

Capítulo 1. Introducción

La época actual es conocida con nombres tales como la era del Ciberespacio, etapa de la Infraestructura de Información Global, era de la Información, época de la Súper Carretera de la Información y era del Ínter espacio. Todos estos términos tienen su origen en el desarrollo y expansión de las redes de computo y su esencia es la información. Información que fluye sobre las redes y esta presente en nuestra computadora personal gracias a la rápida proliferación de servicios como Internet y World Wide Web.

El hecho de tener acceso gratuito a algunos recursos de Internet ha permitido que la búsqueda, generación, evaluación, utilización y recuperación de información se haya convertido en un ejercicio popular entre científicos, estudiantes, hombres de negocios y usuarios comunes, ávidos de información.

En la década de 1984 a 1994, Internet creció de 1000 servidores a 2,000,000 [Oostendorp et al.1994. Se estima que aproximadamente 20 millones de accesos directos ocurren cada día en Internet y que casi 3 millones de URLs están disponibles.

Este fenómeno de la información es único en la historia en términos de cantidad de información disponible, variedad de formas en que se representa dicha información, almacenamiento de la misma, codificación, distribución y manejo. Pero esta "explosión de información" trae consigo muchos problemas incluyendo que tipo de búsquedas deben aplicarse para encontrar información útil y relevante de entre todo ese mar de datos? ,Que técnicas deben implementarse para optimizar el tiempo de búsqueda y acceso? y como recuperar datos que no son texto?

Estos problemas se presentan debido a la naturaleza propia de Internet, ya que es una colección desorganizada de sitios y documentos cuyo crecimiento exponencial ha sobrepasado la habilidad que cualquier usuario experto de bases de datos pudiera tener. Así que es necesario dar a cada usuario la habilidad de organizar una perspectiva individual de Internet. Así mismo es necesario proveer al usuario de medios para manejar dicha información. En este contexto surgen las bibliotecas digitales como una tecnología capaz de organizar la información de una manera eficiente y de proporcionar al usuario nuevas técnicas de adquisición, nuevos métodos de almacenamiento y preservación, nuevas técnicas de clasificación y catalogación y principalmente nuevos métodos

de Interacción Hombre-Computadora [Fox et al. 1995; Fox y Marchionini 1998].

Pero ¿cómo construir una biblioteca digital?, ¿cuánto ancho de banda o que tipo de computo se requiere?, como administrar y acceder las grandes cantidades de información multimedia existente?, bajo que plataforma o arquitectura se debe trabajar?, ¿Qué tecnología desarrollar pare que el usuario indique lo que busca y lo encuentre sin necesidad de tener que supervisar dicha búsqueda?

Muchas disciplinas se han visto involucradas en la creación de bibliotecas, desde economía, psicología, sociología, hasta áreas específicas de la computación como base de datos, Interacción Hombre-Computadora, recuperación de información e incluso la aplicación de tecnologías nuevas como son los agentes.

En particular, el concepto de agente aplicado al desarrollo de bibliotecas digitales permitirá que las tareas de búsqueda, recuperación, filtrado e interpretación de información se optimicen y/o automaticen logrando con esto una comunicación más rápida y segura entre los usuarios y la biblioteca.

En esta tesis se presenta la incorporación de un tipo de agentes conocido como agente móvil a una biblioteca digital específica, la Biblioteca Digital Florística. En la sección 1.3 se profundiza más sobre el tema.

Con el fin de motivar los objetivos del presente proyecto, a continuación se presentan brevemente algunos conceptos básicos de las dos áreas involucradas: bibliotecas digitales y agentes. En los capítulos 2 y 3 se estudian estos conceptos con mayor detalle.

1.1 BIBLIOTECAS DIGITALES

En esta sección se presentan algunos conceptos importantes en el área de bibliotecas digitales.

