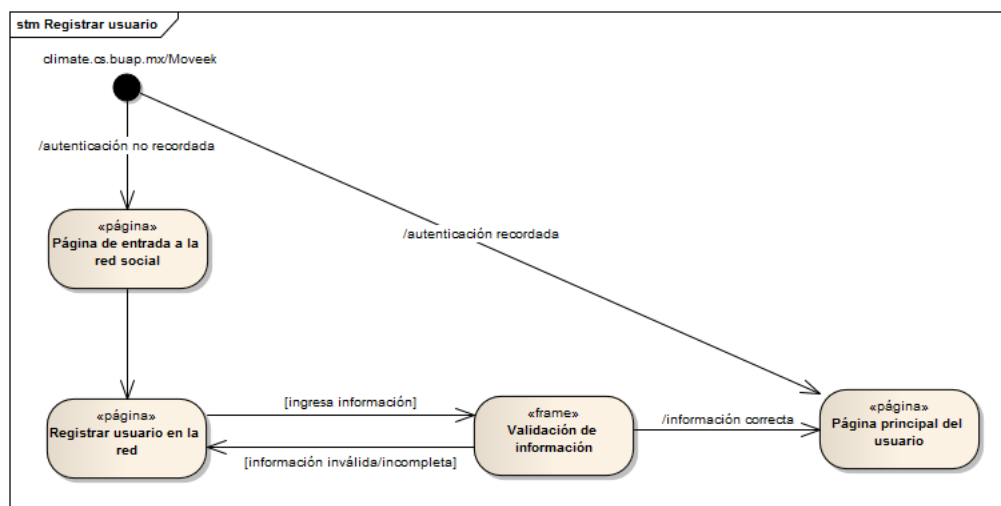


Figura C.10: Modelo de navegación para registrar un usuario.

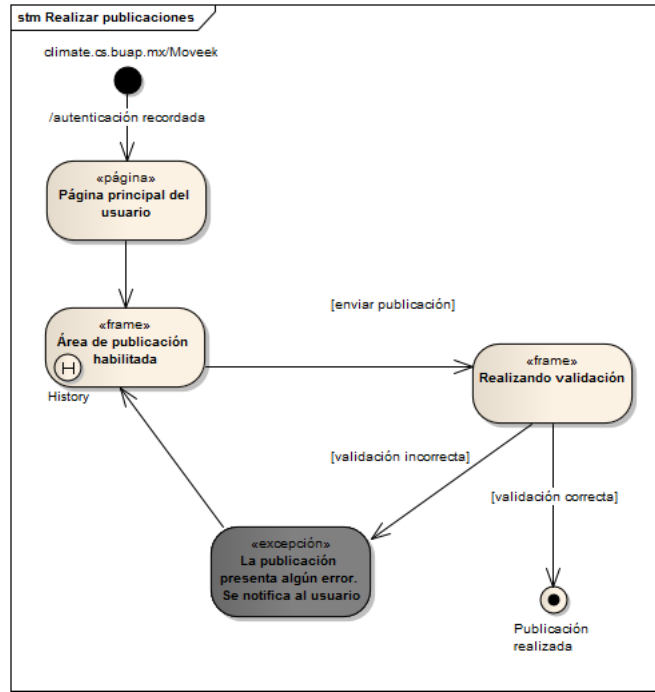


Figura C.11: Modelo de navegación para realizar una publicación.

