INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY
CAMPUS MONTERREY

**PROGRAMA DE GRADUADOS EN MECATRÓNICA,
TECNOLOGÍAS DE INFORMACIÓN**

NAVEGACIÓN HABLADA DE PDLib

TESIS

**PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL
GRADO ACADÉMICO DE:**

MAESTRO EN CIENCIAS EN TECNOLOGÍA INFORMÁTICA

POR:

KARLA PATRICIA SINTA COSME

MONTERREY, N.L. **DICIEMBRE 2008**

INSTITUTO TECNOLÓGICO DE ESTUDIOS SUPERIORES DE MONTERREY
CAMPUS MONTERREY

DIVISIÓN DE MECATRÓNICA Y TECNOLOGÍAS DE INFORMACIÓN

PROGRAMA DE GRADUADOS EN MECATRÓNICA,
TECNOLOGÍAS DE INFORMACIÓN

Los miembros del comité de tesis recomendamos que la presente tesis de la Ing. Karla Patricia Sinta Cosme sea aceptada como requisito parcial para obtener el grado académico de Maestro en Ciencias en Tecnología Informática.

COMITÉ DE TESIS

Dra. Martha Patricia Verdines Arredondo
Asesor principal

Dr. Juan Carlos Lavariega Jarquín
Sinodal

MC Martha Sordia Salinas
Sinodal

Dr. Joaquin Acevedo Mascarúa
Director de Investigación y Posgrado
Escuela de Ingeniería

Diciembre de 2008

NAVEGACIÓN HABLADA DE PDLib

POR:

KARLA PATRICIA SINTA COSME

TESIS

Presentada al Programa de Graduados en Mecatrónica, Tecnologías de Información

Este trabajo es requisito parcial para obtener el grado de

Maestro en Ciencias en Tecnología Informática

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS SUPERIORES DE MONTERREY

CAMPUS MONTERREY

Monterrey, N.L., Diciembre 2008

Dedicatoria

En memoria de mi Madre:
Hoy, mañana, eternamente, y aún después…
Vivirás en mí.
Porque eres la estrella que ilumina el camino de mi vida, y
Mi más grande inspiración…..
¡Te Amaré por siempre!

Agradecimientos

Escribir una tesis es un desafío inigualable: Todo un cúmulo de emociones que emanan durante esta larga trayectoria, y el esfuerzo continuo por vencer los obstáculos presentados, se conjuntan hasta llegar a la recta final. Desde el principio creí en este gran desafío. Hubo otras personas que también lo vieron así; es a ellos, en especial, a quienes debo la más profunda gratitud.

A Mi asesora de tesis: La Dra. Martha Patricia Verdines Arredondo por hacerme creer en la realización de este proyecto y por ser la persona quien me motivó para poderlo culminar. “Mi más sincero agradecimiento”.

Al Dr. Juan Arturo Nolazco y al Dr. David Garza, por creer en mí para la realización de este trabajo de tesis y por respaldarme siempre durante mis estudios de maestría.

A Mi coordinador de Maestría: El Dr. José Raúl Pérez Cázares. Quiero hacer un reconocimiento especial para él, quien siempre me apoyó para culminar mis estudios de posgrado. Gracias por sus conocimientos transmitidos y por toda la atención que ha puesto en mí. “Toda mi admiración y respeto”.

A Petry: Mi Ángel. Un gran ejemplo de vida y entereza. Gracias por toda su enseñanza y por todo su amor. “Ni la distancia, ni el tiempo harán que te olvide, porque en mi corazón permanecerás por siempre…Que Dios te aguarde”.

A Pechi y Mirna: Dios las tenga en su gloria y las bendiga, porque fueron y seguirán siendo una parte importante de mi vida.

A mis abuelos: Gracias por su amor y su entrega, pero sobre todo gracias por estar siempre conmigo. “Los Quiero Mucho”.

A mi papá: Gracias por dejarme volar y permitirme crecer, por su respeto y su cariño incondicional. “Te Quiero Mucho”.

A mis tíos: Gracias por hacerme parte importante de sus vidas.

A mis grandes amigos: Luis Angel, Adrián, Eilen, Ile, Chucho; los que ayer y hoy siguen estando conmigo. Sin su empuje y entusiasmo, yo no habría logrado lo que hasta ahora he ganado. “Espero contar con ustedes toda la vida”.

Mi gratitud especial a Flor, mi fiel amiga: “Gracias por tu respaldo y por el ánimo que me transmites en cada momento de mi vida, es una dicha enorme contar siempre contigo”.

Y, mi eterno agradecimiento a ese maravilloso ser divino: “Dios mío, gracias por permitirme este momento para expresarte mi gratitud, por esta grandiosa experiencia de vida, amor y fortaleza. Gracias por mi Fe”

Resumen

La era del cómputo móvil se expande cada vez más, gracias al advenimiento de tecnologías robustas de comunicaciones y el avanzado diseño de los nuevos equipos de mano o PDAs. Conectarse a las fuentes de información a través de la red, sin importar distancia y tiempo, es hoy una acción natural, y un ejemplo de ello lo constituye el acceso a las bibliotecas digitales a través de un dispositivo móvil. En los últimos años las bibliotecas digitales han despertado un gran interés, no sólo entre la comunidad de informáticos, sino también en la de bibliotecarios, por los grandes beneficios que estas ofrecen a sus múltiples usuarios. Considerando lo antes mencionado y tomando en cuenta el hecho de que los dispositivos móviles ofrecen un gran abanico de posibilidades para el acceso a la información y servicios en ambientes digitales, este proyecto está enfocado en la implementación de una interfaz de navegación hablada en dispositivos móviles para bibliotecas digitales personales, cuyos criterios de "búsqueda de información" permitan el reconocimiento de la voz independientemente del locutor, a través de un aprendizaje deductivo-inductivo, basado tanto en la transferencia de los conocimientos que un experto humano posee a un sistema informático, como en los conocimientos necesarios que el sistema pueda, automáticamente, conseguir a partir de ejemplos reales sobre la tarea que se desea modelar. Específicamente, para la implementación de la aplicación, en este proyecto, se utilizó como cliente móvil un dispositivo PDA de tipo Pocket PC de Microsoft y para efectos de la navegación, ésta fue basada en la infraestructura existente de la Biblioteca Digital PDLib.

Con esta implementación se logró establecer la navegación hablada para aplicaciones móviles, estableciendo un canal de comunicación apto, independientemente de las plataformas destinadas a cada una de las aplicaciones involucradas, de tal modo que cualquier dispositivo móvil de tipo PDA, sea capaz de interactuar con un sistema reconocedor de voz, sin importar el tipo de Sistema Operativo a utilizar, y sin importar también el lenguaje de programación utilizado para el diseño de las interfaces de navegación hablada, obteniendo como resultado el hecho de que cualquier dispositivo móvil con la tecnología implementada de un sistema de navegación de voz, pueda adaptarse de forma práctica a cualquier sistema de bibliotecas digitales en donde se permita la interacción de dispositivos móviles de tipo PDAs.

Tabla de Contenido

Lista de Figuras

Lista de Tablas

Capítulo 1

Introducción

Actualmente, los usuarios de servicios de información acceden a una gran variedad de contenido a través de diferentes sistemas de cómputo. Por tal motivo, la necesidad de dispositivos que puedan acceder a la información, desde cualquier parte y con disponibilidad en todo momento, es cada vez mayor. Los únicos dispositivos que reúnen estas características son los dispositivos móviles; hoy en día han evolucionado de tal forma que son ya parte esencial de la vida diaria de muchos usuarios, que generalmente navegan utilizando diferentes interfaces adaptadas según el tipo de móvil que estén utilizando.

Generalmente, la interfaz de usuario gráfica es el vehículo más común para facilitar una interacción humano-computadora. Sin embargo, para el caso particular de los dispositivos móviles, hay situaciones en las cuales estas interfaces son inadecuadas y la necesidad de utilizar aplicaciones existentes en una modalidad no visual es clara. Un ejemplo, es cuando cierta tarea requiere que la atención visual del usuario esté dirigida en alguna otra parte con excepción de la pantalla de su computadora. Otro ejemplo es cuando el usuario es invidente o discapacitado visualmente o bien, el uso de aplicaciones con el espacio limitado de la pantalla, lo que provoca una mayor interacción por parte de los usuarios, y mayor inversión de tiempo tan solo al recorrer una pantalla, dado que muchas de las interfaces no son de tan fácil acceso. (Boyd, Boyd & Vander-heiden, 1990; Buxton, 1986).

Arons et al. (1995), señala que, éste es el momento correcto para la integración de voz y audio dentro de interfaces de usuario y aplicaciones. Esto es particularmente verdadero para los dispositivos móviles tales como PDAs donde el despliegue es típicamente uno de los componentes más costosos, y frecuentemente estipula el tamaño del dispositivo en conjunto. Una interfaz auditiva podría bajar el costo de este tipo de dispositivos, hacerlos realmente de acuerdo al tamaño de una palm, y permitir que sean utilizados a pesar de que exista movilidad (por ejemplo, mientras que conduce o cultiva un huerto sin ser forzado a mirar una pantalla).

De este modo, el audio está llegando por sí mismo como un medio para interactuar con los dispositivos de cómputo móvil. Esto es debido en parte al progreso tecnológico y en parte a la apreciación intensificada del uso multimedia para el enriquecimiento de presentaciones auditivas. En la actualidad, ha habido un considerable progreso en el diseño y la implementación de las interfaces de voz y audio. Sin embargo, en un diseño de interfaz de usuario tradicional, la voz y el audio para la interacción con el usuario se han pensado

como algo adicional a las interfaces de usuario gráficas, o como un substituto deficiente cuando las GUIs (Graphical User Interfaces) no están disponibles (Arons et al., 1995).

La meta de proporcionar el acceso no visual a las interfaces gráficas puede sonar como una expresión contradictoria. Son numerosos los diseños de interfaces que surgen de la traducción de una interfaz interactiva, representada espacialmente y compacta visualmente, dentro de una interfaz no visual eficiente, intuitiva y no intrusiva (Edwards & Mynatt, 1994).

1.1 Planteamiento del Problema

Los dispositivos móviles ofrecen un gran abanico de posibilidades para el acceso a la información y servicios en ambientes digitales. Sus características son muy distintas a las de una PC (el tamaño, las interfases, el modo de interacción, etc.) y también lo es el uso que se les da. Sin embargo, el reducido tamaño de las terminales móviles implica necesariamente que la interacción mediante interfases tradicionales, como el teclado, resulte incómoda. Si al tamaño de las terminales se agrega su actual complejidad, se puede observar que muchas veces en ellas no hay espacio físico para poner botones o desplegar menús.

El habla es la forma de comunicación natural de los seres humanos y no cabe duda que si se provee de interfaces habladas a las máquinas, se podría mejorar de forma sustancial el modo de interactuar con los dispositivos, además de facilitar el acceso a los sistemas y de reducir el tiempo de aprendizaje para su uso. El habla puede ser también un elemento que facilite que determinados grupos de personas, como discapacitados o ancianos, puedan acceder a sistemas y aplicaciones (Alvarez, 2006).

En el marco de este proyecto se han considerado diversos aspectos, con la finalidad de definir, desarrollar y evaluar una aplicación que sirva como referencia para la realización de soluciones móviles de datos accesibles mediante el habla.

1.2 Objetivos

El objetivo de este proyecto se centra en implementar una interfaz de navegación hablada en dispositivos móviles (específicamente PDAs) para bibliotecas digitales personales, cuyos criterios de "búsqueda de información" permitan el reconocimiento de la voz, a través de un aprendizaje deductivo-inductivo, basado tanto en la transferencia de los conocimientos que un experto humano posee a un sistema informático, como en los conocimientos necesarios que el sistema pueda, automáticamente, conseguir a partir de ejemplos reales sobre la tarea que se desea modelar, con el apoyo de componentes de reconocimiento de los ya existentes.

1.3 Contribuciones

De forma muy general, las contribuciones de este proyecto de investigación, se centralizan principalmente en dos aspectos fundamentales:

1. La adaptación de una interfaz de voz para dispositivos móviles PDAs, que sea de fácil acceso para el usuario, y al mismo tiempo que proporcione la ayuda necesaria para que el usuario pueda realizar su navegación sencillamente.
2. La adaptación de un sistema de reconocimiento de voz, que pueda ser invocado independientemente de la plataforma sobre la que se encuentre la aplicación de bibliotecas digitales personales.

1.4 Limitaciones

Existen algunos factores que pudieran ser considerados como limitaciones dentro de la realización de este proyecto, inclusive, algunos expertos no apuestan mucho al desarrollo de la navegación hablada, dado que por sus características de uso, este tipo de aplicaciones se centra en un número reducido de usuarios. Este tipo de sistemas de navegación hablada tiene la gran desventaja de no tener aún un diseño para "usuarios independientes", por lo que el tener que entrenar el sistema de reconocimiento de voz, para cada uno de los usuarios, resulta una tarea bastante compleja desde el tiempo que se le invierte al entrenamiento de los comandos de voz, hasta la infraestructura diseñada para montar un reconocedor de voz eficiente.

Pasando a la parte de limitaciones en el uso de la biblioteca digital de PDLib, principalmente para este proyecto, existe una limitante absoluta que típicamente se refleja en su arquitectura propuesta; con esto lo que se quiere dar a entender es que la implementación de la aplicación de navegación hablada, por los componentes de software involucrados, puede interactuar únicamente con el entorno de PDLib, es decir, a pesar de que cualquier dispositivo móvil de tipo PDA, cumple con las características adecuadas para interactuar con el reconocedor de voz; por el lado de PDLIb existe una arquitectura específica para solo unos cuantos dispositivos con características específicamente definidas tanto de hardware como de software, adaptables exclusivamente al entorno de PDLib, por lo que con esto se restringe a que otras bibliotecas digitales, que no posean la infraestructura de PDLib, sean adaptables a la arquitectura de navegación hablada de este proyecto.

1.5 Definiciones

Esta sección describe los conceptos clave utilizados en este proyecto de investigación.

Biblioteca Digital. Es una colección organizada de documentos digitales que ofrece diversos servicios a los usuarios, como envío, clasificación, búsqueda, recuperación y administración; facilitando actividades de estudio e investigación colaborativa entre usuarios distribuidos geográficamente.

Dispositivos Móviles: Son aparatos pequeños, con algunas capacidades de procesamiento, móviles o no, con conexión permanente o intermitente a una red, con memoria limitada, diseñados específicamente para una función, pero que pueden llevar a cabo otras más generales.

Interfase: Es el dispositivo de hardware o protocolo de programación encargado de realizar la adaptación que haga posible la conexión entre dos sistemas o elementos de la unidad central de procesamiento, entre unidades o con el usuario.

Interfaz: es la forma en que los usuarios pueden comunicarse con una computadora, y comprende todos los puntos de contacto entre el usuario y el equipo.

Interfaz Auditiva: Es el espacio de la interacción entre usuarios y sistemas donde se desarrollan los intercambios entre aplicaciones de audio y habla.

Interfaz de Usuario Gráfica: Es el artefacto tecnológico de un sistema interactivo que posibilita, a través del uso y la representación del lenguaje visual, una interacción amigable con un sistema informático.

Multimedia: Es un término que se aplica a cualquier objeto que usa simultáneamente diferentes formas de contenido informativo como texto, sonido, imágenes, animación y video para informar o entretener al usuario.

Navegación: Se utiliza el término navegación para describir la exploración de una aplicación, como una página Web, saltando de un punto a otro de la página, o de una página a otra.

Palm: Computadora de pequeño tamaño, algo mayor que un paquete de cigarrillos, que se puede llevar en la palma de la mano (palm) y que, además de otras funciones, permite la conexión con Internet.

PC: Computadora personal que permite la interacción de un usuario con un sistema a la vez.

PDA: Es una computadora de mano con un sistema de reconocimiento de escritura. Hoy día se puede usar como una computadora doméstica (ver películas, crear documentos, juegos, correo electrónico, navegar por Internet, reproducir archivos de audio, etc.).

PDLib: El Concepto PDLib (Personal Digital Library) propone un Sistema de Librería Digital con Acceso Universal. Este es "personal" en el sentido de que cada usuario es proveído por un repositorio cuya documentación es de propósito general. Se dice que es de "Acceso Universal" ya que permite a los usuarios acceder a su biblioteca digital personal desde la mayoría de los dispositivos de cómputo conectados a Internet, incluyendo dispositivos móviles.

Servicios de Información: Realización de un determinado proceso tecnológico de la "actividad de información"' y la entrega a los usuarios de los resultados de dicho proceso, con el fin de satisfacer sus necesidades de información.

Capítulo 2

Marco Teórico

Este capítulo está dividido en varias secciones. La primera de ellas describe de forma general el contexto de las Bibliotecas Digitales, en donde, además de mencionar sus principales características y los servicios que pueden ofrecer, también se definen sus retos y tendencias, mencionando así mismo los componentes involucrados. En la segunda sección de este capítulo se presenta una descripción sistematizada de los estilos de interacciones existentes, así como de cada una de las interfaces de usuario que han sido diseñadas a lo largo de los años, para concretar con el tipo de interfaz y el estilo de interacción utilizados en este proyecto, presentando además un análisis comparativo de los dos tipos de interfaces claves en el diseño de la aplicación. En la tercera sección de este capítulo se especifica una descripción detallada del concepto de Cómputo Móvil; y para concluir, en la cuarta sección se hace una recopilación, en un contexto general, de algunos sistemas de reconocimiento de voz, a manera de ejemplificar el modelo de navegación hablada que se utiliza en dispositivos móviles.

