

## CAPÍTULO 3

### Trabajos relacionados

Para establecer el contexto de este documento, este capítulo describe los trabajos más sobresalientes que incorporan parte de los conceptos descritos en este capítulo y que se relacionan con EGRAI, el proyecto objeto de esta tesis. Se describen investigaciones, aportaciones de proyectos desarrollados en la Universidad de las Américas - Puebla y algunos sistemas comerciales.

#### 3.1 Trabajos relacionados

En relación con los métodos de recuperación de información citados en este capítulo, el sistema SMART es un sistema de búsqueda desarrollado por Gerald Salton en la Universidad de Cornell. Representa los documentos mediante el modelo de espacios vectoriales. La asignación de índices se realiza eliminando las palabras irrelevantes de una lista previamente determinada y aplica el algoritmo de Porter. Considera la asignación de pesos a los términos de los documentos [Salton 1971]. SMART continua siendo utilizado en nuestros días, a pesar de haber sido desarrollado hace ya casi tres décadas.

Los proyectos siguientes, son ejemplo de sistemas diseñados para trabajo en grupo que emplean comunicación asíncrona:

CAIS es un prototipo utilizado por un grupo de usuarios separados por zonas y horarios de trabajo, su objetivo es apoyar la inspección de software, (su nombre proviene de Collaborative Asynchronous Inspection of Software). Emula de forma asíncrona la estructura de la conversación en una reunión de

inspección. CAIS extiende las características del correo electrónico y de los pizarrones de avisos [Mashayekhi et al. 1994].

DIVA es un ambiente para trabajo en grupo que tiene elementos similares a los de una oficina física como personas, cuartos, escritorios o documentos que apoyan el trabajo colaborativo. La comunicación y la cooperación entre los integrantes del grupo se realiza de modo síncrono y asíncrono. En modo asíncrono, se manifiesta mediante el empleo de notas que se colocan en los objetos del ambiente. La cooperación con acceso coordinado a recursos compartidos de la oficina virtual apoya el trabajo por turnos. El sistema informa las últimas actividades de los usuarios y cambios en los documentos. DIVA está implementado en Lisp y emplea el protocolo TCP/IP [Chwelos 94].

El proyecto VSP (Virtual Science Park) surgió en la Universidad de Leeds, aunque actualmente se explota de forma comercial por la compañía VWS Ltd. El sistema proporciona servicios de consultoría en línea, apoya la creación de proyectos colaborativos y atiende ciertas demandas educativas. VSP emplea el concepto de un cuarto de recursos, el cual es un espacio virtual donde los estudiantes pueden almacenar su material y el de otros autores, o tienen acceso a recursos electrónicos de aprendizaje provistos por los profesores. Es un espacio para recibir asesoría y un lugar para trabajar en equipo. La comunicación en el cuarto de recursos es asíncrona [Lau et al. 1999].

A continuación, se enuncian algunos de los proyectos que emplean el paradigma de agentes en procesos de búsqueda y recuperación de información.

NEWT es un sistema cuyos agentes se encargan de filtrar noticias. Cuentan con una representación en la interfaz del usuario. Los agentes son

personajes cuyo vestuario se emplea para reflejar su comportamiento [Sheth 1994] y [Sheth y Maes 1993].

MAXIMS es un agente que asiste al usuario en el manejo de su agenda y correo electrónico. Es una caricatura con diferentes expresiones faciales que indican al usuario el grado de certeza cuando hace una predicción [Lashkari 1994].

Un asistente que apoya la navegación personal y de grupo es VISTABAR. Los integrantes del grupo cooperan de forma asíncrona y construyen estructuras de conocimientos relevantes para el propio grupo. El agente asigna un índice a las páginas de cada uno de los usuarios y lo asiste en la devolución. VISTABAR tiene una representación del usuario final, almacena información relacionada con las actividades de cada usuario y del grupo [Marais y Bharat 1997].