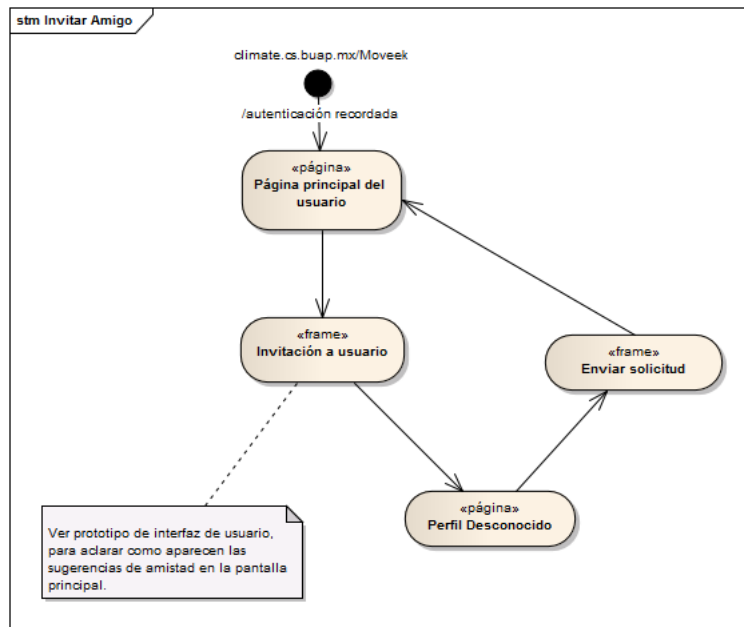


Figura C.12: Modelo de navegación para invitar a un usuario a ser tu amigo.

Apéndice D

Herramienta GATE

En este apéndice explicaremos brevemente la herramienta que utilizamos para realizar la extracción de información.

La extracción de la información tradicional esta basada en una superficie estructural por ejemplo el reconocimiento de Personas, Lugares, Organizaciones, Fechas, Horas, etc. Para la web Semántica, necesitamos la información almacenada en una estructura jerárquica. La idea es que se establezcan los metadatos a los documentos apuntando a conceptos de la ontología. En nuestro caso exportamos la información en la parte textual de la publicación. En este trabajo usamos la herramienta GATE para realizar la tarea de extraer la información de las publicaciones y asociarlas a los términos de nuestra ontología.

GATE Developer es una entorno de desarrollo que proporciona un basto conjunto de herramientas gráficas para la creación, medición y componentes de software para el mantenimiento y procesamiento de lenguaje Natural. GATE Developer es un software libre disponible bajo licencia GNU Lesser General Public Licence 3.0, y puede ser descargado desde la página <http://gate.ac.uk/family/developer.html>.

Para las tareas de web Semántica GATE dispone de tres recursos de procesamiento que utilizamos en este trabajo para la extracción de la información. Estos recursos son los nomencladores(*gazetter*), los clasificadores(*classifier*), y la combinación de ambos. En nuestro caso usamos ámbos recursos, por un lado el nomenclator y el clasificador se analizan el texto de entrada y por otro lado asocian los conceptos encontrados en el texto a la ontología. La Figura D.2 muestra la manera en que se configura el nomenclátor que pide como entrada, entre otras cosas, la referencia a la ontología a la que va a asociar los términos.

En la Figura D.3 se muestra que, a paritr del nomenclator se crea el clasificador. Una vez configuradas ambos recursos de procesamiento se procede a crear el pipeline que realizara el procesamiento del texto completo.

En la Figura D.4 se muestra el pipeline de la aplicación que cargamos programáticamente