2.1 Bibliotecas Digitales

La palabra biblioteca tiene su origen en dos voces griegas: *biblio*, que significa libro, y *tekes*, que quiere decir, caja. Por consiguiente, una biblioteca, partiendo de sus voces de origen, es un espacio en donde se guardan textos. Tradicionalmente se interpreta este concepto como una colección privada o pública de libros, pero generalmente se relaciona con inmensas colecciones de texto, administradas por instituciones privadas o públicas (Flores, 2006).

Partiendo de lo anterior, una biblioteca digital puede ser vista como una abstracción del concepto original de biblioteca en el ámbito de información electrónica.

Una definición formal de biblioteca digital se describe a continuación:

"*Biblioteca Digital es una colección organizada de documentos digitales que ofrece diversos servicios a los usuarios, como envío, clasificación, búsqueda, recuperación y administración; facilitando actividades de estudio e investigación colaborativa entre usuarios distribuidos geográficamente"* (Garza et al., 2004).

Algunas características han sido obtenidas de varias discusiones acerca de bibliotecas digitales (Arms, 1995; Graham, 1995a; Chepesuik, 1997; Lynch & García-Molina, 1995):

1. Las bibliotecas digitales son la cara digital de las bibliotecas tradicionales que incluyen tanto colecciones digitales como tradicionales, abarcando material en papel y electrónico.

2. Las bibliotecas digitales incluyen material digital que existe fuera de las fronteras físicas y administrativas de cualquier biblioteca digital.

3. Las bibliotecas digitales incluyen todos los procesos y servicios que constituyen la columna vertebral y el sistema nervioso de las bibliotecas. Sin embargo, tales procesos tradicionales, a pesar de que constituyen la base del trabajo de la biblioteca digital, deben ser revisados y mejorados para adecuar las diferencias entre los nuevos medios digitales y los medios fijos tradicionales.

4. Las bibliotecas digitales idealmente proveen una vista coherente de toda la información contenida en una librería, sin importar su forma o formato.

5. Las bibliotecas digitales apoyan tanto a las comunidades particulares, como a los distritos electorales, aún y cuando estas comunidades puedan ser dispersadas extensamente a través de la red.

6. Las bibliotecas digitales deben integrar las habilidades de los bibliotecarios y de los informáticos, para poder ser factibles.

Las bibliotecas digitales están construidas por una comunidad de usuarios. Sus capacidades funcionales apoyan las necesidades de información y las aplicaciones de esa comunidad. Básicamente, se puede decir que la biblioteca digital es una extensión, una mejora e integración de una variedad de instituciones de información como lugares físicos en donde los recursos son seleccionados, coleccionados, organizados, preservados y accedidos en apoyo de una comunidad de usuarios (Fox & Sornill, 1999).

A continuación se describen algunos servicios básicos que ofrecen las bibliotecas digitales (Adam et al., 1996).

Creación de documentos digitales. Este servicio permite la creación de un documento digital a partir de la conversión del mismo documento almacenado en otro formato. Esto se hace con la finalidad de tener disponible diferentes versiones (formato) del mismo documento (Ej.: pdf, txt, doc).

Clasificación e indexamiento. Los documentos almacenados (en sus diferentes formatos) en la biblioteca digital deben ser clasificados (al igual que en una biblioteca normal) e indexados periódicamente, con la finalidad de mantener los servicios de búsqueda con la información más actualizada (cada vez que se agrega un nuevo documento, este debe ser indexado).

Búsqueda y recuperación. Una biblioteca digital debe proporcionar servicios de búsqueda y recuperación de documentos de manera fácil e intuitiva para el usuario. Los mecanismos de búsqueda pueden ser variados, pero en la mayoría de las bibliotecas digitales se permite buscar por palabras contenidas en el documento (búsqueda básica) y en los meta datos del mismo (búsqueda avanzada).

Distribución. Los usuarios de la biblioteca digital deben disponer, de forma rápida y segura, de los documentos almacenados.

Administración y control de acceso. Una biblioteca digital debe contar con un sistema de control de acceso a los documentos, así como de un método fácil de administración y configuración de usuarios y características de la biblioteca digital.

Personalización. Las bibliotecas digitales deben proporcionar personalización para satisfacer las preferencias de los usuarios y/o grupos de usuarios.

Existen también las bibliotecas digitales que proporcionan los servicios tradicionales de una biblioteca digital, pero con la diferencia de que a cada usuario se le proporciona un repositorio de almacenamiento de documentos personales, se le permite personalizar la clasificación de documentos (crear, recuperar, renombrar y eliminar colecciones) y también se les permite interactuar con otras bibliotecas digitales (colectivas o personales) (Escoffié, 2006).

La creación de bibliotecas digitales efectivas representa serios retos. La integración de medios digitales dentro de colecciones tradicionales no es sencilla, comparada con otros medios de publicidad, debido a su naturaleza única de información digital (esto es menos rígido, fácilmente copiado y remotamente accesible para múltiples usuarios simultáneamente). Algunos de los más importantes retos que encaran el desarrollo de bibliotecas digitales se enlistan a continuación (Cleveland, 1998).

Arquitectura Técnica. Las bibliotecas digitales necesitan mejorar y actualizar arquitecturas técnicas para adaptar material en formato digital. La arquitectura debe incluir componentes tales como:

- Redes locales de alta velocidad y rápida conectividad.
- Bases de Datos relacionales que soportan una gran variedad de formatos digitales.
- Motores de búsqueda con texto completo para indexar o proveer acceso a los recursos.
- Una variedad de servidores como servidores Web y servidores FTP.
- Funciones de administración de documentos electrónicos que ayudarán en la administración de los recursos digitales.

Con un coordinado esquema de biblioteca digital, serán requeridos también algunos estándares comunes, para permitir que las bibliotecas inter operen y compartan recursos. El problema sin embargo, es que a través de múltiples bibliotecas digitales, exista una amplia

diversidad de diferentes estructuras de datos, motores de búsqueda, interfaces, vocabularios controlados, formatos de documentos, y así sucesivamente. Debido a esta diversidad, sería un esfuerzo imposible, federar todas las bibliotecas digitales, tanto nacional como internacionalmente. De este modo, la primer tarea sería encontrar razones suficientes para federar bibliotecas digitales particulares en un sistema.

Colecciones digitales constructivas. Uno de los más grandes problemas en el incremento de bibliotecas digitales es la construcción de colecciones digitales. Obviamente para cualquiera que esté disponible, esto debe tener eventualmente una colección digital crítica para hacerlo verdaderamente útil. Existen esencialmente tres métodos para construir colecciones digitales (Cleveland, 1998):

1. *Digitalización.* Convertir documentos y otros medios de colecciones existente a forma digital.
2. *Adquisición de trabajos digitales originales.* Creados por publicistas y escolares. (Por ejemplo: libros electrónicos, journals, y datasets.
3. *Acceso a material externo.* Aunque este método puede no exactamente constituir parte de una colección local, es también un método para aumentar la cantidad del material que será disponible para los usuarios locales. Uno de los principales problemas es conocer el grado en el cual las bibliotecas convertirán material existente y adquirirán trabajos digitales originales, en lugar de señalarlos simplemente como información externa. Con base en esto, existen algunos factores en los que podrían basarse los coleccionistas del material específico, procesado por una institución determinada:

    - *Fortaleza de la colección.* Una biblioteca particular con una colección fuertemente enfocada podría ser la responsable de digitalizar porciones selectivas y adicionar nuevos trabajos digitales.
    - *Colecciones Únicas.* Si una biblioteca tiene solo copias de otras, existe obviamente una razón para digitalizar.
    - *Prioridades de las comunidades de usuarios.* Tales prioridades justificarán la contención de material local.
    - *Porciones manejables de colecciones.* Cuando no existe un criterio primordial, entonces el material puede ser dividido entre las instituciones, simplemente acordando lo que es razonable para cualquier institución que colecciona o digitaliza.
    - *Arquitectura técnica.* El estado de una biblioteca de arquitectura técnica será también un factor para seleccionar lo que será convertido. Una biblioteca debe tener una arquitectura técnica hasta la tarea de la ayuda de una colección digital particular.
    - *Habilidades del personal.* Las Instituciones cuyo personal no tiene las habilidades necesarias, no puede convertir un nodo importante en un esquema nacional.

Digitalización. Lo que este término significa, puesto simplemente, es la conversión de cualquier medio fijo o análogo (libros, artículos, fotos, pinturas, micro-formas) en formato

electrónico, por medio del escaneo, muestreo, o de hecho aún re-tecleando. Un obvio obstáculo para la digitalización es que esto es muy costoso (Cleveland, 1998).

Existen muchos enfoques para decidir qué partes de una colección se deben digitalizar:

- Conversión retrospectiva de colecciones. Esencialmente iniciando en la A y finalizando en la Z.
- Digitalización de una particular colección especial o de una porción.
- Lo más destacado de una colección variada, digitalizando particularmente buenos ejemplos de algunas colecciones fuertes.
- Materiales de alto uso, haciendo a aquellos materiales que tienen mayor demanda, más accesibles.
- Un enfoque "ad-hoc" en donde se digitalizan y almacenan los materiales, según como sean requeridos.

Metadatos. El metadato es un dato que describe el contenido y los atributos de cualquier objeto particular en una biblioteca digital, el cual es importante para describir los recursos y el uso de cualquier documento. Sin embargo, mientras que existen estándares formales de metadatos para bibliotecas, tales expedientes desperdician mucho tiempo al crear y requerir especialmente personal entrenado, por lo que la falta de estándares ideales para metadatos comunes, definidos para su uso en algún contexto específico es otra barrera más al acceso y uso de la información en una biblioteca digital, o en un coordinado esquema de biblioteca digital (Cleveland, 1998).

Nombramiento, identificadores y persistencia. El nombramiento es un asunto que va ligado a los metadatos. Los nombres son cadenas que únicamente identifican objetos digitales, y forman parte de cualquier documento de metadatos. Los nombres son tan importantes en una biblioteca digital, como lo es un número de ISBN en una biblioteca tradicional. Es necesario asignar nombres adecuados para identificar únicamente objetos digitales para propósitos como:

- Citas
- Recuperación de información
- Para hacer links entre objetos,
- Para propósito del manejo de derechos reservados.

Por otro lado, un esquema de identificadores únicos es requerido; uno que tenga persistencia más allá de la organización que se origina y que no esté atado a localizaciones o procesos específicos. Estos nombres deben seguir siendo válidos siempre que los documentos se muevan de una locación a otra, o sean migrados de un medio de almacenamiento a otro.

El tema de nombramiento persistente, es un problema organizacional, en lugar de un problema de ingeniería. Técnicamente, es posible un sistema manejador de nombres, sin embargo, los identificadores únicos solo persistirán si alguna institución toma la responsabilidad de su administración y migración desde una tecnología actual hasta

generaciones sucesivas de tecnologías. Así una meta de un esquema de biblioteca digital coordinada sería identificar una institución o instituciones, que se encargarían de publicar, resolver y migrar un sistema de nombres únicos (Cleveland, 1998).

Derechos reservados/administración de derechos. El problema para las bibliotecas es que, a diferencia de los negocios privados o de los editores que poseen información propia, son, en general, simplemente "vigilantes de la información"; dado que no poseen derechos reservados del material con el que cuentan. Es inverosímil que las bibliotecas digitalicen y provean acceso libremente a los materiales que cuentan con derecho de autor en sus colecciones. En lugar de esto, tendrán que desarrollar mecanismos para la administración de los derechos reservados, mecanismos que les permitan proveer información sin la violación de tales derechos, conocido como manejo de derechos.

Algunas funciones en el manejo de los derechos, podrían incluir por ejemplo:

- Rastreo de usabilidad.
- Identificación y autenticación de usuarios.
- Proporcionar el estatus de los derechos reservados de cada objeto digital, y las restricciones en su uso.

Preservación. En la preservación de materiales digitales, el problema real es la obsolescencia técnica. La obsolescencia técnica en la era digital es como la deterioración del papel en la era del formato impreso. Las bibliotecas en la era pre-digital tuvieron que preocuparse por la climatización y por la des-acidificación de libros, sin embargo, la preservación de la información digital significará constantemente prometer nuevas soluciones técnicas. Existen tres tipos de preservación:

1. *La preservación del medio de almacenamiento.* Después de un largo tiempo, los materiales almacenados en medios más viejos, podrían perderse debido a que no contarán más con el soporte de hardware o software para ser leídos. Así, las bibliotecas tendrán que seguir moviendo la información digital de un medio de almacenamiento a otro.
2. *La preservación de acceso al contenido.* Mientras que los archivos pueden ser movidos de un medio de almacenamiento físico a otro, algunos formatos se vuelven obsoletos. Este es un problema mucho mayor al de las tecnologías de almacenamiento obsoletas. Una solución es hacer una migración de los datos, es decir, traducir los datos de un formato a otro, conservando la habilidad de los usuarios para recuperar y desplegar el contenido de la información. Hasta el momento nadie sabe realmente cual es la mejor forma de migrar la información digital. Inclusive, si actualmente existiera una tecnología adecuada, la información tendrá que ser migrada de un formato a otro, por muchas generaciones, pasando una enorme y costosa responsabilidad a las que vengan después.
3. *La preservación de materiales de medios fijos a través de tecnología digital.* Esta inclinación involucra el uso de tecnología digital como un remplazo para los medios de preservación actual, tales como las micro-formas. Una vez más,

no existen, todavía, estándares comunes para el uso de medios digitales como un medio de preservación y esto es confuso si los medios digitales tienen hasta ahora la tarea de preservación a largo plazo. Los estándares de preservación digital serán requeridos para almacenar y compartir constantemente los materiales digitalmente preservados.

En un sistema coordinado, las bibliotecas conjuntamente pueden:

- Crear políticas para la preservación a largo plazo.
- Asegurarse de que las copias permanentes redundantes sean almacenadas en instituciones designadas.
- Ayudar a establecer los estándares de preservación para almacenar y compartir constantemente los materiales digitalmente preservados.

Fox & Sornill (1999), refieren otros tipos de retos que consideran primordiales en el éxito que puedan tener las bibliotecas digitales para alcanzar altos niveles de efectividad, y al mismo tiempo poder proporcionar facilidades de uso a una comunidad diversa. Estos retos están enfocados básicamente en la recuperación de la información para el desarrollo de las bibliotecas digitales, como se menciona a continuación:

- *Tratar con colecciones de información que se encuentren distribuidas en la naturaleza.* Esto es uno de los requerimientos fundamentales para la tecnología de bibliotecas digitales, de este modo el manejo apropiado de tales colecciones es un problema desafiante.
- *Trabajar con un número de bibliotecas digitales, cada una construida separadamente, de tal modo que los sistemas de información sean verdaderamente heterogéneos.* La integración requiere soporte por lo menos de algún escenario popular para los sistemas que esperan diferentes tipos de cadenas de comunicación (Ejemplo: responder a diferentes protocolos y lenguajes de querys); se requiere también que exista una variación de tipos de cadenas y estructuras, y que se combinen ambas en términos de representaciones de datos y metadatos. Para enfrentar este problema, un acercamiento ha sido desarrollar un lenguaje de descripción para cada biblioteca digital y para construir sistemas de búsqueda federada que puedan interpretar este lenguaje de descripción. Sin embargo, cuando el contenido de la biblioteca digital es altamente complejo, existe la necesidad de enriquecer lenguajes de descripción y sistemas más poderosos para interpretar y soportar las operaciones.

2.2 Interfaces de Usuario

Sharp, Rogers & Preece (2007), definen que el Diseño de Interacción consiste en diseñar productos interactivos para apoyar la forma en que las personas se comunican e

interactúan en su vida diaria y en su vida laboral. Este se basa en una comprensión de capacidades y deseos de las personas y en los tipos de tecnologías disponibles para el diseño de la interacción, así como el conocimiento relacionado a cómo identificar los requerimientos y su evolución dentro de un diseño adecuado.

Winograd (1997), señala que el Diseño de Interacción se basa en el diseño de espacios para la comunicación e interacción humana.

Thackara (2001), afirma que el Diseño de Interacción es el porqué y el cómo de nuestras interacciones diarias usando computadoras. Esto está relacionado con la forma en cómo crear, con calidad, las experiencias de los usuarios y requiere tener en cuenta un número de factores interdependientes, incluyendo el contexto de uso, el tipo de actividades, las diferencias culturales y los grupos de usuarios.

El diseño de interacción es fundamental para todas las disciplinas, campos y enfoques que están relacionados con la investigación y el diseño de sistemas de cómputo para las personas.

Una forma de conceptualizar el espacio de diseño se define en términos de las interacciones del usuario, con un sistema o producto. Esto puede ayudar a los diseñadores a formular un modelo conceptual para determinar qué tipo de interacción usar, y porqué, antes de definir una interfaz particular. Sharp, Rogers & Preece (2007), sugieren cuatro estilos fundamentales de interacción, que no tienen que ser mutuamente excluyentes.

1. **Interacción Basada en Comandos.** Donde los usuarios emiten instrucciones a un sistema. Esto puede ser dado de varias maneras: tecleando comandos, seleccionando opciones de menús en un ambiente de ventanas o en un "touch screen", hablando en voz alta, presionando botones o usando una combinación de teclas de función.

2. **Interacción Basada en Diálogo.** Donde los usuarios tienen un diálogo con un sistema. Los usuarios pueden hablar a través de una interfaz o haciendo preguntas de cierto tipo, para las cuales el sistema responde a través de texto o salida del habla.

3. **Interacción Basada en Manipulación.** Donde los usuarios interactúan con objetos en un espacio virtual o físico, manipulando tales objetos. Por ejemplo, abriendo, sosteniendo, cerrando, colocando.