En el proyecto JITIR, (*Just In Time Information Retrieval*), los agentes tienen la misión de recuperar información de su ambiente en tiempo real. Se trata de un software que representa y recupera información basado en el contexto local del usuario y de forma pro activa y de manera no intrusa, presenta información valiosa al usuario. Su objetivo es reducir el esfuerzo cognitivo que requiere el usuario para encontrar, evaluar y acceder a la información.

Se maneja información con valor de uso potencial para situaciones posteriores. Para ello, los agentes observan el ambiente y pueden tomar acciones para modificarlo sin intervención directa del usuario. Los agentes en JITIR tratan de satisfacer una consulta que crean de forma automática que representa las necesidades del usuario. Son un medio para proveer información útil o de apoyo relevante para la realización de una tarea de recuperación, estableciendo un contexto [Rhodes 2000].

Los agentes dan al usuario sugerencias previas acerca del contenido de los documentos que desea recuperar. La información que se provee en este ambiente proviene de diversas bases de datos a cuyos datos se les ha asignado índices previamente.

El objetivo del programa DAML (Darpa Agent Markup Language), es desarrollar un lenguaje para representar las relaciones semánticas de una página web, en formas legibles para los agentes. El proyecto considera la inclusión de métodos para entender las fuentes dinámicas de información y la interoperabilidad entre agentes de forma semántica, considerando el contexto actual y posterior Internet [Lee et. al. 2001].

EDGES (**ED**ucational **aGENTS** **S**erver) es un sistema distribuido con agentes, permite al usuario acceder a colecciones de documentos de temas específicos de web, filtrando los documentos obtenidos a través de un método de búsqueda. Para extraer información relevante, utiliza el modelo de espacios vectoriales junto con una red neuronal tipo SOM, (Self - Organizing Map).

Surge como una alternativa en el contexto de inteligencia artificial para proveer algoritmos útiles para organizar o agrupar los datos, dada las características de adaptabilidad, aprendizaje y organización de las redes neuronales. En especial, una red SOM es una red neuronal artificial que trata de construir una representación de alguna característica del vector de entrada durante su etapa de aprendizaje. Se emplea para capturar el significado de las palabras, y construye un mapa de palabras denominado SOM semántica.

El filtrado de documentos es una tarea implementada con agentes, los cuales construyen una base de datos con documentos representados como vectores. EDGES está compuesto por servidores que contienen la solicitud del usuario. Si el usuario lo desea, los agentes consultan la Web utilizando AltaVista

y distribuyen los resultados a agentes especializados encargados de filtrar información.

Se emplea la red neuronal para establecer relaciones en el espacio de información constituido por la colección de documentos. El usuario consulta esta estructura con los agentes, mismos que usan el conocimiento almacenado en la red para filtrar los documentos recuperados en la búsqueda. El sistema está desarrollado en Java y con la tecnología aglets.

EDGES esta compuesto de servidores que contienen los componentes siguientes:

- Un conjunto de agentes maestros, uno para cada usuario. Un agente maestro ayuda al usuario a obtener la información requerida: obtiene la consulta del usuario y crea el agente de búsqueda para que busque en otros servidores EDGES. El agente maestro puede crear la interfaz de usuario y presentar los resultados al usuario.
  - Agentes de búsqueda, entregan la consulta a los servidores EDGES, junto con el URL de los documentos filtrados.
  - Un conjunto de agentes especializados, los cuales periódicamente reorganizan la base de datos y repiten la etapa de aprendizaje de la red neuronal.
  - Un directorio de otros servidores EDGES, usados por agentes maestros y agentes de búsqueda para saber las direcciones de los otros servidores EDGES [Rizzo et al. 1998].

El modelo BBD por las siglas de los autores y colaboradores (Belkin, Brooks, y Daniels), propone la idea de agentes como intermediarios entre la biblioteca y sus usuarios. Considera a los agentes como una colección de

expertos funcionales. Para satisfacer las necesidades de información de los usuarios, los agentes incluyen una descripción del problema, el modelo del usuario y una estrategia de recuperación.