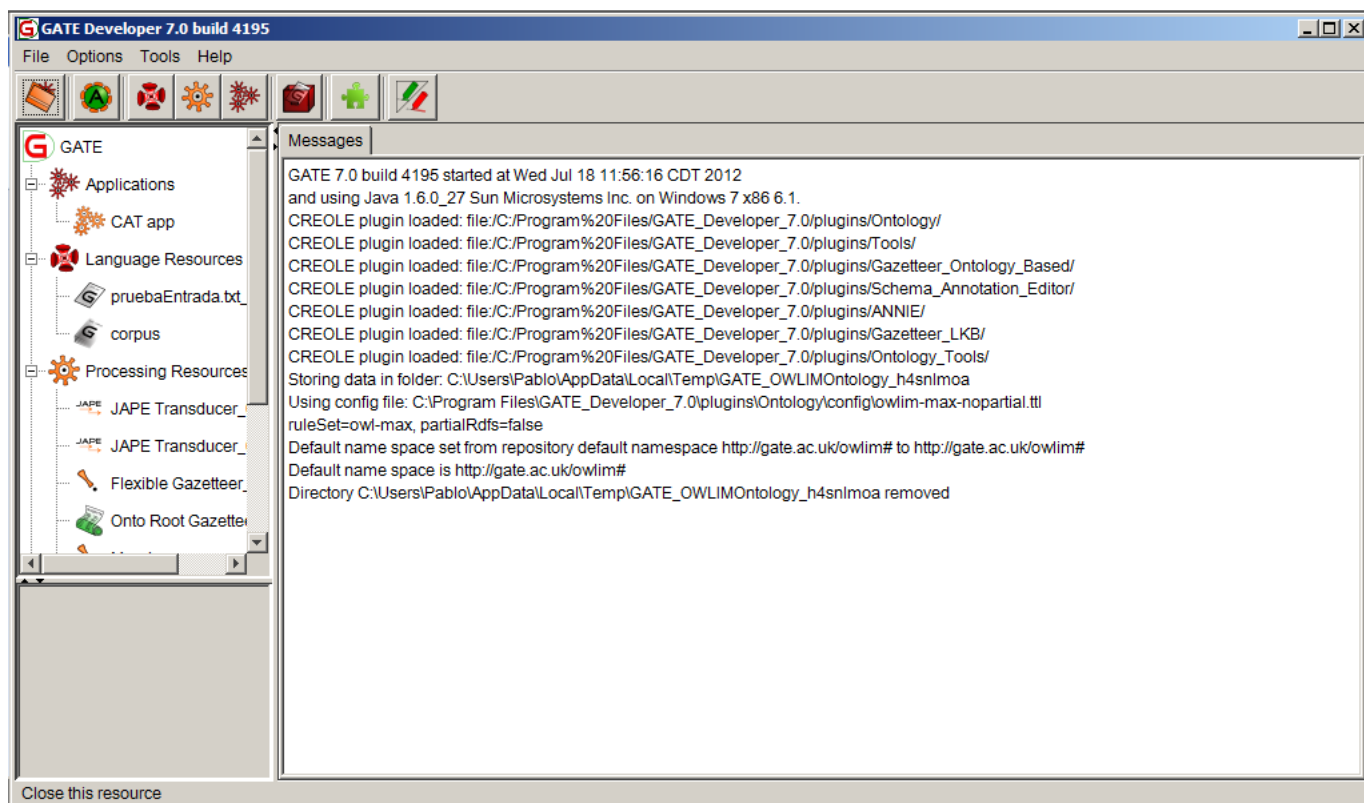


Figura D.1: Pantalla principal de GATE.

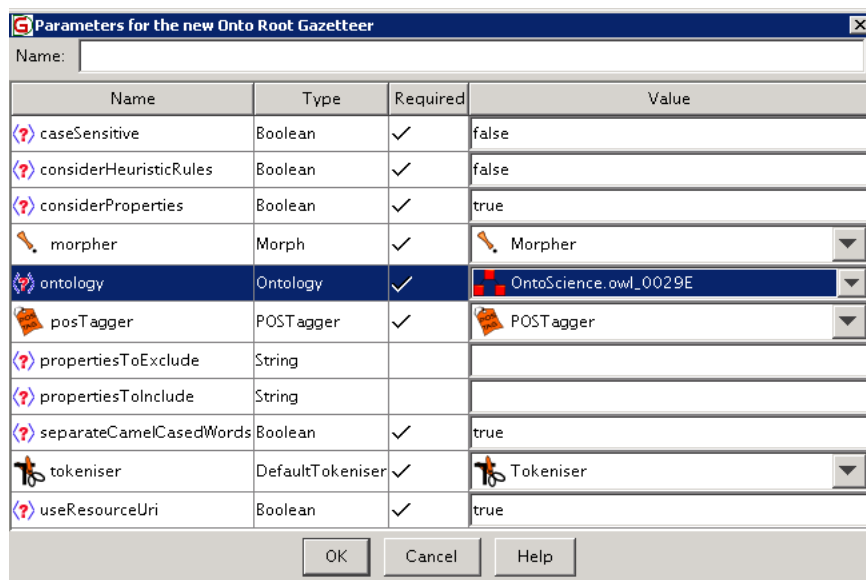


Figura D.2: Asociación de la ontología al nomenclator.

que se uso para realizar las tareas de la anotación semántica mencionadas en el Capítulo 4.