4. **Interacción Basada en Exploración.** Donde los usuarios se mueven a través de un ambiente virtual o un espacio físico. Los ambientes virtuales incluyen mundos en tercera dimensión y sistemas de realidad virtual.

Existen muchos tipos de interfaces que pueden ser usadas para diseñar las experiencias del usuario. Algunos de los tipos están enfocados principalmente en una función (por ejemplo, para ser inteligentes, para ser adaptativas, para ser ambientadas), mientras que

otras se enfocan en el estilo de interacción usado (por ejemplo, comandos, graficas, multimedios), el dispositivo de entrada/salida usado (por ejemplo, PC, microwave). La Tabla 2.1 muestra varias décadas en que fueron desarrollados diferentes tipos de interfaces, así mismo se presenta una breve descripción de cada una de ellas.

TIPO DE INTERFAZ	DÉCADA	DESCRIPCIÓN
Command	1980's	La línea de comandos maneja interfaces que requieren que el usuario teclee comandos que son típicamente abreviaciones y combinaciones de teclas. Algunos comandos son también una parte fija del teclado, mientras que otras teclas de funciones pueden ser programadas por el usuario como comandos específicos.
WIMP/GUI	1980's	Permiten al usuario interactuar con un sistema. Específicamente, se vuelven posibles, nuevas formas de visualizar el diseño, como el uso del color, la tipografía y las imágenes. El WIMP original comprende: Ventanas, Iconos, Menús y dispositivos de punteros.
Advanced Graphical	1990's	Incluye animaciones interactivas, multimedios, ambientes virtuales y visualizaciones. Muchos usuarios afirman que este tipo de interfaz posee grandes beneficios comparados con la Interfaz de Usuario Gráfica tradicional.
Web	1990's	Enfocada en una estructura de información bien definida para que el usuario pueda navegar y acceder fácil y rápidamente. Provee hyperlinks hacia diferentes sitios o páginas de texto.

Tabla 2.1: Tipos de Interfaces.

TIPO DE INTERFAZ	DÉCADA	DESCRIPCIÓN
Speech (voice)	**1990's**	Es aquella en donde el usuario habla con un sistema que tiene una aplicación de lenguaje de habla como un servicio de telefonía. Comúnmente usada para investigar a cerca de información específica de algo. Esto es una forma específica de interacción de lenguaje natural que se basa en el tipo de interacción de conversación, en donde los usuarios hablan y escuchan a través de una interfaz (en lugar de tener que teclear o escribir sobre la pantalla). La tecnología de "speech" también ha avanzado en aplicaciones que pueden ser usadas por personas con discapacidades, incluyendo las aplicaciones de procesadores de palabras de reconocimiento de voz, lectores de web y software de reconocimiento de voz para operar sistemas de control de casa, incluyendo luces, TV, estéreo y otros aparatos del hogar.
Pen, Gesture, and Touch	**1990's**	Diseñadas para permitir a las personas escribir, dibujar, seleccionar y mover objetos utilizando dispositivos de entrada. Las "touchscreens" han sido diseñadas para permitir a los usuarios usar la punta de sus dedos para seleccionar opciones en una interfaz y mover objetos alrededor de un tablero interactivo. Usando diferentes formas de entrada, se pueden permitir mayores grados de libertad para la expresión del usuario y la manipulación de los objetos.
Appliance	**1990's**	Incluye electrodomésticos para el hogar, lugares públicos o carros (ejemplo: máquinas de lavado, fotocopiadoras) y productos de consumo personal (ejemplo: MP3 player, reloj digital, cámara digital). La mayoría de las personas que las usa para obtener algo específico dado en un período corto de tiempo.

Tabla 2.1: Tipos de Interfaces (continuación).

TIPO DE INTERFAZ	DÉCADA	DESCRIPCIÓN
Mobile	**2000´s**	Son diseñadas para dispositivos portátiles, tales como los PDAs y los teléfonos celulares. En particular los PDAs se han expandido considerablemente y son ahora comúnmente usados en restaurantes para tomar la orden, en rentas de carros para checar la hora de devolución del auto, etc. Un número de controles físicos han sido desarrollados para interfaces móviles, la preferencia y la habilidad para usar estos dispositivos de control varía, dependiendo de la destreza y del compromiso del usuario cuando utiliza el dispositivo portátil.
Multimodal	**2000's**	Similar a las interfaces de multimedios, éstas, siguen el principio "más es más" para proveer experiencias del usuario más enriquecidas y complejas. Ellos lo hacen así para multiplicar las formas en que la información es experimentada y controlada en la interface a través del uso de diferentes modalidades, por ejemplo: el tacto, la vista, el sonido, el habla. Las técnicas de interfaces que han sido combinadas para este propósito incluyen el habla y los gestos, la mirada y los gestos y el apuntador y el habla.
Shareable	**2000's**	Son diseñadas para más de una persona que los usa. A diferencia de las PCs, las laptops y los dispositivos móviles (que están dirigidos a usuarios únicos), estas interfaces típicamente proveen múltiples entradas y algunas veces permiten simultáneas entradas para grupos específicos. Esto incluye grandes formas de despliegue, por ejemplo: SmartBoards, en donde las personas utilizan sus propios apuntadores o gestos, y tableros interactivos, en donde grupos pequeños pueden interactuar con información desplegada en la superficie usando su dedo como apuntador.

Tabla 2.1: Tipos de Interfaces (continuación).

TIPO DE INTERFAZ	DÉCADA	DESCRIPCIÓN
Tangible	**2000's**	Esta, es un tipo de interacción basada en sensores, donde los objetos físicos, por ejemplo, ladrillos, bolas y cubos son acoplados con representaciones digitales. Cuando una persona manipula los objetos físicos, esto es detectado por un sistema de cómputo a través de un mecanismo sensitivo incrustado en el objeto físico, causando un efecto digital que ocurre, tal como un sonido, animación o vibración. Los efectos digitales pueden tomar lugar en un número de medios y lugares, o ellos pueden ser incrustados en el objeto físico por sí mismo.
Augmented and Mixed Reality	**2000's**	Son aquellas en donde las representaciones virtuales son impuestas en dispositivos y objetos físicos y en la realidad mezclada, en donde las vistas del mundo real están combinadas con vistas de un ambiente virtual. Las herramientas de oficina, como los libros, documentos y papers, fueron integradas con representaciones virtuales, usando proyectores y video cámaras. Ambos documentos virtuales y reales fueron combinados. La realidad aumentada ha sido experimentada mayormente en el área de la medicina, en donde los objetos virtuales, por ejemplo los rayos X y scanners son sobrepuestos en la parte del cuerpo de un paciente para ayudar al médico a comprender qué es lo que va a examinar u operar.

Tabla 2.1: Tipos de Interfaces (continuación).

TIPO DE INTERFAZ	DÉCADA	DESCRIPCIÓN
Wearable	**2000's**	Uno de los primeros desarrollos en cómputo "ponible" fue la cabeza –y gafas- con cámaras montadas que permitían al usuario grabar lo que veía y el acceso a la información digital mientras estaba en movimiento. Nuevas tecnologías de despliegue y comunicación de redes inalámbricas presentan muchas oportunidades para pensar acerca de cómo incrustar tales tecnologías en las ropas que usan las personas. Las aplicaciones que han sido desarrolladas incluyen diarios automáticos para los usuarios que guardan la fecha de lo que está sucediendo y lo que ellos necesitan hacer a través del día, y guías de turistas que informan a los usuarios de información relevante cuando ellos caminan a través de cualquier lugar público.
Robotic	**2000's**	Los Robots han estado notablemente como caracteres en películas de ciencia ficción, pero también jugando un papel importante como parte de las líneas de ensamblaje de manufactura, como investigadores remotos de locaciones peligrosas, como incendios. Las interfaces de consola han sido desarrolladas para habilitar humanos para robots de navegación y control en terrenos remotos, usando una combinación de joysticks y controles de teclado junto con cámaras e interacciones de sensores. El enfoque ha estado en diseñar interfaces que habilitan a los usuarios para conducir efectivamente y mover un robot remoto con ayuda del video real y mapas dinámicos.

Tabla 2.1: Tipos de Interfaces (continuación).

Para este proyecto se implementó una combinación de interfaces: por un lado fue creada una interfaz móvil dado que la aplicación fue desarrollada exclusivamente para dispositivos PDAs, específicamente para una Pocket PC; por otro lado, dado que es un sistema de navegación hablada, fue necesario también el diseño de una interfaz de Voz, cuyo tipo de interacción está basada en comandos, de tal forma que el usuario pueda navegar utilizando comandos a través del habla. Técnicamente este tipo de sistema es muy flexible, ya que permite al usuario tomar iniciativa y especificar mayor información en una

sentencia determinada, además de que el diseño de la interfaz móvil es sumamente sencillo puesto que lo más importante aquí es el menor uso de botones y controles que son remplazados por los comandos que el usuario emite.

Muchas tecnologías de interfaces avanzadas que han sido investigadas por décadas, recientemente han madurado al punto de sostener la promesa de futuras mejoras a grupos de interfaces. Un excelente ejemplo de ello, que podrá cambiar la manera en que el usuario utiliza los componentes de software de una aplicación, consiste en la interpretación del lenguaje hablado: Sistemas que permiten el uso del habla continua, en un diálogo naturalmente abierto entre el usuario y los componentes que permitirán el control y uso de sistemas afines sin la necesidad de un teclado o un mouse (Bullinger, 1999).

Los sistemas de reconocimiento de voz (como comúnmente se les conoce), están ahora disponibles con 30,000 a 40,000 vocabularios de palabras. Tales sistemas aún no están diseñados para usuarios independientes, lo cual quiere decir que cada usuario debe "entrenar" el sistema para el reconocimiento de su propia voz. Como tal, el reconocimiento de voz podría no ser tan apropiado para los usuarios, sin embargo, en sistemas distribuidos, por ejemplo, con "usuarios a largo plazo", el reconocimiento de voz simplifica la interfaz para dichos usuarios y posiblemente también acelere la entrada para tareas de texto intensivo como la "lluvia de ideas" (Carey, 1997).

El lenguaje hablado es una tecnología de interfaz poderosa; sin embargo la tecnología y aún la velocidad de reconocimiento es secundaria para el usuario. Lo más importante radica básicamente en el diseño del diálogo del usuario y la sencillez de este diálogo: la posibilidad para "hablar a través de"; por ejemplo, la habilidad para reconocer lenguajes continuos o para descubrir palabras específicas en sentencias de usuario y máxima flexibilidad del diálogo entre el usuario y el sistema (Bullinger, 1999).

En las interfaces de voz intervienen diversas tecnologías, las más frecuentes son (Casanovas, 2005):

- **Detección de tonos (DTMF):** El usuario oye una voz que le da las instrucciones y pulsa el teclado del terminal para escoger las opciones. El sistema reconoce la opción dada por el usuario a partir del tono pulsado.
- **Reconocimiento de voz (ASR):** El usuario hace peticiones o responde con la voz para dar instrucciones o elegir opciones. El sistema reconoce lo que dice el usuario.
- **Síntesis de voz (TTS):** La voz que oye el usuario no está pregrabada, es voz sintetizada, útil para dar respuestas con valores variables.
- **Verificación de la persona que habla (SV):** Es la vertiente biométrica del reconocimiento de voz que permite reconocer a la persona a través de las características de su voz.

En la Tabla 2.2. se presenta un análisis comparativo de las características principales que poseen las Interfaces de Usuario involucradas en el sistema de navegación de este

proyecto (Esto abarca la interfaz gráfica de PDLib y la nueva implementación propuesta de navegación hablada).

CARACTERISTICAS	INTERFAZ GRAFICA	INTERFAZ DE VOZ
De Lenguaje	Escrito. Utiliza técnicas de escritura para plasmar el lenguaje hablado.	Hablado. Utiliza la capacidad de articular sonidos como la voz.
De Comunicación	Se vale de símbolos e imágenes para su navegación.	Facilidad de comunicación a través del habla (de fines prácticos).
De Portabilidad	Menor.	Mayor.
De Vista	Existe información visual de las acciones y modos del usuario.	Poca o nula Información visual, todo se maneja a través de comandos de voz.
De Usabilidad	Fáciles de usar. Intuitivamente obvias para el usuario en relación al funcionamiento de sus componentes.	Prácticas. El usuario no necesita conocer mucho de la aplicación para poder navegar con comandos de voz.
Del Tipo de Usuario	Todo tipo de usuario vidente.	Mayormente utilizado por los invidentes.
De Espacio	El usuario dispone de un espacio visual.	El usuario no dispone de un espacio visual que escanear.
De Permanencia de información	La información permanece presente.	La información se presenta y desaparece.

Tabla 2.2: Análisis comparativo de la Interfaz Gráfica vs. La Interfaz de Voz.

CARACTERISTICAS	INTERFAZ GRAFICA	INTERFAZ DE VOZ
De Margen de Error	Si es una interfaz bien diseñada, no se presentan errores en la navegación.	Errores continuos en el reconocimiento de la voz, independientemente del diseño de interfaz.
De Adaptabilidad del Usuario	Por lo regular, se adapta fácilmente, si se trata de una interfaz "amigable".	De principio, los usuarios no están familiarizados con relación a la utilización del sistema.
De Navegación	Generalmente de fácil navegación, si sus componentes son atractivos para el usuario.	Muchas veces, los usuarios se pierden, no encuentran lo que necesitan.

Tabla 2.2: Análisis comparativo de la Interfaz Gráfica vs. La Interfaz de Voz (Continuación).

2.3 Cómputo Móvil

El término "Cómputo Móvil" implica un concepto inmediato: hacer cómputo "sin alambres", sin cables, sin que el equipo personal esté visiblemente conectado a algo. Involucra dos avances tecnológicos fundamentales: las redes inalámbricas para transmisión de datos (conocidas genéricamente como wireless) y la miniaturización de los componentes de un equipo de cómputo, al grado de portarlos como un accesorio más del vestir. De manera secundaria, pero no por ello menos importante, tiene que ver con el continuo desarrollo de aplicaciones y sistemas operativos (software) más sofisticados y especializados.

En la actualidad los grandes avances en microelectrónica y las redes de comunicaciones, han hecho posible que los dispositivos móviles provean al usuario mejores prestaciones y servicios.

Los PDAs, cuyas siglas en inglés significan Asistente Personal Digital, son dispositivos portátiles que fueron originalmente diseñados como organizadores personales, pero que al paso de los años se hicieron más versátiles (Greg, 1999). Los principales usos y tareas de un PDA básico incluyen muchas características como: calculadora, reloj y calendario, juegos, acceso al Internet, envío y recepción de e-mails, radio, video, libreta de direcciones, hoja de cálculo, entre otros. Los modelos más nuevos tienen distintas capacidades de audio y pantalla a color, permitiéndoles ser usados como teléfonos móviles, web browsers o media players. Muchos PDAs pueden acceder al Internet, intranets o extranets vía Wi-Fi o Wireless Wide-Area Networks (WWANs).

Greg (1999), señala que como el uso y la funcionalidad de los dispositivos móviles se incrementan, las organizaciones de TI encaran serios retos en estrategias de desarrollo para integrarlos y administrarlos, y al mismo tiempo, maximizar su efectividad. Por ejemplo, las organizaciones de TI atienden varios tipos y líneas de productos de dispositivos móviles, los cuales frecuentemente corren en diferentes plataformas y usan diferentes aplicaciones, tecnologías de acceso a redes, y así sucesivamente.

Estos dispositivos se pueden ver como la evolución de las tradicionales agendas electrónicas, con funciones limitadas y sin capacidad de programación, hacia ordenadores personales de prestaciones limitadas pero con posibilidades de programación, lo que los convierte en herramientas de propósito general para poder ser aplicadas a cualquier necesidad de una organización.

Actualmente se tienen diferentes tipos de PDAs con características funcionales muy específicas, como se muestran en la Tabla 2.3.

TIPO	PANTALLA	TECLADO	SISTEMA OPERATIVO
Palm	Táctil	No	Palm OS
Pocket PC	Táctil	No	Windows CE
Psion	Táctil	Sí	EPOC

Tabla 2.3: Tipos de PDA existentes en el mercado.

A pesar de que cada tipo de dispositivo tiene un Sistema Operativo diferente, y que estos a su vez son incompatibles entre sí, existen características comunes que los identifican, entre las más destacadas, se encuentran las siguientes:

- Soportan gran variedad de nuevas tecnologías, tales como dispositivos USB, DVD, infrarrojos, Ethernet y GPS.
- Incluyen programas de navegación web con ciertas restricciones a nivel de cliente (no todos soportan ejecución de código en cliente, como pudiera ser JavaScript).
- Proporcionan clientes de correo electrónico. Estos dispositivos tienen módulos de expansión para ampliar sus capacidades. A través de estos módulos, se puede conectar al PDA ampliaciones de memoria, tarjetas módem, red, etc.

Para fines de este proyecto, se utilizó específicamente un PDA de tipo Pocket PC, el cual tiene la funcionalidad de Windows, opciones de expansión, y comparte herramientas de desarrollo con otras plataformas del mismo Windows, corriendo versiones de Pocket de MS Outlook, Internet Explorer, Word, Excel, Windows Media Player, por mencionar algunas.

2.4 Sistemas de Reconocimiento de Voz

Dos soluciones software: Talks y Mobile Speak, y una solución hardware, Owasys, destacan sobremanera en el área de tecnología móvil con reconocimiento de voz.

Talks ("TALKS for Series and Nokia Communicator", 2008) añade a un móvil Symbian de la Serie 60 o 90 la capacidad del habla. El principal cometido de esta aplicación es que una persona invidente sepa en todo momento sobre qué elemento de la interfaz del móvil se encuentra y cuáles son las acciones disponibles en el contexto actual. Talks utiliza el sintetizador de voz ETI Eloquence (*SpeechWorks Releases*, 2003).