1.1.1 Definición de Biblioteca Digital

Hasta hoy, se han producido y coleccionado miles de libros, cientos de miles de artículos e innumerables estudios formales e informales además de reportes, revistas, periódicos y videos. Nuestra capacidad para absorber esta información y usarla de manera inteligente es estrecha no sólo por la cantidad y variedad de datos disponibles, sino también por las complejas relaciones que existen entre los tipos de información y las dificultades que resultan para poder interpretarlos.

Para poder almacenar y organizar la información se cuenta con bibliotecas. La misión fundamental de las bibliotecas y los bibliotecarios es proporcionar acceso intelectual y físico al registro humano así como la preservación del mismo. Una biblioteca es fundamentalmente un conjunto organizado de recursos, que incluye componentes físicos, tales como espacio, equipo y medios de almacenamiento, componentes intelectuales tales como política de colección que determinan que materiales serán incluidos y esquemas organizacionales que determinan como va a ser accesada la colección; y gente que maneja las componentes físicas e intelectuales e interactúan con los usuarios para resolver problemas de información.

Las funciones de las bibliotecas convencionales son la colección, organización y representación, el acceso, la recuperación y el análisis, la síntesis y diseminación de la información.

Pero la "explosión de la información" de los últimos años ha motivado que surja un nuevo concepto conocido como biblioteca digital.

Existen muchas definiciones de lo que es una biblioteca digital. Términos como biblioteca electrónica y biblioteca virtual son a veces usados sinónimamente. En particular [Gladney et al. 1994] propone una definición que precise el término y que es ampliamente aceptada en el ambiente de bibliotecas digitales. "Una biblioteca digital es un ensamblaje de computo digital, mecanismos de comunicación y software necesario para reproducir, emular y extender los servicios provistos por bibliotecas convencionales como son la colección, catalogación, búsqueda y diseminación de la información " Un servicio completo de biblioteca digital debe estar acompañado de los servicios esenciales de las bibliotecas tradicionales y además explotar las ventajas de almacenamiento, búsqueda y comunicación digital. Otras definiciones incluyen las de [fox et al. 1993], [Birmingham 1994], [Marchionini y Maurer 1995], [Nurnberg et al. 1995] y [Graham 1995] y se presentan en el capítulo 3. La misión de las bibliotecas digitales es adquirir información, organizarla, hacerla disponible y preservarla. Además de diseminar el conocimiento y estrechar la comunicación y colaboración entre los usuarios de la misma. Pero, cuáles son entonces las diferencias entre una biblioteca tradicional y una digital? Se sabe que una biblioteca digital es vista popularmente como una versión electrónica de una biblioteca tradicional, pero reemplazar el papel por almacenamiento electrónico acarrea cuatro grandes diferencias:

1. Los trabajos (libros, revistas, periódicos, discos, videos, etc.) están almacenados en forma digital.
2. Existencia de una interacción directa del usuario con la fuente de información con el fin de obtener el material deseado.
3. El material siempre es copiado de una versión original.

4. Oportunidades para los usuarios tanto de usar como de depositar información. [Birmingham 1994] menciona algunas características que distinguen a las bibliotecas digitales de las convencionales: Proporcionan información en cualquier tiempo y en cualquier lugar. Proporcionan acceso a colecciones de información multimedia, construidas bajo la integración de texto, imagen, gráficas, audio y video. Soportan acceso y representación personalizada de la información, incluyendo soporte para recuperar sólo información relevante. Mejoran las actividades colaborativas tales como la investigación, el aprendizaje y el diseño, al reducir las barreras de tiempo y distancia.

Para [Gladney et al. 1994] las ventajas de una biblioteca digital sobre una biblioteca física son similares a aquellas que cualquier base de datos digital tiene sobre su contraparte en:

Rápido acceso a la colección de datos.

Mejoramiento en la funcionalidad de revisión y búsqueda.

Rápida distribución desde los puntos de creación y los puntos de almacenamiento al punto de uso.

Un tema particularmente importante en el desarrollo de bibliotecas digitales es la recuperación de información. Esta área ha contribuido significativamente a facilitar el acceso a grandes depósitos de información (tales como bibliotecas digitales).