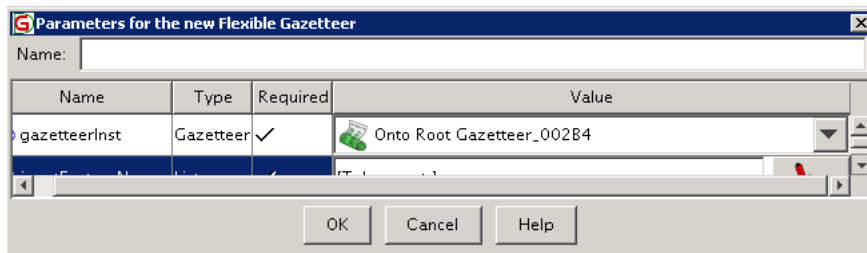


Figura D.3: Asociación de la ontología al nomenclátor.

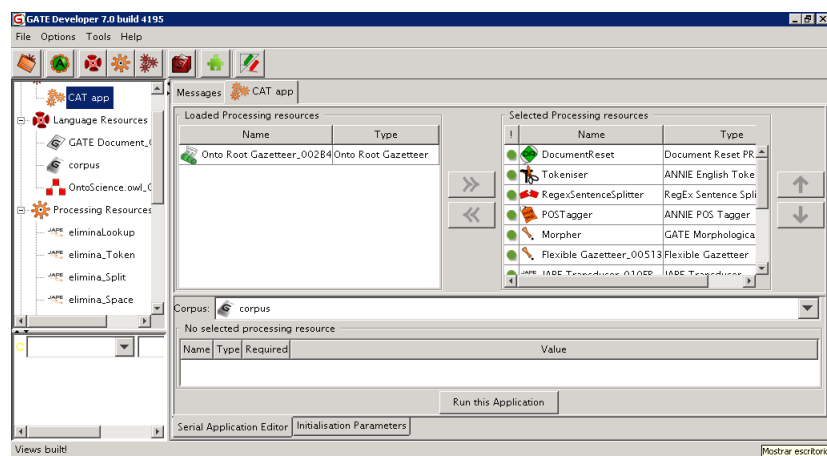


Figura D.4: Pipeline completo para la extracción de la información.

Apéndice E

WSDL

```
1 <?xml version='1.0' encoding='UTF-8'?><!-- Published by JAX-WS RI at http://jax-
   ws.dev.java.net. RI's version is Metro/2.1.1-b09 (branches/2.1-6834;
   2011-07-16T17:14:48+0000) JAXWS-RI/2.2.5-promoted-b04 JAXWS/2.2. --><!--
   Generated by JAX-WS RI at http://jax-ws.dev.java.net. RI's version is Metro
   /2.1.1-b09 (branches/2.1-6834; 2011-07-16T17:14:48+0000) JAXWS-RI/2.2.5-
   promoted-b04 JAXWS/2.2. --><definitions xmlns:wsu="http://docs.oasis-open.
   org/wss/2004/01/oasis-200401-wss-wssecurity-utility-1.0.xsd" xmlns:wsp="http
   ://www.w3.org/ns/ws-policy" xmlns:wsp1_2="http://schemas.xmlsoap.org/ws
   /2004/09/policy" xmlns:wsam="http://www.w3.org/2007/05/addressing/metadata"
   xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/" xmlns:tns="http://
   gestionOntologia/" xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns="http
   ://schemas.xmlsoap.org/wsdl/" targetNamespace="http://gestionOntologia/"
   name="operacionesProtege">
2 <types>
3 <xsd:schema>
4 <xsd:import namespace="http://gestionOntologia/" schemaLocation="http://
   localhost:8080/ServiciosSemanticos/operacionesProtege?xsd=1"/>
5 </xsd:schema>
6 </types>
7 <message name="crearInstancia">
8 <part name="parameters" element="tns:crearInstancia"/>
9 </message>
10 <message name="crearInstanciaResponse">
11 <part name="parameters" element="tns:crearInstanciaResponse"/>
12 </message>
13 <message name="crearInstanciaConInstancia">
14 <part name="parameters" element="tns:crearInstanciaConInstancia"/>
15 </message>
16 <message name="crearInstanciaConInstanciaResponse">
17 <part name="parameters" element="tns:crearInstanciaConInstanciaResponse"/>
18 </message>
19 <message name="EsConsistente">
20 <part name="parameters" element="tns:EsConsistente"/>
21 </message>
22 <message name="EsConsistenteResponse">
23 <part name="parameters" element="tns:EsConsistenteResponse"/>
```

```

24 </message>
25 <message name="asociarClaseUsuario">
26 <part name="parameters" element="tns:asociarClaseUsuario"/>
27 </message>
28 <message name="asociarClaseUsuarioResponse">
29 <part name="parameters" element="tns:asociarClaseUsuarioResponse"/>
30 </message>
31 <message name="PublicacionCubreAmbito">
32 <part name="parameters" element="tns:PublicacionCubreAmbito"/>
33 </message>
34 <message name="PublicacionCubreAmbitoResponse">
35 <part name="parameters" element="tns:PublicacionCubreAmbitoResponse"/>
36 </message>
37 <message name="AutorRealizaPublicacion">
38 <part name="parameters" element="tns:AutorRealizaPublicacion"/>
39 </message>
40 <message name="AutorRealizaPublicacionResponse">
41 <part name="parameters" element="tns:AutorRealizaPublicacionResponse"/>
42 </message>
43 <message name="ObtenerPublicacionesTermino">
44 <part name="parameters" element="tns:ObtenerPublicacionesTermino"/>
45 </message>
46 <message name="ObtenerPublicacionesTerminoResponse">
47 <part name="parameters" element="tns:ObtenerPublicacionesTerminoResponse"/>
48 </message>
49 <portType name="operacionesProtege">
50 <operation name="crearInstancia">
51 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    crearInstanciaRequest" message="tns:crearInstancia"/>
52 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    crearInstanciaResponse" message="tns:crearInstanciaResponse"/>
53 </operation>
54 <operation name="crearInstanciaConInstancia">
55 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    crearInstanciaConInstanciaRequest" message="tns:crearInstanciaConInstancia
    "/>
56 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    crearInstanciaConInstanciaResponse" message="tns:
    crearInstanciaConInstanciaResponse"/>
57 </operation>
58 <operation name="EsConsistente">
59 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    EsConsistenteRequest" message="tns:EsConsistente"/>
60 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    EsConsistenteResponse" message="tns:EsConsistenteResponse"/>
61 </operation>
62 <operation name="asociarClaseUsuario">
63 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    asociarClaseUsuarioRequest" message="tns:asociarClaseUsuario"/>
64 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    asociarClaseUsuarioResponse" message="tns:asociarClaseUsuarioResponse"/>
65 </operation>
66 <operation name="PublicacionCubreAmbito">