Mobile Speak para Pocket PCs (*Mobile Magnifier para Pocket PCs V2.0*, 2008) es un lector de pantallas que se instala en un PDA con Windows Mobile Pocket PC, o en un teléfono-PDA Pocket PC, que te permite usar el dispositivo incluso si no puedes leer la pantalla. La información visual que se muestra en la misma, se presenta mediante salida de voz sintetizada generada mediante el uso de tecnologías de conversión de texto a voz (TTS) y se envía al altavoz del dispositivo o a auriculares. Los contenidos de la pantalla se pueden mostrar también en Braille si el dispositivo móvil se conecta a una línea Braille. Las salidas de voz y de Braille pueden utilizarse simultánea o independientemente.

La empresa OWASYS (*OWASYS 22C y 112C,* 2003) ha creado móviles adaptados para gente invidente. El modelo 22C es un móvil sin pantalla que viene acompañado de un sintetizador de voz proporcionado por Babel (*The Owasys 22C,* 2007) y provee las funcionalidades típicas de cualquier móvil convencional, como mensajería de texto, lista de contactos y telefonía.

Por otro lado, IBM dio a conocer un sustancial avance en tecnología de voz que permite a los conductores de automóviles y a los usuarios de dispositivos de mano enunciar comandos naturalmente sin tener que memorizar comandos específicos predeterminados. Lanzado como parte del paquete de software Embedded ViaVoice 4.4 de IBM, constituye un avance tecnológico significativo para la tecnología de habla incorporada en dispositivos y en sistemas de navegación de automóviles. ("Avances en software de reconocimiento de voz", 2006).

Anteriormente, los usuarios debían aprender, memorizar y usar una serie fija de frases y comandos para interactuar con los sistemas de reconocimiento de habla. Por ejemplo, al pedir "Radio 104.3 FM," la nueva tecnología pionera de IBM permite a los conductores decir simplemente "sintonice 104.3" o "ponga la estación de radio en 104.3" o "cambie la estación de radio a 104.3". Una amplia variedad de comandos intuitivos pueden cambiar la estación de radio a la sintonía deseada, eliminando la necesidad de memorizar una lista de comandos específicos. Embedded ViaVoice 4.4 de IBM trae "reconocimiento de comandos en forma libre" que utiliza el modelado de lenguaje estadístico avanzado y la interpretación semántica para habilitar la comprensión del lenguaje natural entre el usuario y el sistema de reconocimiento de voz. El reconocimiento de comandos en forma libre permite a los

usuarios utilizar frases intuitivas que no son “memorizadas” para los dispositivos de control, tales como sistemas de radio o de navegación en automóviles, o comandos en dispositivos de mano.

En resumen, el mercado de las aplicaciones de reconocimiento de voz continúa creciendo en todo el mundo. Sin embargo, y a pesar de los importantes recursos dedicados a las iniciativas de I+D en este campo, el uso masivo del reconocimiento de tecnologías de voz sigue sin eclosionar. Hoy por hoy muchos clientes de centros de llamadas prefieren hablar con agentes humanos para evitarse los reiterados mensajes de “No le he entendido bien”.

En el entorno de dispositivos móviles el índice de aciertos de los programas de reconocimiento de voz es muy alto una vez que se les ha “entrenado” en captar la voz del usuario. Estos programas resultan especialmente útiles en entornos industriales donde por ejemplo los usuarios tienen que tener las manos ocupadas, donde se necesiten mecanografiar textos repetitivos o para el uso de personas discapacitadas. Sin embargo, lo cierto es que la gran mayoría de usuarios siguen prefiriendo el teclado como dispositivo de entrada del ordenador.

Capítulo 3

Navegación Hablada de PDLib

En este capítulo se presenta una descripción de cada uno de los componentes que integran la Arquitectura de la aplicación de "Navegación Hablada de PDLib" en dispositivos móviles, así mismo se describe la forma en que estos interactúan, y se especificarán las diversas funciones que realizan tales componentes.

También se identifican los retos de la interacción de Clientes móviles con Servidores Remotos, utilizando un mecanismo de comunicación para enviar y recibir las peticiones realizadas por el Cliente móvil en el formato que el Servidor de Reconocimiento de Voz atenderá para su "traducción"; así como la adaptación de los Clientes mismos para adaptarse a cualquier plataforma (Linux, Windows, Macintosh, etc.).

3.1 Arquitectura Existente de PDLib

La plataforma PDLib (ver Figura 3.1) propone un Sistema de Librería Digital con Acceso Universal ("PDLib: The Personal Digital Library Project Webpage",2006). Este es "personal" en el sentido de que cada usuario tiene acceso a la plataforma de un repositorio cuya documentación es de propósito general. Se dice que es de "Acceso Universal" ya que permite a los usuarios acceder a su biblioteca digital personal desde la mayoría de los dispositivos de cómputo conectados al Internet, incluyendo dispositivos móviles.

En esta sección se describirá brevemente la arquitectura de PDLib presentando dos formas tradicionales de su diseño: la Figura 3.2 –Visión General de PDLib- y la Figura 3.3– su arquitectura detallada-.

La información presentada a continuación está basada específicamente en los trabajos de García (2004); Alvarez et al. (2005).

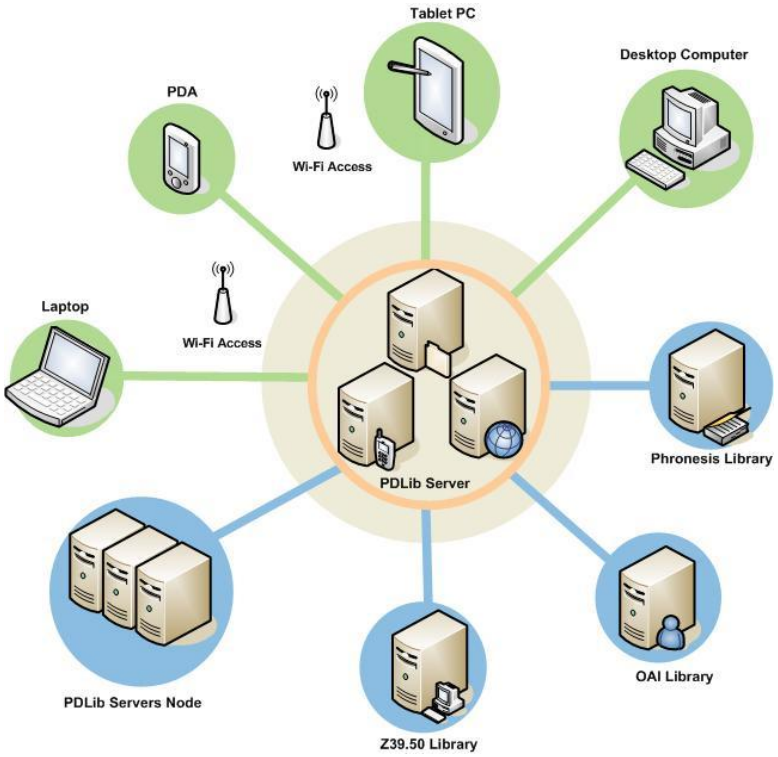

Figura 3.1: Concepto PDLib.

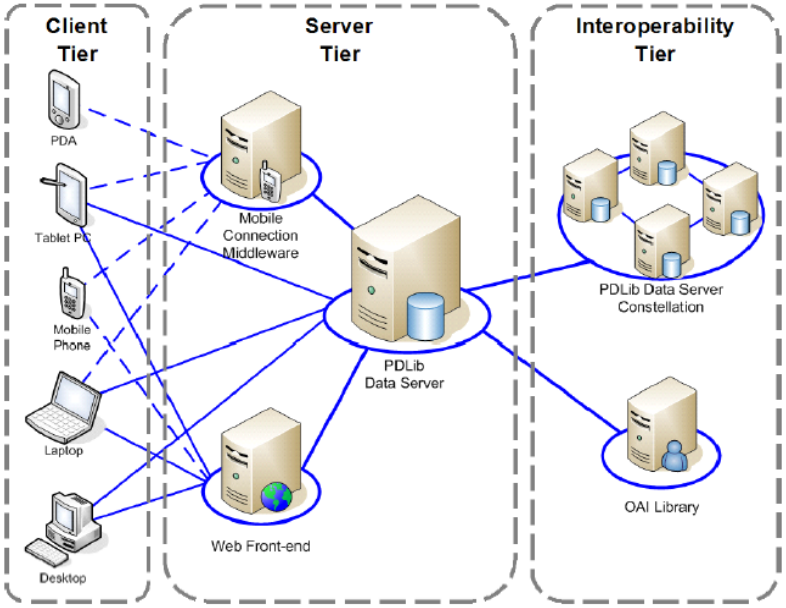

Figura 3.2: Visión General de PDLib.

La arquitectura de PDLib está diseñada para proveer servicios de librerías a muchos tipos de dispositivos (ej.: desktop, laptop, PDA y teléfonos celulares), con múltiples Sistemas Operativos (ej.: Windows, Linux, Mac OS, Palm OS y Windows CE). El concepto de operación detrás de esta arquitectura es también proveer a cada dispositivo de la mejor forma posible para interactuar con los servicios ofrecidos por el Data Server.

PDLib, está compuesto de tres capas:

- ***Client Tier.*** En esta capa se incluyen la variedad de dispositivos con los que un usuario de PDLib puede interactuar.
- ***Server Tier.*** Infraestructura servidor que provee los servicios a los clientes.
- ***Interoperability tier.*** Incluye los servicios de interoperabilidad con otros servidores PDLib y sistemas de bibliotecas digitales que cumplen con el estándar OAI-PMH ("OAI Initiative.Open archives initiative").

Los dispositivos de la capa cliente se comunican con la capa del servidor (server tier) para obtener acceso a los diferentes servicios de la biblioteca digital PDLib. El tipo de acceso de los clientes con el servidor varía de acuerdo a las características del dispositivo. Los accesos pueden ser cualquiera de los siguientes tipos (ver Figura 3.3):

- ***Acceso por Middleware.*** Da soporte a dispositivos móviles, especialmente a aquellos con recursos de cómputo limitados (ej.: PDA, teléfonos móviles con soporte HTTP).
- ***Acceso por Web.*** Provee acceso mediante HTTP a cualquier dispositivo que incluya un navegador Web (p.e: WML/HTML micronavegadores).
- ***Acceso directo.*** Proporciona acceso directamente con el data server a aplicaciones con requerimientos muy particulares.

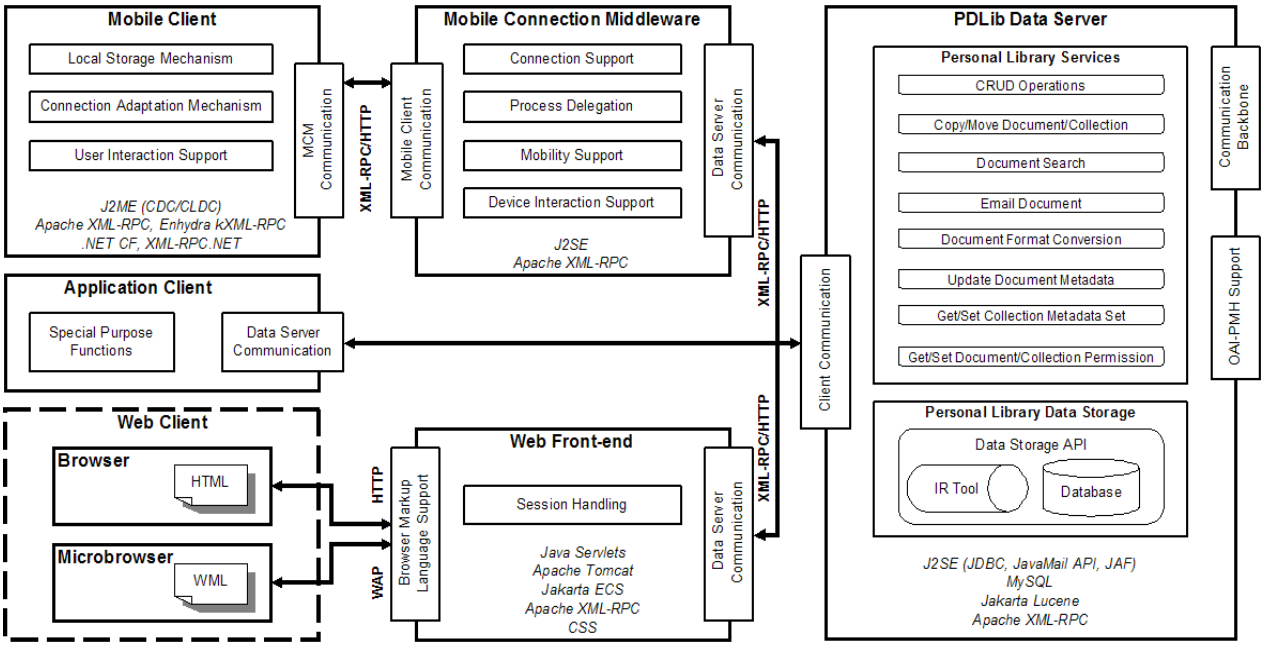

Figura 3.3: Arquitectura detallada de PDLib.

Los componentes de PDLib, requeridos para su interacción e implementaciones de prototipo, son descritos en las siguientes secciones.

3.1.1 Cliente

Las aplicaciones cliente de un ambiente móvil pueden ser clasificadas de acuerdo a su arquitectura del lado cliente en: clientes ligeros (thin clients) y clientes pesados (thick clients); y de acuerdo a su movilidad en: clientes fijos (fixed clients) y clientes móviles (mobile clients). Esta clasificación de clientes se puede apreciar mejor en la Figura 3.4 propuesta por Alvarez et al. (2005).

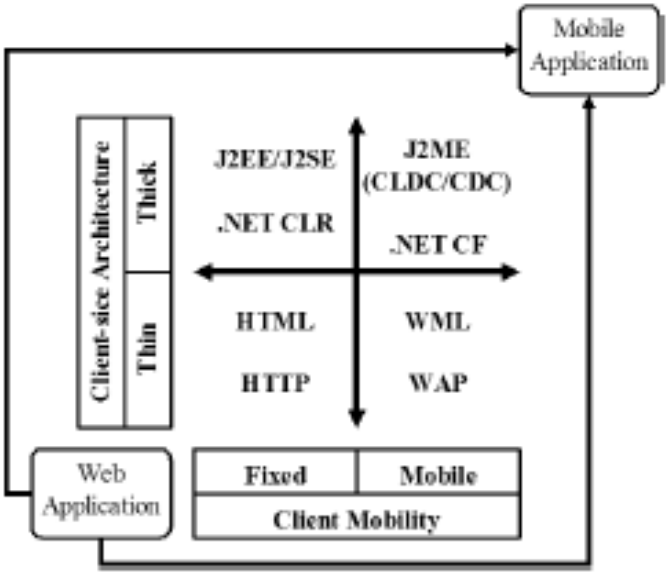

Figura 3.4: Clasificación de clientes en PDLib.

Con el objetivo de proveer un medio de acceso apropiado para los dispositivos mostrados en la Figura 3.2, los siguientes tipos de aplicaciones cliente fueron definidos (ver Figura 3.3):

- ***Clientes Web (clientes ligeros fijos y móviles).*** En esta categoría se encuentran los dispositivos capaces de mostrar el contenido de una página en formato HTML o WML. A pesar de que estos clientes no cuenten con todas las características de un cliente pesado, proveen soporte de interacción básica con PDLib. Estos clientes Web se comunican con el Data Server a través del Web Front-end.
- ***Aplicación cliente (clientes pesados fijos).*** Este tipo de aplicación está diseñada para ejecutarse en computadoras de escritorio o portátiles que tiene pocas limitaciones de recursos de cómputo. Estas aplicaciones tienen la capacidad de comunicarse directamente con el Data Server y presentar interfaces gráficas más ricas que las de un cliente Web.
- ***Clientes móviles (clientes pesados móviles).*** Estos clientes fueron diseñados para lidiar con las limitaciones inherentes al ambiente móvil. Esto implica una redefinición de las funcionalidades que proporcionan los clientes pesados fijos, con la finalidad de proveer una abstracción de una biblioteca digital en el dispositivo móvil. Estos clientes requieren de un middleware (MCM) para la comunicación con el Data Server.

Uno de los principales retos de PDLib está en los clientes móviles, puesto que ellos tratan con las limitaciones del ambiente móvil. Para complementar sus tareas, los clientes móviles desarrollan las siguientes funciones:

- ***Mecanismos de Almacenamiento Local.*** Almacenar documentos en el dispositivo móvil en modo "offline" (fuera de línea o desconectado de la red).
- ***Mecanismos de Adaptación de Conexión.*** Estos mecanismos proveen un tiempo de respuesta constante a pesar de la variabilidad de conexión para conexiones inalámbricas, por interactuar con el MCM (Ver sección 3.1.3).
- ***Soporte de Interacción de Usuarios.*** Diseño de Interfaces Gráficas que permiten al usuario navegar a través de la biblioteca digital desde su dispositivo móvil.
- ***Configuración de Imagen.*** Las imágenes juegan un rol de aplicación importante, puesto que las más representativas resaltan la experiencia del usuario. Sin embargo, las imágenes adicionan un overhead al tamaño de la aplicación, tanto en almacenamiento permanente como en memoria; por lo que para reducir las demandas de memoria, las imágenes son recuperadas desde el almacenamiento permanente por medio de una pantalla básica, como requeridas por interacción del usuario. Una vez traídas, las imágenes permanecerán en memoria, para su uso posterior.
- ***Servicios de Librería.*** Los servicios de librería están disponibles a través de los objetos de la interfaz de usuario (ejemplos: combo boxes, listas, menús y botones).
- ***Navegación de la Biblioteca***. El cliente móvil facilita la interacción entre los usuarios y la biblioteca digital con una interfaz de navegación parecida a la interfaz provista por los sistemas de archivos para acceder a las colecciones y documentos de la biblioteca.