1.1.2 Recuperación de información

Un reto significativo en bibliotecas digital es presentar la vasta y creciente cantidad y variabilidad de información de Internet de una manera amigable y práctica para los usuarios. Para lograr esto, se deben determinar factores clave tales como las fuentes candidatas más promisorias para buscar información, requerimientos para acceder a esas fuentes, disponibilidad pública, costos por acceder dicha información y lenguajes en los cuales la información está almacenada.

En el futuro, nuestro trabajo será mediado por un acceso rápido y coordinado de la información compartida. A través de bibliotecas digitales, la gente colaborará con colegas sin importar las distancias geográficas y/o temporales; usarán esas bibliotecas para organizar espacios de información personal y para leer, escribir, enseñar, aprender y crear. Su trabajo intelectual será compartido con otros por medio de la biblioteca, y sus contribuciones e interacciones serán elementos de una biblioteca accesible global y universal que podrá ser usada por mucha gente diferente de muchas comunidades distintas [Schnase et al. 1994].

El uso de equipo de computo moderno para procesar información tiene dos efectos: por un lado facilita la generación, colección y almacenamiento de más información, pero por otro lado complica la tarea de absorber y usar dicha información; así que es necesario implementar técnicas útiles de búsqueda, filtrado y recuperación de datos.

Uno de los mayores retos en el desarrollo de bibliotecas digitales es incrementar la efectividad y la velocidad con la cual la información es buscada y recuperada. Esto permitirá avanzar en las áreas de estudio existentes, promoverá la interacción entre disciplinas y habilitará nuevas áreas de estudio.

Afortunadamente las técnicas de búsqueda de información están mejorando la manera en la cual la información es almacenada y procesada. En particular los equipos de procesamiento de información modernas y las bases de datos actuales pueden organizar y almacenar grandes cantidades de información y proporcionar un rápido acceso a los registros almacenados.

En la sección 3.5 se revisan algunos avances importantes en el área de recuperación de la información.

1.2 AGENTES

Existen otras clasificaciones de agentes incluyendo las de [Brustolini 1991], y [Nwana 1996].

El presente trabajo se centra en un tipo particular de agente que es el agente móvil o de red. Se pretende facilitar y agilizar la recuperación de información de bibliotecas digitales incorporando el concepto de agentes móviles.

1.2.3 *Uso de agentes en bibliotecas digitales*

Dado que "el principal propósito de una biblioteca digital es ofrecer servicios de información a usuarios finales en ambientes altamente distribuidos" [Sánchez y Leggett 1997]; es necesario introducir tecnologías que faciliten y agilicen los servicios que proporciona la biblioteca.

Una solución prometedora para el problema de localizar, recuperar y procesar grandes cantidades de información es el uso de agentes [Knoblock et al. 1994, Sánchez 1994]. En este sentido los agentes funcionan como asistentes del usuario. Las misiones que pueden ser delegadas a los agentes de una biblioteca digital incluyen [Sánchez 1994]:

Notificación al usuario cuando la información de interés halla sido

hallada o actualizada.

Filtrado de la información recuperada de acuerdo a las necesidades o preferencias del usuario.

Manejo de los procedimientos administrativos en la biblioteca (como procedimientos de pago y derechos de autor).

Recomendación a los usuarios basándose en el conocimiento que tiene de la biblioteca o en base al uso observado por otros usuarios o por contacto con otros usuarios (o agentes).

Las tareas específicas que un agente puede realizar en una biblioteca digital son:

El manejo de las peticiones particulares de los usuarios.

La selección del conjunto apropiado de fuentes de información.

La ejecución de estrategias efectivas de búsqueda sobre la red.

El despliegue y resumen de la información de varias maneras.

La diferenciación de los contenidos y la organización de colecciones particulares.

El entendimiento de las relaciones entre colecciones de datos.

La posesión de métodos alternativos para resumir y desplegar información.