```

```

67 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    PublicacionCubreAmbitoRequest" message="tns:PublicacionCubreAmbito"/>
68 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    PublicacionCubreAmbitoResponse" message="tns:PublicacionCubreAmbitoResponse
    "/>
69 </operation>
70 <operation name="AutorRealizaPublicacion">
71 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    AutorRealizaPublicacionRequest" message="tns:AutorRealizaPublicacion"/>
72 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    AutorRealizaPublicacionResponse" message="tns:
    AutorRealizaPublicacionResponse"/>
73 </operation>
74 <operation name="ObtenerPublicacionesTermino">
75 <input wsam:Action="http://gestionOntologia/operacionesProtege/
    ObtenerPublicacionesTerminoRequest" message="tns:ObtenerPublicacionesTermino
    "/>
76 <output wsam:Action="http://gestionOntologia/operacionesProtege/
    ObtenerPublicacionesTerminoResponse" message="tns:
    ObtenerPublicacionesTerminoResponse"/>
77 </operation>
78 </portType>
79 <binding name="operacionesProtegePortBinding" type="tns:operacionesProtege">
80 <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="document
    "/>
81 <operation name="crearInstancia">
82 <soap:operation soapAction=""/>
83 <input>
84 <soap:body use="literal"/>
85 </input>
86 <output>
87 <soap:body use="literal"/>
88 </output>
89 </operation>
90 <operation name="crearInstanciaConInstancia">
91 <soap:operation soapAction=""/>
92 <input>
93 <soap:body use="literal"/>
94 </input>
95 <output>
96 <soap:body use="literal"/>
97 </output>
98 </operation>
99 <operation name="EsConsistente">
100 <soap:operation soapAction=""/>
101 <input>
102 <soap:body use="literal"/>
103 </input>
104 <output>
105 <soap:body use="literal"/>
106 </output>
107 </operation>
108 <operation name="asociarClaseUsuario">