3.1.2 Web Front-End

El Web Front-end transforma los servicios de biblioteca digital personal dentro de una aplicación Web. Para el soporte de clientes móviles y fijos, el Web Front-end genera código WML o HTML de acuerdo al dispositivo que lo solicita, es decir, si la petición la hace un microbrowser desde un dispositivo ligero (thin client), el Web Front-end responderá a la petición en formato WML, mientras que si la petición la hace un navegador Web desde una computadora de escritorio, la respuesta que enviará el Web Front-end será en formato HTML.

Para mantener la interacción de un cliente Web con el Web Front-end a través de una petición HTTP, se usa el mecanismo de manejo de sesiones.

3.1.3 Mobile Connection Middleware

Uno de los principales problemas a resolver para proveer los servicios del data Server a clientes móviles es el hecho de que el data Server fue diseñado para soportar dispositivos fijos (ej. computadoras de escritorio). Existe una notable diferencia en recursos de cómputo entre dispositivos móviles (ej. PDAs) y dispositivos fijos. Esta disparidad de los recursos dificulta la adaptación del data Server con las capacidades de los dispositivos móviles y señala la necesidad de un componente middleware como intermediario para la interacción de ambos. Específicamente para PDLib, se ha diseñado el Mobile Connection Middleware (MCM), el cual es responsable de proveer las siguientes funcionalidades:

- ***Soporte de Conexión.*** Realiza tareas que proporcionen la adaptabilidad que el cliente móvil necesita para contrarrestar la variabilidad y limitaciones de ancho de banda que caracterizan a las conexiones inalámbricas.
- ***Delegación de Procesos.*** Ejecuta funciones que demandan una excesiva cantidad de recursos computacionales para el dispositivo móvil.
- ***Soporte de movilidad.*** Desarrolla operaciones para acelerar la recuperación de información desde el data Server, para que ésta sea almacenada en servidores caché, los cuales están cercanos al usuario. Cuando el usuario cambia de localidad, éste deberá necesariamente soportar la migración de información entre diferentes instancias del MCM.
- ***Soporte a la interacción de dispositivos.*** Aplica técnicas de adaptación de contenido de acuerdo a las características del dispositivo sobre el cual éste es diseñado para el despliegue de información.

3.1.4 Data Server

El data Server provee los servicios de una biblioteca digital personal, almacenamiento de datos y soporte de interoperabilidad a través del OAI-PMH ("OAI Initiative. Open archives initiative"). El data Server provee las siguientes funcionalidades:

- ***Servicios de Biblioteca Personal.*** El data Server proporciona la creación, recuperación, actualización y eliminación de operaciones sobre los objetos de librerías guardados en el almacenamiento de datos personal (colecciones, documentos y metadatos). El data Server también provee servicios para copiar y mover documentos o colecciones, buscar documentos en la biblioteca personal, enviar documentos a la biblioteca personal de otros usuarios, enviar documentos vía correo electrónico a cualquier usuario que posea una cuenta de e-mail, conversión de formato de documentos, cambiar la colección del conjunto de metadatos, editar metadatos, y conceder o revocar permisos para acceso al contenido de la biblioteca.

- ***Almacenamiento de Datos Personales.*** Este servicio provee almacenamiento estructurado con recuperación de información escalable, apta para bibliotecas digitales, mediante el uso de un motor de búsqueda de texto. Se utiliza una base de datos para almacenar los objetos de las bibliotecas digitales personales, para representarlos y relacionarlos entre ellos con el propósito de realizar búsquedas de texto completo en los documentos y en sus metadatos.
- ***Backbone de Comunicación.*** Establece un backbone de comunicación con otros servidores de datos.
- ***Soporte OAI-PMH.*** El data Server muestra los metadatos de los documentos de la biblioteca digital a través del protocolo OAI-PMH ("OAI Initiative.Open archives initiative"). En adición, los usuarios de otras bibliotecas digitales que cumplan con este protocolo pueden acceder a los documentos PDLib (con permisos de acceso público) y viceversa.

Actualmente el sistema PDLib sigue siendo objeto de investigación y un esfuerzo de desarrollo en donde interesantes retos relacionados con bibliotecas digitales y cómputo móvil son explorados.

3.2 Arquitectura Propuesta para el Modelo de Navegación Hablada

Para este proyecto se diseñó un modelo de navegación de tipo Cliente-Servidor, en donde se establece el acceso a un reconocedor de voz existente en un Servidor remoto: este último será invocado desde un cliente móvil, de tal forma que los usuarios naveguen a través de él, utilizando "palabras habladas". El caso de uso que se muestra en la Figura 3.5 describe la interacción de los actores involucrados en el servicio de voz para esta arquitectura propuesta.

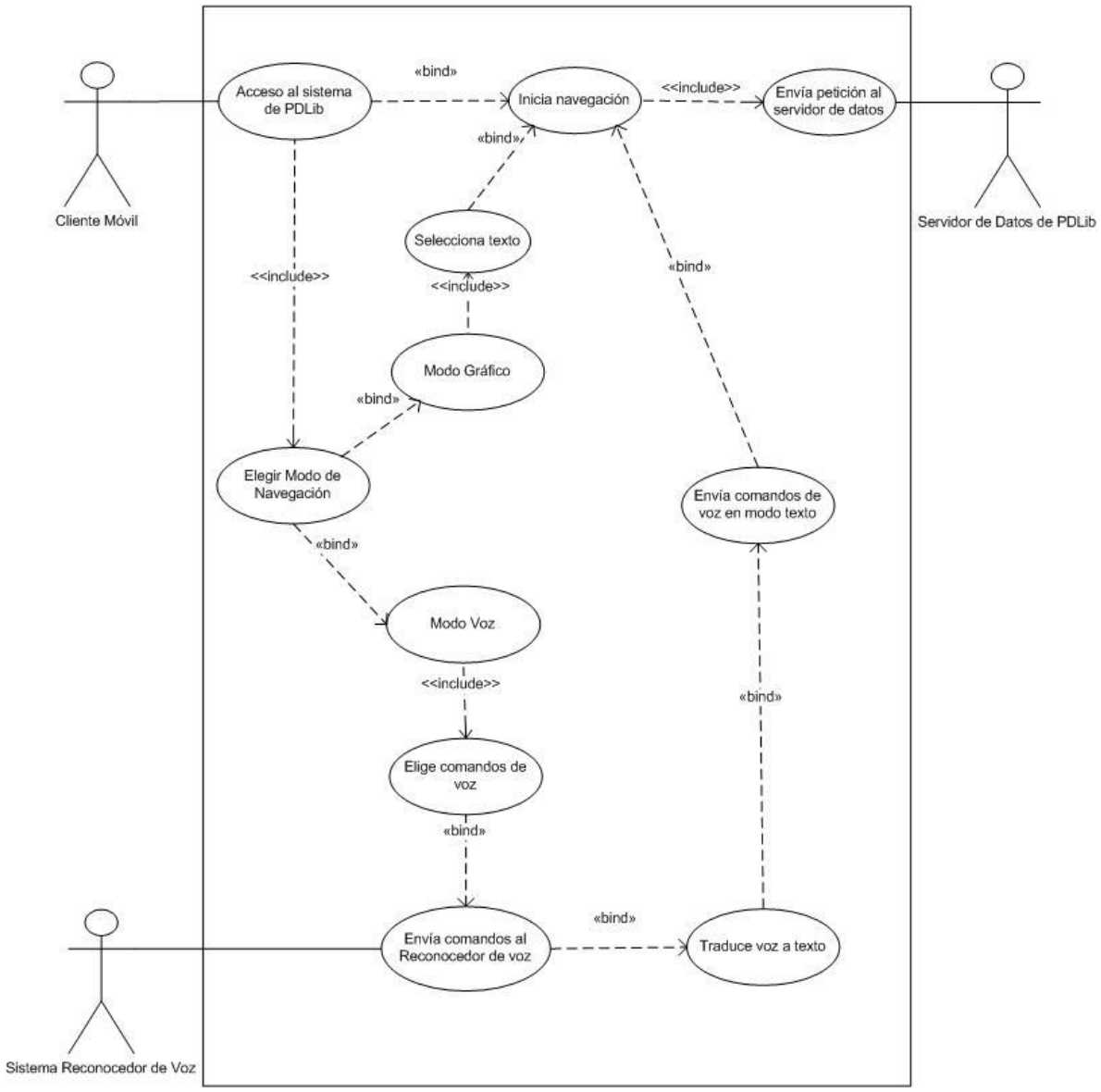

Figura 3.5: Caso de Uso del Modelo de Navegación Hablada.

Específicamente, este proyecto propuesto se centró, primeramente en diseñar un modelo de navegación hablada, en donde el usuario pueda ejecutar comandos de voz desde su dispositivo móvil, a través de una interfaz de voz, la cual interactúa directamente con la interfaz gráfica diseñada por PDLib.

En la Figura 3.6 se presenta una vista general en la que el usuario interactúa con el reconocedor de voz, sugerido en este proyecto.

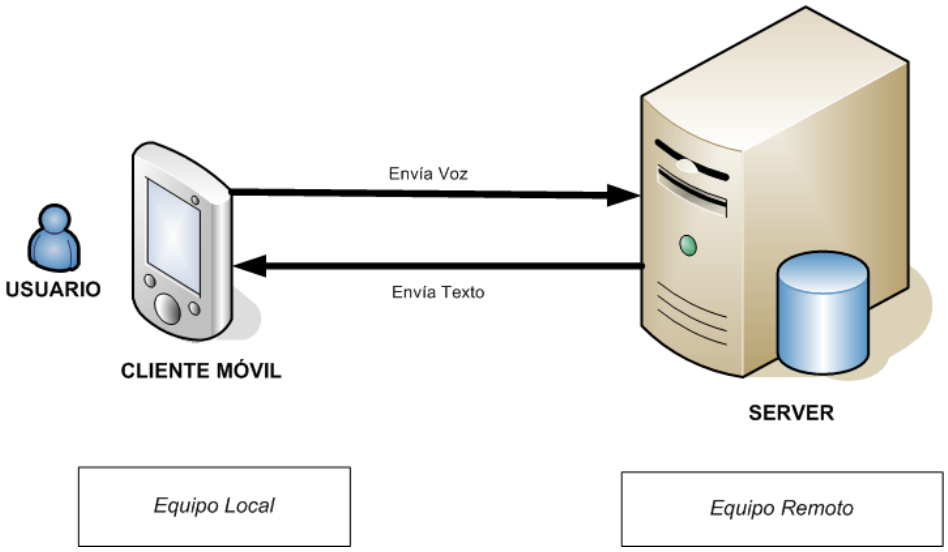

Figura 3.6: Vista general de interacción entre el Usuario y el Reconocedor de Voz.

Cuando el usuario navega a través de PDLib en *modo voz*, se hace la invocación al servidor en donde se encuentra hospedado el reconocedor de voz; esto es, el usuario "graba", a través del cliente, un comando de voz que es enviado automáticamente al servidor, quien lo transformará en *modo texto*, para devolverlo nuevamente al cliente móvil, quien finalmente realizará, de forma inmediata, la ejecución del mismo para iniciar la navegación.

El *modo voz* y el *modo texto* son dos formas representativas que el usuario tendrá como opciones para poder navegar a través de las interfaces de PDLib; sin embargo es muy importante recalcar que, todos los comandos que el usuario utilice para navegar, serán enviados al servidor de datos de PDLib en *modo texto*, siendo esto totalmente transparente para el usuario.

La Arquitectura Propuesta es la arquitectura típica de un sistema Cliente-Servidor (ver Figura 3.7), en donde la mayor parte de la aplicación corre en el lado del Servidor (servidor pesado), el cual recibe e interpreta los requerimientos del Cliente, respondiendo a sus peticiones, siguiendo políticas propias de prioridad de asignación.

El cliente no es más que la interfaz de voz (cliente ligero) quien envía peticiones al servidor donde se encuentra hospedado el Reconocedor de Voz, cuya función es recibir los comandos de voz que envía el Cliente, para que sean convertidos en modo texto, y en ese mismo formato devolverlos nuevamente al Cliente para que puedan ser enviados y reconocidos por el Servidor de Datos de PDLib, quien tiene conocimiento de las bibliotecas digitales que se encuentran disponibles y para quien será transparente el servicio de Reconocimiento de Voz.

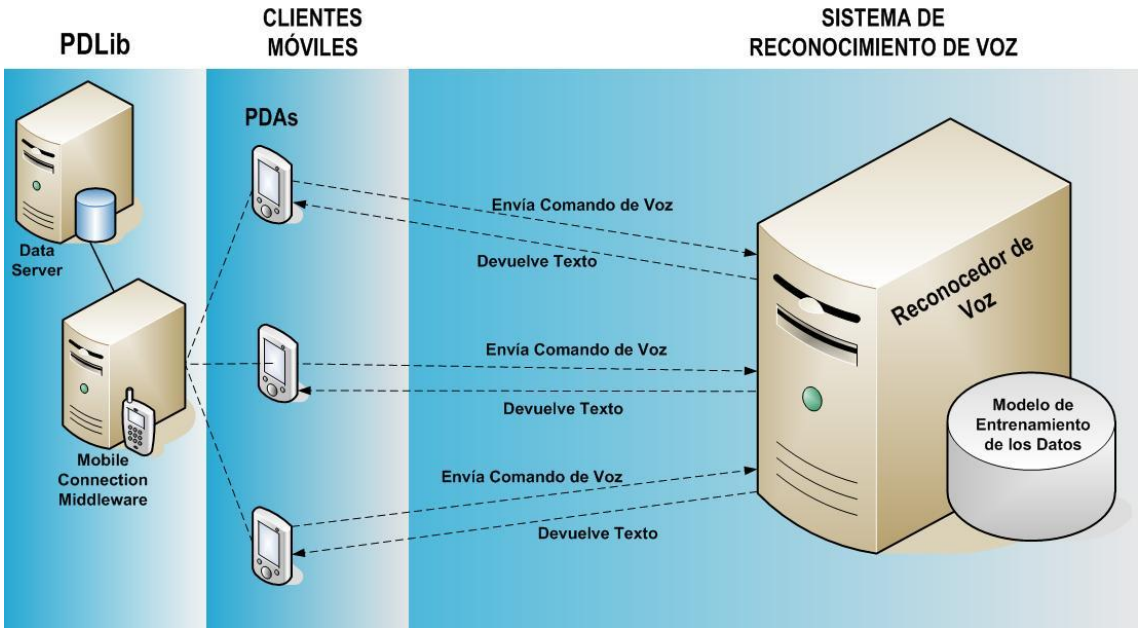

Figura 3.7: Arquitectura del Modelo de Navegación Hablada.

Básicamente la aplicación está implementada para un Cliente ligero que contiene las interfaces de navegación, y un Servidor pesado, que traduce comandos de voz a texto.

Para explicar detalladamente la arquitectura de la aplicación en términos de cada uno de los elementos involucrados, así como la funcionalidad de cada uno de ellos, ésta se describe en tres capas:

- Infraestructura de PDLib
- Cliente Móvil (PDA)
- Sistema de Reconocimiento de Voz

La capa "Infraestructura de PDLib" es la que se describe en la sección 3.1 de este capítulo. Esta capa muestra la infraestructura que soporta el funcionamiento para los Clientes y su interacción con PDLib. La capa "Cliente Móvil" muestra los dispositivos del tipo PDA (específicamente Pocket PC) con los que un usuario puede acceder a PDLib, mediante navegación hablada. Por último la capa "Sistema de Reconocimiento de Voz" es un Servidor pesado que hospeda al Reconocedor de Voz, cuya función es esencial para poder interpretar los comandos de voz que el usuario utilizará para la navegación a través del Cliente móvil. Estas dos últimas capas conforman la arquitectura detallada de navegación hablada (ver Figura 3.8), la cual fue diseñada con el objetivo de proveer un medio de acceso apropiado a la infraestructura de PDLib (Es importante destacar aquí que la infraestructura de PDLib fue desarrollada independiente del desarrollo de este nuevo modelo de navegación hablada).

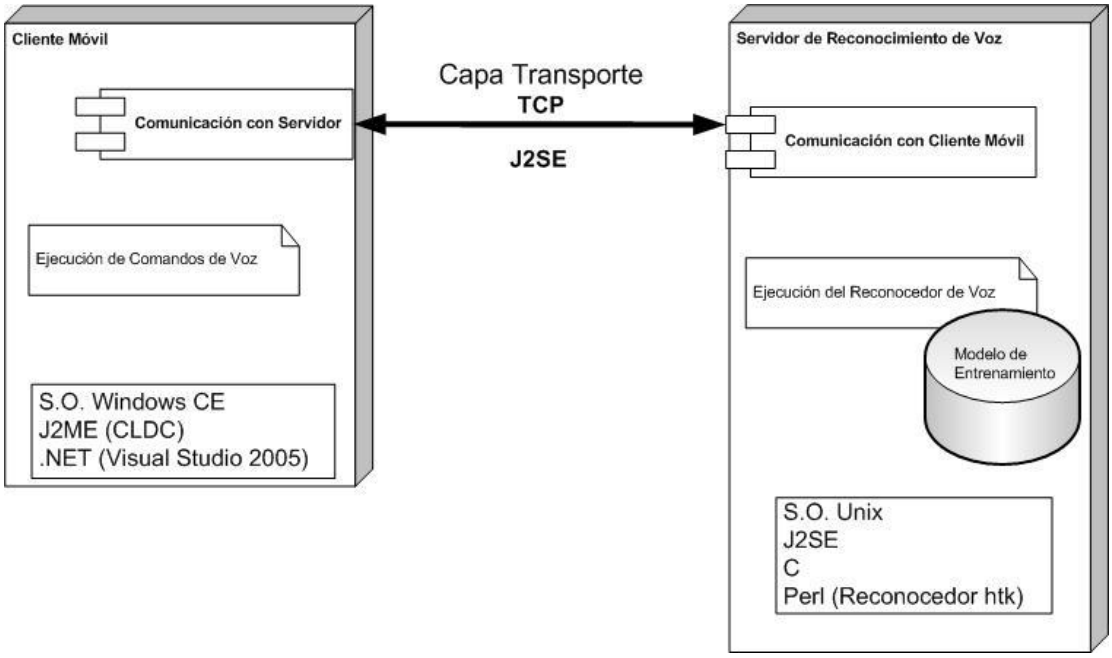

Figura 3.8: Arquitectura detallada del Modelo de Navegación Hablada.