Entre las tareas mencionadas se encuentra la búsqueda de información; sin embargo, buscar información distribuida en servidores remotos es uno de los mayores problemas en la programación de agentes, debido a la necesidad de tomar en cuenta tanto la gran cantidad de información involucrada como el hecho de tener repositorios de datos distribuidos.

El sistema ideal debe poseer tanto conocimiento específico del dominio en el cual opera y de los intereses del usuario, como la habilidad de filtrar datos basado en este conocimiento.

Además de estos problemas, se deben tomar en cuenta detalles técnicos inherentes a las redes de computo, como son: la velocidad de acceso en la red, la compatibilidad entre plataformas y arquitecturas y los problemas con las conexiones remotas (vía MODEM. De lo anterior puede apreciarse que la tecnología de agentes representa un gran desafío para la investigación en computación.

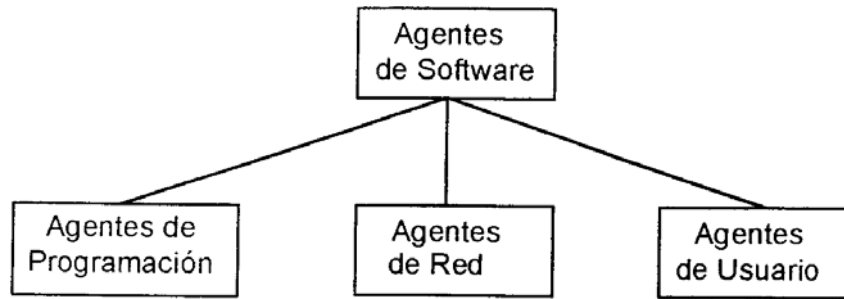


Figura 1.1 Clasificación de Agentes (Adaptada de [Sánchez 1997]).

1.3 ALCANCE DE LA INVESTIGACIÓN

En esta sección se presenta el panorama general bajo el cual se desarrolla este trabajo de tesis.

1.3.1 La Biblioteca Digital Florística

La Iniciativa de la Biblioteca Digital Florística (FDL) es un proyecto multiinstitucional y multidisciplinario que proporciona espacios de colaboración virtual que incluyen colecciones de información botánica actualizada y una variedad de servicios de biblioteca digital. Los proyectos Flora de Norte América (FNA) y Flora de China (FOC) forman parte de esta iniciativa.

Construir una biblioteca digital florística requiere la colaboración y comunicación de autoridades especializadas en la materia y de procesos complejos de revisión. Solo el proyecto de FNA involucra a más de 800 científicos, incluyendo taxonomistas y biólogos. Se pretende que FNA sea una biblioteca digital botánica distribuida, con información actualizada sobre aproximadamente 20,000 especies de plantas vasculares y briofitas recolectadas en el Norte de México, Estados Unidos, Cayos de la Florida, Islas Aleutianas, Canadá, Groelandia, Islas de Miquelon y Saint Pierre [FNA 1996; Schnase et al. 1997].

La biblioteca digital FDL consistirá de una gran colección de documentos en una gran variedad de medios y formatos incluyendo texto, mapas e ilustraciones, e integrará herramientas computacionales y servicios de biblioteca [Schnase et al. 1997].

Es de esperarse que un proyecto de tal magnitud oErexca retos y dificultades. Algunas interrogantes serian: ¿cómo mejorar la introducción masiva de datos?, ¿cómo lograr una buena digitalización de la información? ¿, Cómo recuperar información útil pare el usuario?

Esta tesis esta enfocada a tratar el problema de recuperar información útil

para el usuario en el ambiente de computo distribuido de FDL.

Como se ha mencionado en secciones anteriores una de las tareas mas cotidianas en el ambiente de bibliotecas digitales es la búsqueda y recuperación de información relevante, pero dada la gran cantidad de datos que se van a manejar en esta biblioteca, ¿cómo lograr que una persona encuentre la información deseada sin tener que pasar horas realizando búsquedas exhaustivas y sin tener que ser un experto en bases de datos? Que técnica de recuperación de información debe implementarse para agilizar la búsqueda y recuperación de datos? ¿De qué manera pueden incorporarse los agentes móviles a una biblioteca digital de estas características?