Para la comunicación entre los agentes y el usuario se emplean enunciados escritos en lenguaje natural, mediante un lenguaje controlado. Es un sistema con un tiempo de respuesta lento, puede tardar incluso días para atender una solicitud del usuario [Logan et al. 1994].

Entre los proyectos en bibliotecas digitales que incorporan el paradigma de agentes se encuentran:

La biblioteca digital SARA, (*Synthetic Aperture Radar Atlas*), explora un método basado en agentes móviles para procesar datos de forma autónoma y descubrir conocimientos. Los principales componentes de la biblioteca son archivos distribuidos de imágenes de la Tierra [Yang et al. 2000].

En el proyecto "Modelo de una estructura de hipermedios basada en agentes para el diseño y desarrollo de bibliotecas digitales", se propone que las bibliotecas digitales sean diseñadas y desarrolladas como ambientes de información, compuestas por diferentes agentes que realizan tareas distintas, por ejemplo: agentes que apoyan en los servicios de recuperación de información, agentes que asesoran para proveer servicios de almacenamiento de información, entre otras [Salampasis et. al. 1996].

El proyecto UMDL, (University of Michigan Digital Library), emplea una arquitectura basada en agentes para brindar las características de modularidad, escalabilidad y extensibilidad. Los agentes tienen *autonomía*, (entendida en este contexto como la habilidad de razonar sobre el uso de los recursos de la biblioteca) y *negociación*, (considerada porque los agentes son entidades autónomas que negocian con otros agentes para ganar el acceso a los recursos y capacidades del sistema).

La arquitectura contempla un mantenimiento de los agentes, un corredor y equipos de formación de agentes para encontrar fuentes potenciales de información, asisten en el manejo de los servicios y en el comercio de bienes de información respetando la privacidad y propiedad intelectual [Birmingham 1995].

En nuestros días, los recursos tecnológicos como la web en ocasiones facilitan la búsqueda de material bibliográfico relevante. En [Holmberg 2000] se sugiere como un lugar apropiado para llevar a cabo esta actividad a los sitios de los editores.

En la Universidad de las Américas - Puebla, la iniciativa U-DL-A (University Digital Libraries for All), está dirigida hacia la construcción de colecciones o acervos digitales y a la provisión de servicios para el uso de éstas, con el propósito de atender las demandas de una comunidad académica universitaria [Sánchez y Leggett 1997]. EGRAI es parte de esta iniciativa.

En U-DL-A, se han realizado investigaciones previas en recuperación de información, groupware y agentes, que han contribuido al desarrollo de esta investigación. En seguida se enuncian.

MiEGA, (Espacios grupales de aprendizaje en la biblioteca digital de la UDLA-Puebla), es un ambiente para un grupo donde se encuentran y colaboran usuarios de bibliotecas digitales. El ambiente brinda un espacio agradable, un foro de discusión y apoya la construcción del conocimiento del grupo mediante el registro y la consulta de minutas [Pérez 2000].

El sistema VIAJERUS emplea recuperación personalizada de información distribuida. Utiliza las ventajas de las bibliotecas digitales al utilizar sus recursos de comunicación y colaboración mediante agentes móviles. En

este ambiente, los agentes son una opción para la creación de interfaces de usuario, para la generación de servicios personalizados y apoyan el desarrollo del cómputo distribuido [Chevalier 2000].

SYREX, es un sistema de Servicios de recomendaciones para usuarios de colecciones digitales de la biblioteca digital U-DL-A. Los agentes de SYREX emplean el modelo del usuario para buscar documentos de su interés en colecciones de la biblioteca. Considera un mecanismo de recomendación combinado: recomendación basada en contenido y recomendación colaborativa. Para recuperar información, implementa el modelo de espacios vectoriales [Ramírez 2001].