En las siguientes subsecciones se da una descripción de los componentes involucrados en la arquitectura del modelo de navegación hablada y la forma en que estos interactúan.

3.2.1 Cliente

Para fines de este proyecto se eligió un tipo de cliente móvil ya que de este modo, un usuario en movimiento puede tener interacción con PDLib a través de una Interfaz diseñada específicamente para la navegación hablada. Para realizar su tarea, los clientes móviles poseen las siguientes funcionalidades:

- Comunicación con el sistema reconocedor de voz. Se encarga de enviar los comandos de voz al reconocedor de voz.
- Comunicación con el servidor de datos de PDLib. Se encarga de enviar los datos convertidos en "modo texto".
- Interfaz de Usuario. Interfaz de voz que permite a un usuario navegar a través de PDLib (utilizando comandos de voz), por medio del dispositivo móvil.

Los clientes móviles con los que se plantea tener acceso a la navegación hablada se restringen exclusivamente a dispositivos de tipo PDAs con conectividad Wi-Fi. La razón fundamental ha sido la mayor capacidad de procesamiento de estos equipos, comparados con otros dispositivos móviles; a su vez, de entre los distintos modelos de PDA se han

seleccionado aquellos que soportan el sistema operativo Pocket PC, por la mayor facilidad de desarrollo del soporte multimedia -acceso al altavoz y micrófono- necesarios en el sistema. Desde un punto de vista económico, también existe una razón primordial para su elección, ya que un PDA de precio medio en el mercado, tiene una capacidad de proceso que se puede equiparar a la de una PC con un procesador Pentium II a 200MHz, lo que hace que resulten especialmente interesantes.

Uno de los problemas principales que se plantearon al momento de decidir la tecnología a utilizar para la navegación hablada, fue la necesidad de que fuera independiente de las plataformas, para que se pudieran enfocar los esfuerzos de desarrollo a un solo prototipo y que no se tuviera la necesidad de dar mantenimiento a un lenguaje por cada plataforma. Es por esto que se tomó la decisión de utilizar a Java como el lenguaje de programación, pues además de que actualmente se cuenta con máquinas virtuales y API's de desarrollo que funcionan en las diferentes plataformas (Linux, Windows, Macintosh, etc.), también proporciona la posibilidad de usarlo en versiones más limitadas de dispositivos móviles (*Sun Microsystems*, 2008).

3.2.2 Reconocedor de Voz

La parte esencial de esta arquitectura lo constituye el Servidor de Reconocimiento de Voz. La Figura 3.9 muestra la segmentación de aplicaciones del Servicio de Voz en donde el Servidor es el encargado de atender a las peticiones del Cliente para traducir mediante un modelo de entrenamiento de datos, los comandos de voz (utilizados en la navegación de PDLib), que éste le envía.

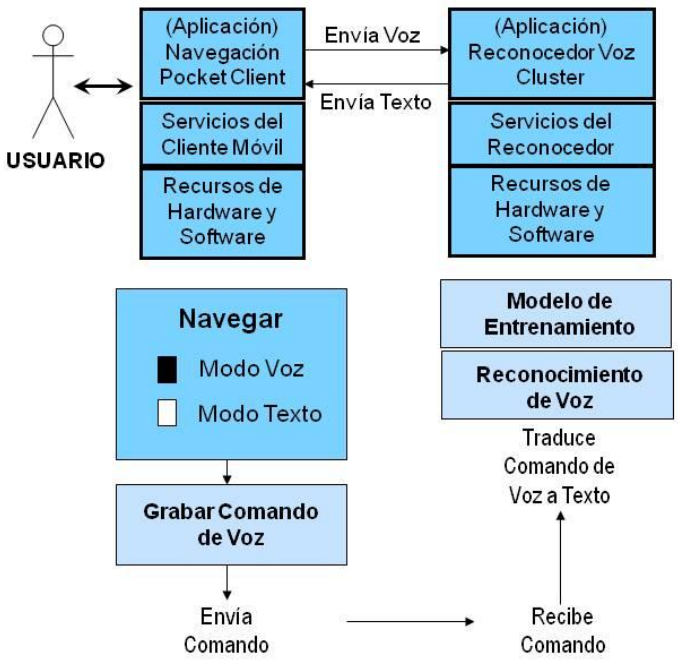

Figura 3.9: Segmentación de Aplicaciones: Servicio de voz.

Un Sistema de Reconocimiento de Voz es conocido formalmente como ASR (*Automatic Speech Recognition*), cuyas siglas en inglés significan Reconocimiento Automático del Habla. Esta tecnología permite a una computadora identificar las palabras que una persona interpreta en un micrófono o teléfono. Lo esperado con este tipo de sistemas es lograr un 100% de exactitud en tal interpretación, independientemente del tamaño del vocabulario, el ruido, el acento del hablante, o las condiciones del canal. Lograr esto en tiempo real también es importante. A pesar de que hay muchas décadas de investigación, una exactitud mayor de 90% sólo se logra cuando la tarea es restringida en alguna manera. Dependiendo de cómo se restrinja la tarea, se pueden alcanzar diferentes niveles de desempeño; por ejemplo, si el sistema se entrena para aprender la voz de un hablante determinado, entonces se pueden distinguir vocabularios grandes, aunque la exactitud será entre 90% y 95% para sistemas comerciales. Para vocabularios grandes y muchos hablantes sobre diferentes canales, la exactitud no es mayor del 87 %, y el procesamiento puede tomar mucho tiempo (Peña, 2004).

El Sistema de Reconocimiento de Voz implicado en esta implementación consiste en el reconocimiento del "lenguaje hablado" basado en una secuencia de segmentos de sonido discretos que son encadenados en el tiempo. Para estos segmentos, llamados fonemas, se asumen características acústicas y articulatorias únicas, que utilizan redes neuronales (para su reconocimiento) a través de la creación de clusters (esta definición se describe más específicamente en la Sección 4.2 del capítulo 4).

3.2.3 Protocolo de Control de Transmisión (TCP)

La interacción Cliente-Servidor es frecuentemente dada usando los protocolos de transporte de la red subyacente. Con la incrementada popularidad del Internet, es ahora más común construir aplicaciones y sistemas utilizando TCP. El beneficio de utilizar este protocolo se enfoca en la confiabilidad que ofrece al trabajar sobre cualquier red (Steen & Tanenbaum, 2002).

De acuerdo con Coulouris, Dollimore, & Kindberg (2001), TCP provee un sofisticado servicio de transporte. Este provee, arbitrariamente una entrega confiable de largas secuencias de bytes a través de la abstracción de la programación basada en *stream*. La confiabilidad garantizada implica la entrega al proceso receptor, de todos los datos presentados por el proceso emisor, en el mismo orden. TCP es conexión orientada. Antes de que cualquier dato sea transferido, los procesos de envío y recepción deben cooperar en el establecimiento de un canal de comunicación bi-direccional. La conexión es simplemente un acuerdo *end-to-end* para efectuar una transmisión de datos confiable. La capa TCP incluye mecanismos adicionales (implementados sobre IP) para encontrar la confiabilidad garantizada. Estos son:

- ***Secuenciación.*** Un proceso de envío TCP divide el *stream* dentro de una secuencia de segmentos de datos y los transmite como paquetes IP. Un número de secuencia es adjuntado a cada segmento TCP.
- ***Control de flujo.*** Es responsabilidad del emisor, no sobrecargar al receptor. Esto es logrado por un sistema de reconocimiento de segmento.
- ***Retransmisión.*** El emisor registra el número de secuencias de los segmentos que este envía. Cuando este recibe un reconocimiento, detecta que los segmentos fueron recibidos exitosamente y éste puede entonces borrarlos desde sus buffers externos.
- ***Buffering.*** El buffer interno en el receptor es usado para balancear el flujo entre el emisor y el receptor. Si el proceso receptor *recibe* operaciones más lentamente que las operaciones de *envío* del emisor, la cantidad de datos en el buffer aumentará.
- ***Checksum.*** Cada segmento lleva un checksum considerando el encabezado y el dato en el segmento.

Para la conectividad de esta aplicación se pensó como forma de comunicación la utilización de TCP con el uso de **sockets**.

Conceptualmente, un socket es una comunicación de punto final para la cual, una aplicación puede escribir datos que son enviados sobre la red subyacente, y desde la cual los datos pueden ser leídos. Un socket forma una abstracción sobre el punto final de comunicación actual que es usado por el sistema operativo local, por un protocolo de

transporte específico. En Steen & Tanenbaum (2002) se resumen las primitivas del socket para TCP que se muestran en la Tabla 3.1.

PRIMITIVAS	SIGNIFICADO
Socket	Crear un nuevo endpoint de comunicación
Bind	Adjuntar una dirección local a un socket
Listen	Anunciar la disponibilidad para aceptar conexiones
Accept	Bloquear llamadas hasta que una requisición de conexión llegue
Connect	Activamente intentar establecer una conexión
Send	Enviar algún dato sobre la conexión
Receive	Recibe algún dato sobre la conexión
Close	Liberar la conexión

Tabla 3.1: Primitivas del Socket para TCP/IP.

Los Servidores generalmente ejecutan las primeras cuatro primitivas, normalmente en el orden dado. Cuando se llama a la primitiva **socket**, el Cliente crea una nueva comunicación punto final para el protocolo de transporte. Internamente, al crear un punto final de comunicación, significa que el sistema operativo local reserva recursos para acomodar el envío y recepción de los mensajes para el protocolo especificado.

La primitiva **bind** asocia una dirección local, con el socket creado nuevamente. Por ejemplo, un servidor podría enlazar la dirección IP de su máquina junto con un número de puerto a un socket.

La primitiva **listen** es llamada solo en caso de comunicación orientada a conexión. Esto es una llamada sin bloqueo que permite al sistema operativo local recibir suficientes buffers para un número máximo especificado de conexiones que el cliente está dispuesto a aceptar.

La primitiva **accept** es una llamada para aceptar bloquear al cliente hasta que una requisición de conexión llegue. Cuando tal requisición llega, el sistema operativo local crea un nuevo socket con las mismas propiedades como el original, y lo retorna al cliente.

Ahora bien, por el lado del cliente, aquí también un socket debe primero ser creado usando la primitiva **socket**, pero explícitamente, no necesariamente enlazando el socket a una dirección local, desde el sistema operativo puede dinámicamente localizar un puerto cuando la conexión está configurada y lista. La primitiva **connect** requiere que el cliente especifique la dirección del nivel de transporte para la cual una requisición de conexión requiere ser enviada. El cliente es bloqueado hasta que una conexión ha sido configurada

exitosamente, después de lo cual ambos lados pueden iniciar en intercambio de información a través de las primitivas **write** y **read**, las cuales establecen el envío y recepción de los datos, respectivamente. Finalmente, el cierre de una conexión es simétrico cuando se usan sockets, y este cierre queda completado cuando tanto el cliente como el servidor llaman a la primitiva **close**.

Lo anterior se puede resumir en la Figura 3.10, en la cual se muestra la transferencia de los datos para la aplicación de navegación hablada.

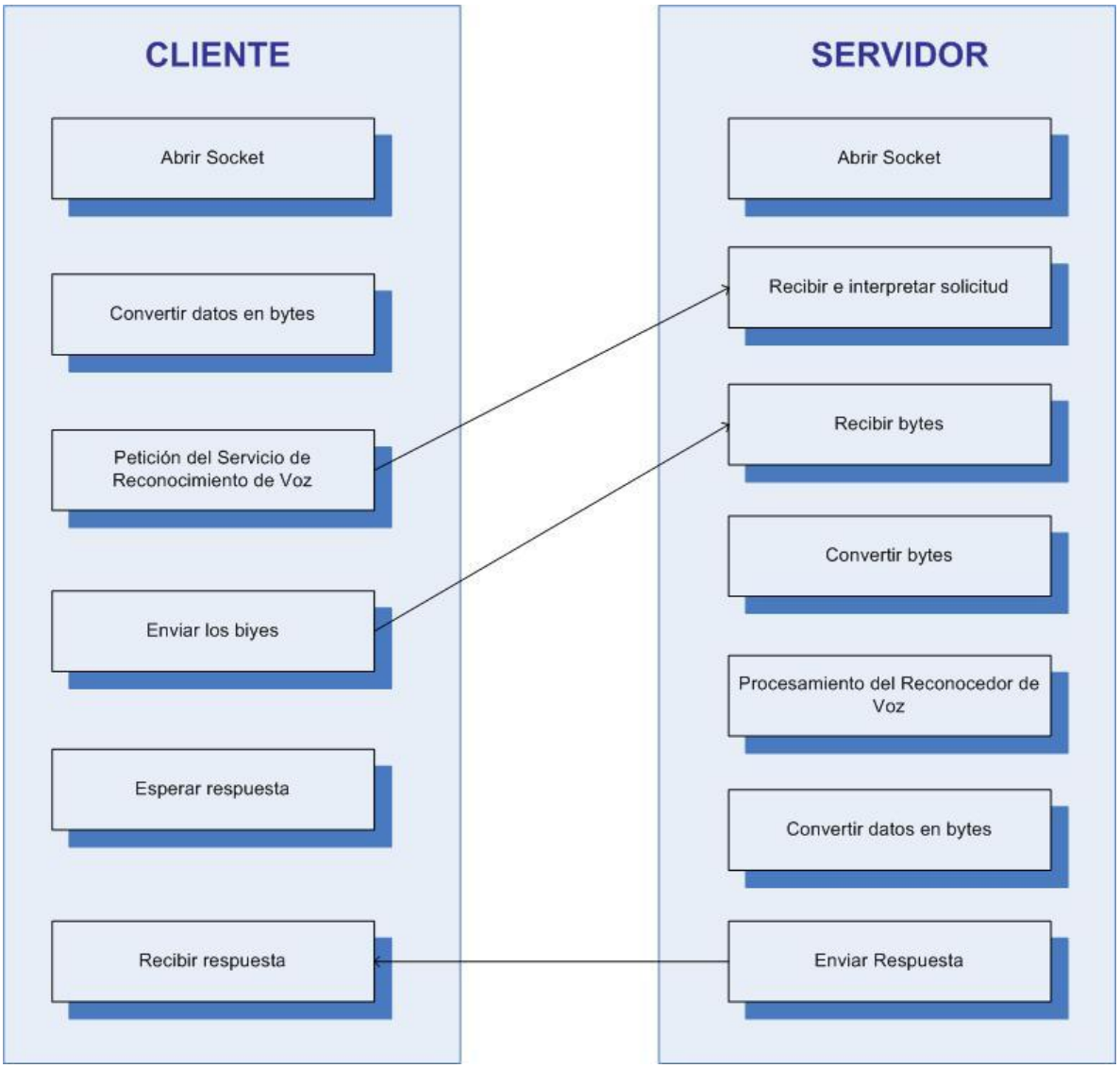

Figura 3.10: Transferencia de los datos.

Capítulo 4

Implementación de la Aplicación

En este capítulo se explica detalladamente cada uno de los pasos requeridos para llevar a cabo el desarrollo de la aplicación, en donde se involucran los distintos componentes principalmente de software. En la primera sección se habla del esquema de desarrollo de la aplicación, en la que se describen las distintas relaciones existentes entre los diversos elementos de desarrollo. En la segunda sección se especifica punto por punto la parte correspondiente al Reconocimiento de Voz, pieza clave para la realización de este proyecto. En la tercera sección se describe la fase de implementación de la aplicación, en la cuarta sección se especifica cómo está integrado el Servidor de Datos de PDLib con el Servidor remoto de Reconocimiento de Voz. Finalmente, en la quinta sección se muestran ejemplos de las interfaces que fueron diseñadas para la navegación de la aplicación.

4.1 Esquema de Desarrollo de la Aplicación

El esquema general mostrado en la Figura 4.1 describe el funcionamiento básico desde el punto de vista de la aplicación. Cabe indicar que este planteamiento es perfectamente válido para desarrollos con clientes de otras plataformas que soporten adicionalmente java.

El cliente dispone de un navegador que interpreta HTML, para solicitar metadatos al servidor web de PDLib; todo esto en conjunto, es con lo que van a interactuar las interfaces diseñadas específicamente para la navegación hablada, las cuales se encuentran contenidas dentro de la aplicación del cliente. Estas interfaces están diseñadas con el framework de .Net, que a su vez, interactúa con el Java 2 Micro Edition (J2ME) para comunicarse con el servidor de voz a través del Java 2 Standard Edition (J2SE), en donde el reconocedor de voz está desarrollado en c y corre en perl, montado sobre plataforma Unix.