1.3.2 Objetivos del Proyecto

Una vez planteado el marco de trabajo en el que se desarrolla esta investigación, se puede mencionar los siguientes objetivos:

Estudiar el paradigma de agentes móviles y su aplicación en bibliotecas digitales.

Analizar el diseño e implementación de sistemas de agentes móviles ya existentes.

Definir tareas que puedan ser ejecutadas por agentes móviles en el marco de bibliotecas digitales florísticas.

Introducir el concepto de agentes móviles a la arquitectura de una biblioteca digital florística distribuida.

Desarrollar agentes móviles que permitan mejorar el funcionamiento de bibliotecas digitales.

Definir las tareas específicas que realizarán los agentes móviles tanto en los nodos cliente como en los nodos servidores que componen una biblioteca digital.

Plantear los problemas de comunicación y seguridad involucrados al implementar código que migra entre máquinas y de que manera pueden atenuarse o eliminarse dichos problemas.

1.3.3 Solución Desarrollada

La investigación presentada en este trabajo representa logros en dos sentidos. Primero, presenta un estudio sobre agentes móviles, sus características, ventajas y lenguajes de programación, entre otros aspectos. Y segundo, plantea una alternativa de solución al proceso de búsqueda, filtrado y recuperación de información de la Biblioteca Digital

Florística (FDL)

Para solucionar el problema de buscar, filtrar y recuperar información, se incorpora una clase de agente móvil de usuario a la arquitectura de FDL. Las instancias de agentes móviles migran a cada uno de los nodos que componen la biblioteca digital en busca de la información de interés para el usuario. El agente obtiene y filtra información relevante, descartando información irrelevante debido a que conoce previamente las preferencias del usuario. Una vez recorridos todos los nodos de la biblioteca (o los nodos especificados por el usuario), el agente regresa a su nodo origen con la información recolectada. Después de presentarle dicha información al usuario, este puede retroalimentar al agente "calificando" cada uno de los documentos recuperados, asignándole mayor peso a los documentos "buenos" y restándole peso a los documentos "malos" para de esta manera ir mejorando o afinando el proceso de recuperación. Finalmente, el agente trabaja de manera semi-autónoma debido a que el usuario puede en cualquier momento suspenderlo o desactivarlo.

1.4 ORGANIZACIÓN DEL DOCUMENTO

Las secciones restantes de este documento están organizadas de la siguiente manera: En el capítulo 2 se trata con mayor detalle la teoría existente sobre agentes móviles. Se describe algunos de los lenguajes que permiten desarrollar este tipo de agentes, las ventajas y desventajas de dichos lenguajes, y se desarrolla una tabla comparativa de acuerdo a varios criterios como el tipo de migración, el tipo de ligado, la estrategia de replicación usada y la estrategia de seguridad incorporada. Se mencionan también algunas de las posibles aplicaciones de este tipo de agentes. La construcción de bibliotecas digitales conlleva retos y dificultades. Algunas de las arquitecturas existentes, trabajos realizados en el área y algunos de los algoritmos de recuperación de información más usados en bibliotecas digitales, es lo que se presenta en el capítulo 3. El disco conceptual del sistema desarrollado se describe en el capítulo 4. En el capítulo 5 se presenta a detalle la implementación de dicho sistema, así como las tecnologías involucradas en el desarrollo del mismo. Finalmente en el capítulo 6, se resume el trabajo realizado, se expone el estado actual del proyecto, se plantea el trabajo a futuro y se presentan los alcances y las limitaciones encontradas para la realización del mismo.

Pérez Lezama, C. V. 1998. **Agentes Móviles en Bibliotecas Digitales**. Tesis Maestría. Ciencias con Especialidad en Ingeniería en Sistemas Computacionales. Departamento de Ingeniería en Sistemas Computacionales, Escuela de Ingeniería, Universidad de las Américas Puebla. Mayo. Derechos Reservados © 1998.