SERGEN, (Servidor General para la Biblioteca Digital de la Universidad de las Américas Puebla), es un proyecto que ofrece una interfaz al administrador mediante la cual puede dar mantenimiento a los servicios de agentes de usuario de la biblioteca digital, y una interfaz para el usuario final que le permite administrar, crear, eliminar o cambiar el estado de sus agentes [Cocoletzi 2001].

VRef (Sistema de Referencias Virtuales) [Sánchez et al. 2001], es un ambiente diseñado para apoyar la colaboración y construcción social de conocimiento, dirigido a referencistas y usuarios de bibliotecas digitales. En él, los usuarios solicitan referencias bibliográficas sobre algún tema de su interés. Se accede al sistema mediante el navegador. Los referencistas son los encargados de atender las solicitudes de los usuarios, aunque éstas pueden ser contestadas por otros usuarios del sistema, quienes desean compartir su experiencia con los demás según su especialidad. A éstos usuarios se les denomina referencistas voluntarios.

El usuario evalúa la lista de referencias bibliográficas, indicando si las referencias resultaron satisfactorias o no. Con la información de solicitudes

previas que han sido evaluadas como satisfactorias, se forma un acervo de conocimientos utilizado y enriquecido por todos los usuarios.

VRef facilita la transición del escritorio del referencia tradicional, a un papel más dinámico, el de un referencista digital [Sánchez et al. 2001]. Dada la importancia para mostrar los conceptos centrales del proyecto EGRAI, éste sistema se describe con mayor detalle en el capítulo siguiente.

En seguida, se mencionan algunos de los sistemas comerciales relacionados con esta investigación.

### **3.2 Sistemas Comerciales**

En esta sección se describen algunos de los sistemas comerciales que se relacionan con el proyecto EGRAI. Por razones de secreto comercial, los detalles de implementación no están disponibles.

AppleSearch es un sistema que recupera texto de ligas. Los agentes examinan este texto y asignan índices a los documentos [Jansen 1996].

Ebrary es un sitio en Internet dedicado a buscar referencias bibliográficas. Es una biblioteca en línea que permite buscar y consultar en colecciones de libros, mapas, periódicos o anuarios, dirigida principalmente a estudiantes, profesores y profesionistas. El sitio se originó a partir de la iniciativa de las editoriales Venture LLC, Pearson PLC y McGraw-Hill.

Los documentos de ebrary están protegidos por los derechos de autor correspondientes y pueden ser consultados completamente en línea, aunque si se desea obtener una copia total o parcial, se debe pagar por ello [en línea].

La compañía InfoWorks Technology mantiene el sitio Internet Library for Librarians, el cual emplea la metáfora de un centro comercial, donde se pueden localizar recursos para recuperar información en línea de almanaques, bibliografías, biografías, diccionarios, directorios o enciclopedias. El sitio ofrece herramientas para apoyar la administración o automatización de los principales departamentos y servicios de bibliotecas tradicionales. Fomenta el trabajo en equipo al formar grupos de bibliotecarios mediante listas de correo electrónico y asociaciones. Cuenta con un sistema de noticias [en línea].

Un sitio en donde los estudiantes pueden buscar información de una colección de libros es QUESTIA. Para ello, el usuario selecciona el tipo de material que desea, (libros, periódicos, o enciclopedias, elige una categoría que generalmente corresponde a un área de estudio, introduce las palabras clave e indica si desea que la búsqueda contemple sinónimos o antónimos de las palabras clave.

Cuando el usuario introduce la consulta, debe indicar si desea que aparezcan todas o algunas de las palabras clave en el documento por recuperar. Se le solicita que indique a qué parte pertenecen, (títulos, encabezados, autores, editores, ISBN o al texto), y que especifique un periodo de tiempo que contemple la fecha de edición de los documentos [en línea].

Los conceptos y trabajos relacionados presentados en el capítulo, sugieren algunas ideas para mejorar y extender los servicios de referencia que actualmente se ofrecen en la biblioteca de la Universidad de las Américas - Puebla. En el capítulo siguiente se presenta el diseño conceptual del ambiente EGRAI, dirigido hacia ese fin.