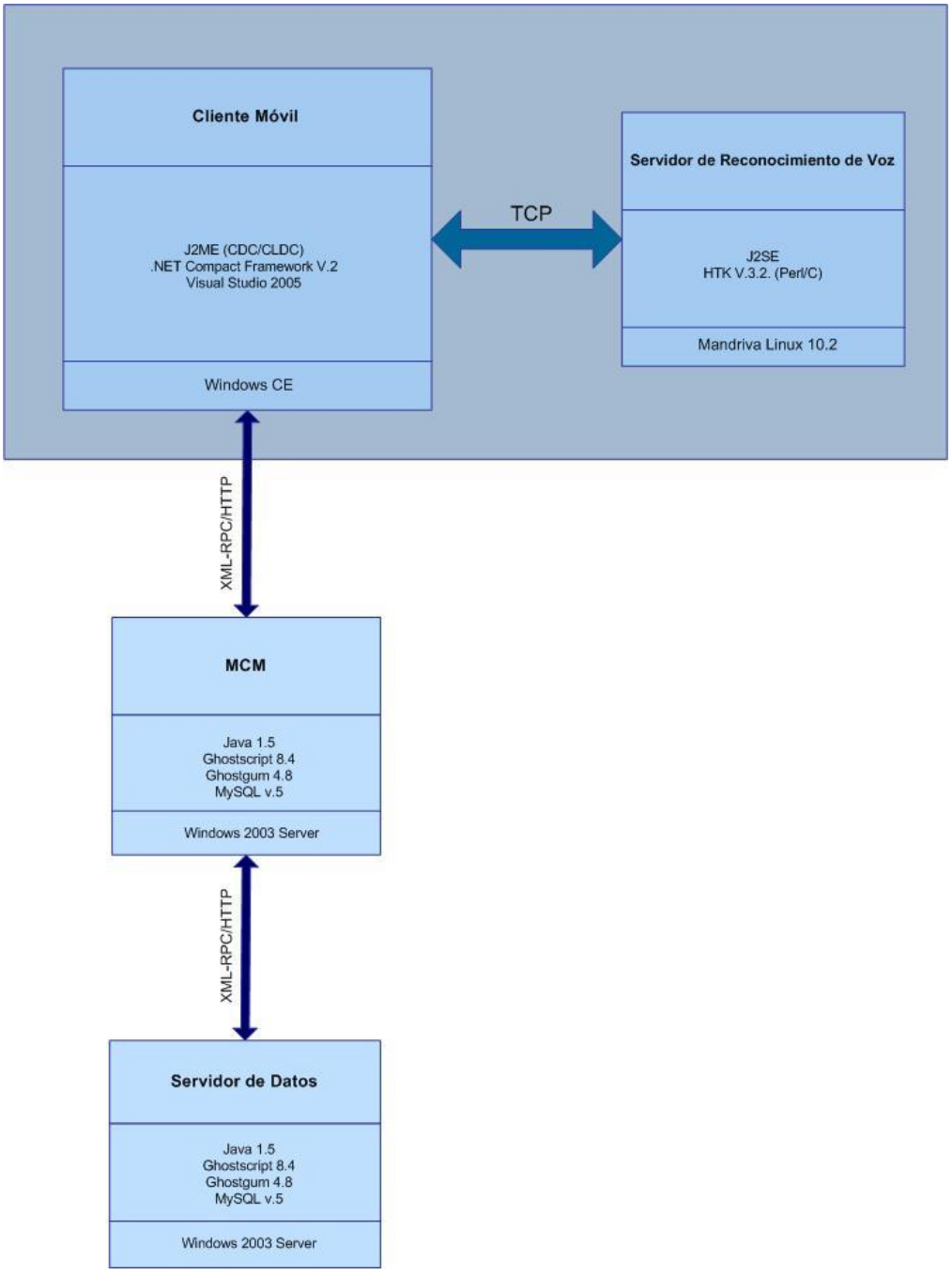

Figura 4.1: Esquema de Desarrollo de Navegación Hablada de PDLib.

Una de las grandes ventajas de esta aplicación es la aportación de portabilidad. De esta manera, se puede lograr la comunicación con el reconocedor de voz, desde cualquier cliente, sea ordenador personal o PDA. Se trata de un esquema de desarrollo de propósito general, y aplicable a cualquier plataforma cliente que disponga de java con un esfuerzo de

adaptación mínimo. No se requiere ninguna configuración específica de los clientes; así un programa .class escrito en un Windows CE puede ser interpretado en un entorno Linux, por ejemplo.

En la Tabla 4.1 se describen las especificaciones de software requeridas por cada uno de los componentes que integran el esquema de desarrollo en el entorno de navegación hablada.

COMPONENTE	ESPECIFICACIÓN	DESCRIPCIÓN
Cliente Móvil	Windows CE	Sistema Operativo abierto, escalable, de 32 bits que permite construir un amplio rango de pequeños dispositivos footprint. Cuenta con un "Constructor de Plataforma" para diversos componentes de Sistemas Embebidos. Incluye además, gran calidad en la productividad de OAL (OEM-Original Equipment Manufacturer-Adaptation Layer), drivers que utilizan menos código de hardware específico de la plataforma, enriquecido conjunto de herramientas de programación para crear código administrado y aplicaciones de código nativo ("Microsoft Windows CE", 2008).
	.NET Compact Framework	Ambiente de Hardware independiente para correr programas en dispositivos móviles. Éste corre sobre Windows CE y depende de un "Common Language Runtime" (CLR) que está diseñado para operar eficientemente cuando se corren programas sobre dispositivos con recursos limitados. El .NET Compact Framework contiene clases exclusivamente diseñadas para esta clase de dispositivos ("Microsoft. NET Compact Framework", 2008).

Tabla 4.1: Especificaciones de Software.

COMPONENTE	ESPECIFICACIÓN	DESCRIPCIÓN
Cliente Móvil	Visual Studio	Conjunto de herramientas de desarrollo para la construcción de aplicaciones Web .NET, Web Services XML, aplicaciones de escritorio y aplicaciones móviles que corren en cualquier plataforma soportada por el .NET Framework; estas plataformas abarcan Windows servers y workstations, Pocket PC, Smartphones y buscadores WWW. Visual Basic, Visual C++, Visual C# y Visual J#, todos usan el mismo ambiente de desarrollo integrado (IDE) ("Microsoft. NET Compact Framework", 2008).
	J2ME	Específicamente está enfocado al vasto espacio de los extremadamente "tiny comodities", tal como un PDA. J2ME mantiene las cualidades por las que la Tecnología de Java se ha vuelto famosa (Harkey et al., 2002): ▪ Consistencia incorporada a través de productos que se ejecutan en cualquier lugar, en cualquier tiempo y sobre cualquier dispositivo. ▪ Portabilidad de código. ▪ "Levantamiento" del propio lenguaje de programación Java. ▪ "Entrega segura" (dentro del entorno de red) ▪ Las aplicaciones J2ME son ascendentemente compatibles para trabajar con J2SE y J2EE.

Tabla 4.1: Especificaciones de Software (continuación).

COMPONENTE	ESPECIFICACIÓN	DESCRIPCIÓN
Servidor de Reconocimiento de Voz	Mandriva Linux	Antes Mandrakelinux y Mandrake Linux. Distribución Linux, que destaca por compilar sus paquetes optimizados para arquitecturas Pentium y procesadores avanzados que son incompatibles con las versiones de CPU viejas como 386. Mamdrakelinux soporta alrededor de 60 idiomas y usa el Centro de Control para la administración de Linux en vez de un editor de textos para cambiar las configuraciones. Tiene muchos programas conocidos como Drakes o Draks, en conjunto llamadas drakxtools, para configurar muchas aplicaciones y la mayoría de ellas pueden correr tanto en modo gráfico como en modo texto ("Mandriva", 2007). Cambios recientes en Mandriva Linux: • kernel 2.6.27 • KDE 4.1 • Mozilla Firefox 3 • OpenOffice.org 3 • Gimp 2.6 • Totem 2.24 • Amarok 2.0 • Arranque más rápido • Control parental mejorado • Instalación personalizada para equipos portátiles Requisitos mínimos: • Sistema operativo: Consola/X11 • Procesador: 1 GHz • Memoria: 512 MB • Espacio libre en disco: 2 GB

Tabla 4.1: Especificaciones de Software (continuación).

COMPONENTE	ESPECIFICACIÓN	DESCRIPCIÓN
Servidor de Reconocimiento de Voz	HTK	HTK, Hidden Markov Model Toolkit. Es un conjunto de herramientas de software para diseñar y manipular HMM (Hidden Markov Models). Originalmente fue creado para aplicarlo al desarrollo de sistemas ASR (Automatic Speech Recognition). Ahora puede utilizarse en cualquier área del conocimiento, la única restricción es que el problema a resolver pueda ser enfocado como un proceso de modelación Estocástico Markoviano. En la actualidad es exitosamente utilizado en: Reconocimiento y síntesis de voz, reconocimiento de caracteres y formas gráficas, análisis de vibraciones mecánicas, etc. El desarrollo de HTK lo lleva a cabo el grupo del habla, visión y robótica del Departamento de Ingeniería de la Universidad de Cambridge (CUED), UK. Actualmente HTK es de libre distribución y su código y librería pueden ser modificados en común acuerdo con el CUED. Además la herramienta se encuentra disponible para utilizarlo en diversas plataformas o sistemas operativos, tales como: Unix, Linux, Windows XP y DOS. HTK dispone de una arquitectura flexible y autosuficiente. Es controlado por módulos de librerías, que alimentan la interfaz de funciones correspondientes: Manejo de archivos, operaciones matemáticas e interacción con el sistema operativo (Carrillo, 2007).

Tabla 4.1: Especificaciones de Software (continuación).

COMPONENTE	ESPECIFICACIÓN	DESCRIPCIÓN
Servidor de Reconocimiento de Voz	J2SE	J2SE es Java 2 Standard Edition ("Sun Developer Network", 2008). Sus principales características de Portabilidad son: • API de JDBC para acceso a bases de datos • CORBA tecnología • Java secuity J2SE consta de dos componentes: Java Básico y Java Desktop. Java Básico proporciona funcionalidad back-end, mientras que Java Desktop proporciona GUI (interfaz gráfica de usuario) funcionalidad. J2SE contiene tanto el J2SE Development Kit (JDK) y Java Runtime Environment (JRE).

Tabla 4.1: Especificaciones de Software (continuación).

4.2 Reconocimiento de Voz

El Procesamiento de Lenguaje Hablado se refiere a las tecnologías relacionadas con el reconocimiento del lenguaje (o reconocimiento de voz), la conversión de texto a habla y el entendimiento del lenguaje hablado (Huang et al., 2001).

El objetivo del reconocimiento de voz es que las computadoras tengan la capacidad para comprender el lenguaje hablado y una vez entendido puedan ejecutar funciones específicas. Básicamente, un sistema de reconocimiento de voz debe cumplir tres tareas fundamentales ("Reconocimiento de Voz", 2008):

1. **Pre-Procesamiento.** Modela las pronunciaciones, dada la gran variabilidad de dichas señales.
2. **Reconocimiento.** Identifica lo que se dijo (traducción de señal a texto).
3. **Comunicación.** Envía lo reconocido al sistema (Software/Hardware).

La Figura 4.2 representa las tres tareas fundamentales de un sistema de reconocimiento de voz.

Figura 4.2: Tareas de un sistema de reconocimiento de voz.

Es difícil desarrollar un programa de computadora que sea lo suficientemente sofisticado para entender un discurso hablado por cualquier persona. Sólo cuando se simplifica el problema — aislando palabras, limitando el vocabulario, o restringiendo el modo en cómo se forman las frases— es posible el reconocimiento del lenguaje hablado. El lenguaje hablado está basado en una secuencia de segmentos de sonido discretos que son encadenados en el tiempo. Para estos segmentos, llamados fonemas, se asumen características acústicas y articulatorias únicas. Mientras que el aparato vocal humano puede producir casi un número infinito de gestos articulatorios, el número de fonemas es limitado. Cada fonema tiene características acústicas distinguibles, y, en combinación con otros fonemas forman unidades más grandes, como sílabas y palabras. Conocer estas características es lo que nos permite distinguir *ba* de *pa* (Huang et al., 2001).

Un sistema típico de reconocimiento de habla consiste de los componentes básicos mostrados en la Figura 4.3. Las aplicaciones se comunican con el decodificador para obtener resultados de reconocimiento que pueden ser usados para adaptar otros componentes en el sistema. Los *modelos acústicos* incluyen la representación del conocimiento acerca de la acústica, la fonética, las variables ambientales, las diferencias de género y dialecto entre los hablantes, etc. Los *modelos de lenguaje* se refieren al conocimiento del sistema de lo que constituye una posible palabra, qué palabras son posibles de co-ocurrir y en qué secuencia. La semántica y las funciones relacionadas a alguna operación que un usuario quiera realizar también pueden ser necesarias para el modelo de lenguaje. Existe mucha incertidumbre en esas áreas, relacionada con las características del hablante, la velocidad y el estilo del habla, el reconocimiento de segmentos básicos del habla, las palabras posibles, las palabras parecidas, las palabras desconocidas, la variación gramática, la interferencia de ruido, el acento no nativo, etc. Un sistema de reconocimiento exitoso debe considerar toda esta incertidumbre (Huang, Acero & Hon, 2001).

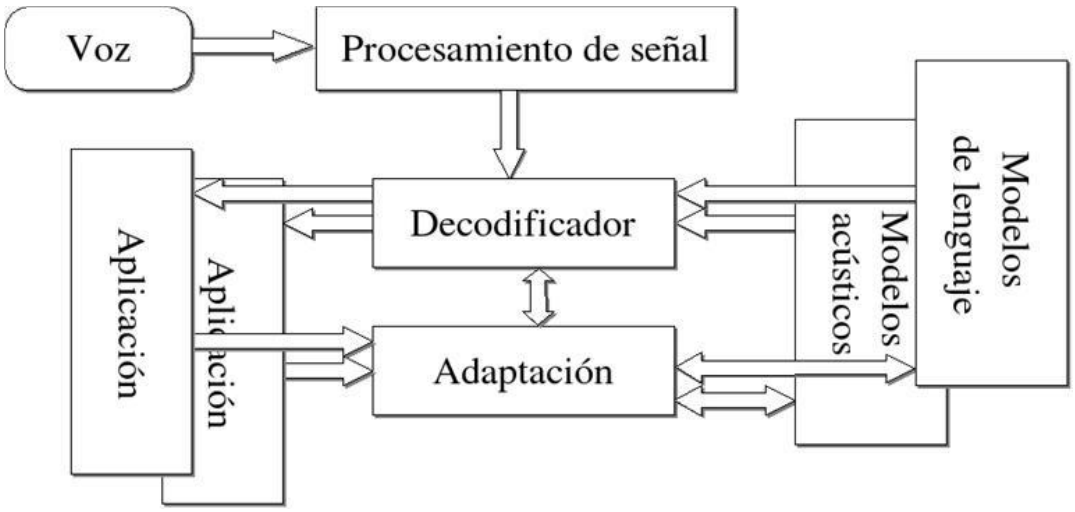

Figura 4.3: Arquitectura Básica de un Sistema de Reconocimiento del Habla.

La señal de voz es procesada en el módulo de procesamiento de señal, que extrae vectores de características para el decodificador. El decodificador usa los modelos acústicos y de lenguaje para generar la secuencia de palabra que tiene la máxima probabilidad posterior para los vectores de características de entrada. También puede proveer información necesaria para el componente de adaptación para modificar el modelo acústico o el de lenguaje y mejorar el rendimiento del sistema ((Huang, Acero & Hon, 2001).

En resumen, los elementos básicos para construir un reconocedor de voz son (Pérez, 2006):

- **Modelo Acústico.** Representación de un sonido basado en datos empíricos. El modelo acústico captura las propiedades acústicas de la señal de entrada, obtiene un conjunto de vectores de características que después compara con un conjunto de patrones que representan símbolos de un alfabeto fonético y arroja los símbolos que más se parecen.
- **Modelo de Lenguaje.** Proporciona la probabilidad de una secuencia de palabras y se obtiene del modelo de entrenamiento.
- **Diccionario de Pronunciación.** Un factor importante en el reconocimiento es el diccionario de pronunciación debido a que contiene la apropiada secuencia de símbolos que componen una palabra.

Estos elementos se relacionan como se representa en la Figura 4.4 (Pérez, 2006):

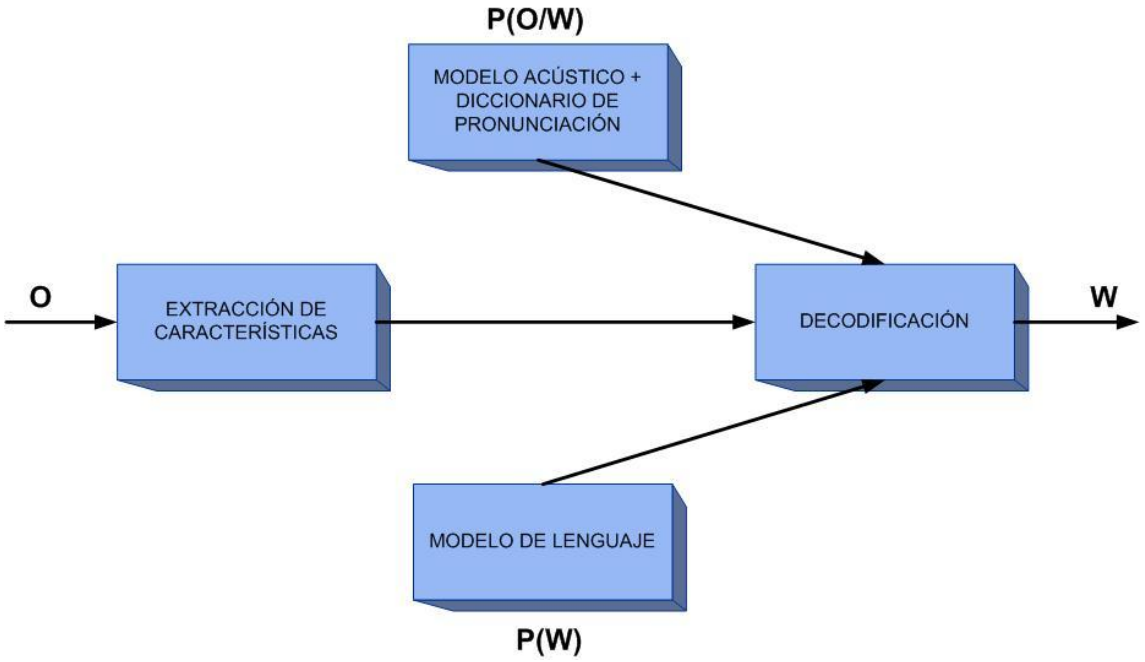

Figura 4.4: Elementos Básicos para construir un Reconocedor de Voz.