```

```

109 <soap:operation soapAction=""/>
110 <input>
111 <soap:body use="literal"/>
112 </input>
113 <output>
114 <soap:body use="literal"/>
115 </output>
116 </operation>
117 <operation name="PublicacionCubreAmbito">
118 <soap:operation soapAction=""/>
119 <input>
120 <soap:body use="literal"/>
121 </input>
122 <output>
123 <soap:body use="literal"/>
124 </output>
125 </operation>
126 <operation name="AutorRealizaPublicacion">
127 <soap:operation soapAction=""/>
128 <input>
129 <soap:body use="literal"/>
130 </input>
131 <output>
132 <soap:body use="literal"/>
133 </output>
134 </operation>
135 <operation name="ObtenerPublicacionesTermino">
136 <soap:operation soapAction=""/>
137 <input>
138 <soap:body use="literal"/>
139 </input>
140 <output>
141 <soap:body use="literal"/>
142 </output>
143 </operation>
144 </binding>
145 <service name="operacionesProtege">
146 <port name="operacionesProtegePort" binding="tns:operacionesProtegePortBinding">
147 <soap:address location="http://localhost:8080/ServiciosSemanticos/
    operacionesProtege"/>
148 </port>
149 </service>
150 </definitions>

```

Figura E.1: Contrato de servicio del servicio Web operacionesProtege

```

1 <?xml version='1.0' encoding='UTF-8'?><!-- Published by JAX-WS RI at http://jax-
    ws.dev.java.net. RI's version is Metro/2.1.1-b09 (branches/2.1-6834;
    2011-07-16T17:14:48+0000) JAXWS-RI/2.2.5-promoted-b04 JAXWS/2.2. --><!--
    Generated by JAX-WS RI at http://jax-ws.dev.java.net. RI's version is Metro
    /2.1.1-b09 (branches/2.1-6834; 2011-07-16T17:14:48+0000) JAXWS-RI/2.2.5-

```

```

promoted-b04 JAXWS/2.2. --><definitions xmlns:wsu="http://docs.oasis-open.
org/wss/2004/01/oasis-200401-wss-wssecurity-utility-1.0.xsd" xmlns:wsp="http
://www.w3.org/ns/ws-policy" xmlns:wsp1_2="http://schemas.xmlsoap.org/ws
/2004/09/policy" xmlns:wsam="http://www.w3.org/2007/05/addressing/metadata"
xmlns:soap="http://schemas.xmlsoap.org/wsdl/soap/" xmlns:tns="http://
anotaciones/" xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns="http://
schemas.xmlsoap.org/wsdl/" targetNamespace="http://anotaciones/" name="
ServiciosAnotacion">
2 <types>
3 <xsd:schema>
4 <xsd:import namespace="http://anotaciones/" schemaLocation="http://localhost
:8080/ServiciosSemanticos/ServiciosAnotacion?xsd=1"/>
5 </xsd:schema>
6 </types>
7 <message name="Anotar">
8 <part name="parameters" element="tns:Anotar"/>
9 </message>
10 <message name="AnotarResponse">
11 <part name="parameters" element="tns:AnotarResponse"/>
12 </message>
13 <portType name="ServiciosAnotacion">
14 <operation name="Anotar">
15 <input wsam:Action="http://anotaciones/ServiciosAnotacion/AnotarRequest" message
="tns:Anotar"/>
16 <output wsam:Action="http://anotaciones/ServiciosAnotacion/AnotarResponse"
message="tns:AnotarResponse"/>
17 </operation>
18 </portType>
19 <binding name="ServiciosAnotacionPortBinding" type="tns:ServiciosAnotacion">
20 <soap:binding transport="http://schemas.xmlsoap.org/soap/http" style="document
"/>
21 <operation name="Anotar">
22 <soap:operation soapAction=""/>
23 <input>
24 <soap:body use="literal"/>
25 </input>
26 <output>
27 <soap:body use="literal"/>
28 </output>
29 </operation>
30 </binding>
31 <service name="ServiciosAnotacion">
32 <port name="ServiciosAnotacionPort" binding="tns:ServiciosAnotacionPortBinding">
33 <soap:address location="http://localhost:8080/ServiciosSemanticos/
ServiciosAnotacion"/>
34 </port>
35 </service>
36 </definitions>

```

Figura E.2: Contrato de servicio del servicio Web servicioAnotacion