Donde:

W = Palabras
O = Medidas Acústicas
P(W) = Probabilidad de que la secuencia de palabras **W** sea pronunciada
P (O/W) = Probabilidad de que cuando una persona pronuncie la secuencia de palabras **W**, se obtenga la secuencia de medidas acústicas **O**.

Esto, explicado de forma general, consiste en utilizar "modelos" para modelar las pronunciaciones, dada la gran variabilidad de dichas señales.

Un sistema de reconocimiento de voz comprende dos etapas ("Reconocimiento de Voz", 2008):

1. **Entrenamiento.** Se le presentan al sistema una cantidad de pronunciaciones (elementos del habla: unidades básicas de las palabras, palabras, frases, oraciones, etc.) que se desea que éste "memorice".
2. **Reconocimiento.** Se le pide al sistema que identifique una pronunciación particular dada, como alguna de las que ya conoce o parecida a las que conoce o simplemente como desconocida. Esto significa que la pronunciación a reconocer no tiene que ser, necesariamente, una de las que se usan en la etapa de entrenamiento.

La información almacenada o retenida por el reconocedor está constituida por propiedades extraídas de todas las pronunciaciones de entrenamiento (No se almacenan las

pronunciaciones, sino propiedades de ese conjunto). Así se evita almacenar datos redundantes y con ello darle al sistema la propiedad de responder en forma rápida, a cualquier solicitud de identificación de alguna señal de entrada, lo ideal es que los sistemas respondan en tiempo real. La estructura general de un sistema de reconocimiento de voz consta de los siguientes módulos ("Reconocimiento de Voz", 2008):

- **Módulo de adquisición de datos.** Realiza la conversión analógica a digital.
- **Módulo de extracción de propiedades de la señal de voz**. Compresión de los datos para obtener un vector de propiedades (energía espectral, tono, formantes, donde empieza el sonido, donde termina el sonido, etc.) de cada segmento y de cada sonido de la pronunciación.
- **Módulo de cuantificación de los sonidos.** Identificar los distintos sonidos utilizando la secuencia de vectores de propiedades obtenida en el módulo anterior. Cada vector está asociado a un sonido del habla, luego la salida de este módulo es una secuencia de valores, donde cada valor representa el sonido con el que está asociado un vector de propiedades. Un mismo valor y por lo tanto un mismo sonido, puede aparecer varias veces en esta secuencia de salida.
- **Módulo reconocedor propiamente dicho.** Identifica una pronunciación dada, como conocida, parecida a una conocida o como desconocida. Para ello recibe desde el módulo de cuantificación la secuencia de valores que corresponde a una mezcla de los sonidos que puede tratar el sistema; estos sonidos individualmente corresponden a un segmento de la señal de la voz pero en conjunto y en la secuencia constituyen la señal completa de la pronunciación que se desea reconocer o memorizar. La complejidad de este módulo depende del tipo de identificación que se requiera.

La Figura 4.5 representa un diagrama del proceso de reconocimiento considerando los puntos que se mencionan arriba, con relación a su estructura general.

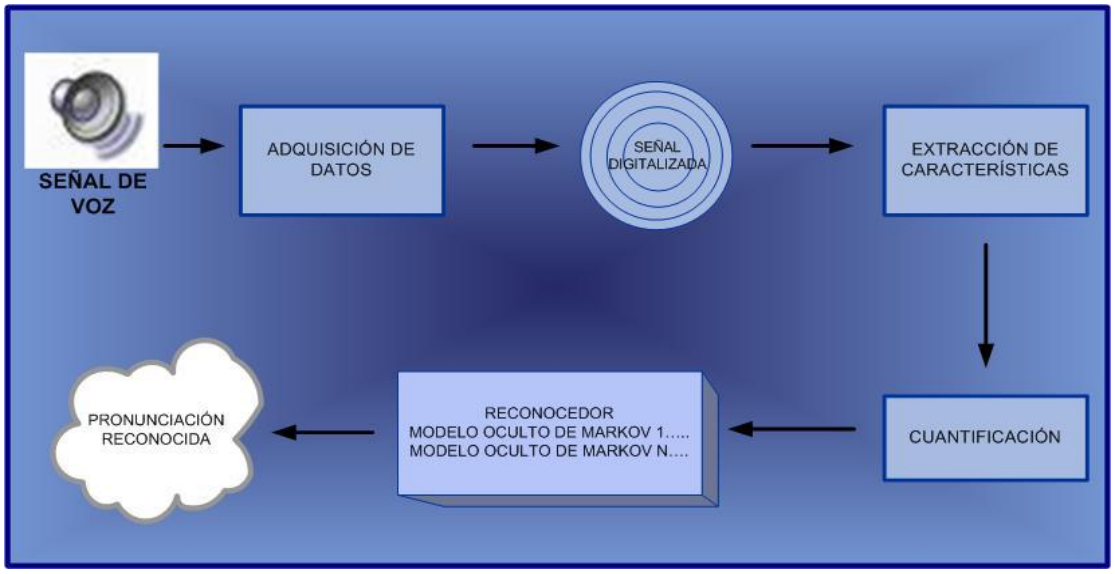

Figura 4.5: Proceso de Reconocimiento.

De acuerdo a Jelinek (1997), para hablar del problema de diseño de un reconocedor de voz se necesita una formulación matemática; para este efecto se propone mapear de una cadena de símbolos a otra, tanto en modelado de pronunciación para Reconocimiento Automático de Voz, como para la corrección de ortografía en OCR (Optical Character Recognition – Reconocimiento óptico de caracteres). En el reconocimiento de voz, dada una cadena de símbolos que representa la pronunciación de una palabra en un contexto, se busca la correspondiente cadena de símbolos en el diccionario de pronunciación. De esta manera se puede hablar de modelos probabilísticos de variación de pronunciación y ortografía, en particular, del modelo de inferencia de Bayes o canal ruidoso (ver Figura 4.6). Jelinek introdujo la metáfora del canal ruidoso en una aplicación del modelo para reconocimiento de voz en 1976. Esta idea consiste en tratar a la entrada (ya sea una mala pronunciación o una palabra mal escrita) como una instancia de la forma léxica (la pronunciación léxica o la correcta ortografía) que ha pasado a través de un canal de comunicación ruidoso. Este canal introduce "ruido" por lo que es difícil reconocer la palabra o elocución original. El objetivo es entonces la construcción de un modelo de canal ruidoso que permita imaginar cómo se modificó la elocución original y así poder recuperarla. Este ruido, tratándose de reconocimiento de voz, puede ser causado por variaciones en la pronunciación, variaciones en la realización del fonema, variación acústica debido al canal (micrófono, teléfono, la red), etc.

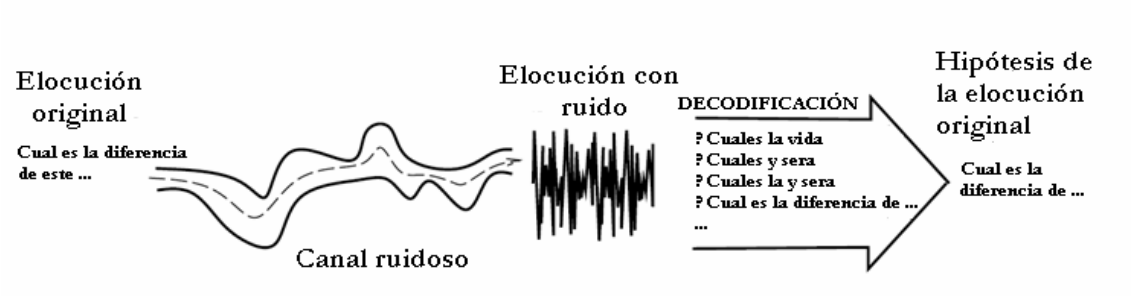

Figura 4.6: Canal Ruidoso.

En la actualidad se investigan algoritmos y modelos que permiten que el habla sea reconocida sin importar el hablante. Uno de los productos en esta área es el HTK, de la Universidad de Cambridge ("HTK Hidden Markov Model Toolkit", 2007), el cual forma parte de la realización de este proyecto. El HTK está diseñado principalmente para construir Modelos Ocultos de Markov basados en herramientas de procesamiento de voz, particularmente de reconocedores. Así, mucho del soporte de infraestructura en HTK está dedicado a esta tarea. Los Sistemas de Reconocimiento de Voz generalmente asumen que la señal de voz es una realización de algunos mensajes codificados como una secuencia de uno o más símbolos. Técnicamente, el rol del reconocedor es el efecto de un mapeo entre las secuencias de vectores de voz y las secuencias de símbolos deseados (Young et al., 2000).

De forma general la tarea realizada en la fase de reconocimiento para este sistema de navegación, fue representar cada palabra del vocabulario del reconocedor como un modelo generativo (que se calculara en la fase de entrenamiento) y posteriormente, se calculó la probabilidad de que la palabra a reconocer fuera producida por algunos de los modelos de la base de datos del reconocedor. Para ello, se asume que durante la pronunciación de una palabra el aparato fonador pudo adoptar sólo un número finito de configuraciones articulatorias o estados, y que desde cada uno de esos estados se produjo uno o varios vectores de observación cuyas características espectrales dependerían (probabilísticamente) del estado en el que se generaron. Vista la generación de la palabra, las características espectrales de cada fragmento de señal dependen del estado activo en cada instante, y las del espectro de la señal durante la pronunciación de una palabra dependen de la función de transición entre estados (Pérez, 2006).

4.3 La Fase de Implementación

Específicamente, la implementación de este proyecto involucró por un lado el entrenamiento previo de los comandos de voz permitidos para la navegación hablada, lo cual se llevó a cabo primeramente creando un archivo de "información" necesaria para el entrenamiento, desarrollo y prueba, definiendo, posteriormente el conjunto de los datos a utilizar para tal entrenamiento, haciendo un re-entrenamiento hasta alcanzar el mejor

desempeño de todas las redes generadas, aunque algunas veces la red que resultó de la primera vuelta tuvo el mejor desempeño. Para la fase de pruebas se asignaron 12 locutores. De los datos de entrenamiento se tomaron las muestras con la finalidad de que la red obtuviera el aprendizaje necesario por cada dato emitido. Es importante mencionar aquí que las redes neuronales utilizadas fueron las necesarias, para asegurar de este modo un mejor modelado y reconocimiento.

Posterior al entrenamiento fue necesario determinar cuál de las iteraciones fue la que tuvo el mejor desempeño en el conjunto de prueba. Para ello se buscó reconocer cada pronunciación en el conjunto de datos de desarrollo usando los pesos de la red de cada iteración. Aquí cabe aclarar que si el número de palabras en cada pronunciación no es conocido antes, entonces se evalúa el desempeño en cada iteración en términos de sustitución, inserción y borrado de errores. Si el número de palabras es conocido con anticipación, entonces sólo se miden los errores de sustitución, con el mismo método. Esto último fue lo que se llevó a cabo debido a que el número de palabras fue conocido. La exactitud de la red se midió como un **100% - (subs+ins+del)**, donde **subs** son los errores de sustitución, **ins** son los errores de inserción y **del** son los errores de omisión, todos ellos en porcentajes. Una vez que se obtuvieron los resultados, se eligió la mejor red con el nivel de exactitud en la palabra. Finalmente, se llevó a cabo la preparación de los datos.

La inicialización y el entrenamiento del modelo utilizó datos asociados con un modelo particular. Durante estos pasos de entrenamiento se asume que los límites fonéticos son definidos y que no hay interacciones entre los modelos vecinos. La re-estimación de parámetros direcciona estos problemas creando un modelo compuesto desde las transcripciones asociadas con este.

Por otro lado fue implementada la integración de los componentes relacionados: en primer lugar fue necesario que dentro de la aplicación se desarrollara la interacción del servidor de reconocimiento de voz con el dispositivo móvil, cuando este último enviara su petición, así mismo, se creó un canal de comunicación seguro, para que de este modo, el servidor de reconocimiento de voz respondiera a las peticiones enviadas por el cliente móvil. La intención de establecer un canal de comunicación con conectividad segura, fue pensada con la finalidad de que los datos enviados llegaran íntegros a su destino final.

Por último se realizó una readaptación de las interfaces ya existentes con la nueva funcionalidad de navegación hablada, para fines prácticos del usuario.

4.4 Integración del Servidor de Datos de PDLib con el Reconocedor de Voz

Dado que la implementación del reconocedor de voz fue parte de otro proyecto, en este capítulo sólo se ofrece una breve descripción de dicha implementación para entender más

fácilmente el contexto de la integración del reconocedor con el servidor de datos de PDLib.

Por el lado del reconocedor de voz, para realizar el reconocimiento del fonema, se utilizó un grupo de redes neuronales, definido como *cluster*, donde cada red neuronal en el grupo está especializada en un contexto del fonema, y donde el contexto es determinado automáticamente por las mismas redes neuronales durante el proceso de entrenamiento. Con esto, se logra que múltiples redes neuronales por fonema sean más eficientes que el esquema de una sola red neuronal para todos los fonemas (Peña, 2004). Una red neuronal es una estructura interconectada de unidades simples de procesamiento, en la que el conocimiento adquirido está asociado con las conexiones entre las unidades. Bengio (1996), propone el uso de redes especializadas en contexto para reconocer cada fonema, donde cada red neuronal es activada por un reprocesamiento especializado indicando a qué fonema corresponde la entrada a través de una señal codificada.

El modelo particular consiste en un método para entrenar cada cluster en un fonema específico, dentro del cual cada red neuronal está especializada en una realización acústica (fono) del fonema correspondiente al cluster al que pertenece la red. Cada red neuronal funciona como predictor no lineal de la forma de onda de la señal del fonema, haciendo una medición del error de esta señal predicha contra la original. Para un fonema se tienen n errores, uno por red neuronal de cada cluster. Estos n errores alimentan a una red neuronal clasificadora, que en última instancia es quien indica a qué fonema pertenece la señal de entrada (Peña, 2004).

Entrando ya en materia de lo relacionado a la integración del Servidor de Datos de PDLib, con el Reconocedor de Voz, dado que este último se encuentra hospedado físicamente en un servidor remoto al Servidor de Datos de PDLib, la forma en cómo se llevan a cabo las peticiones de un servidor a otro, es como lo muestra la Figura 4.7:

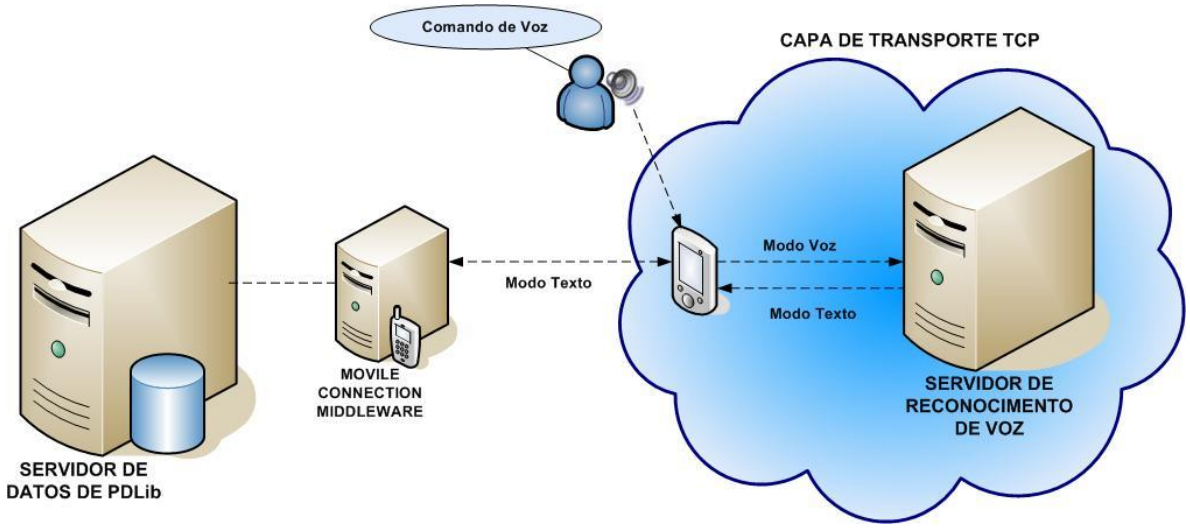

Figura 4.7: Esquema de comunicación entre el Servidor de Datos de PDLib y el Servidor Remoto del Reconocedor de Voz

En términos específicos, la Figura 4.7 describe la forma en la que, una vez que el usuario interpreta comandos de voz, estos se graban en el dispositivo móvil. Una vez grabados dichos comandos son enviados al servidor remoto en donde se hospeda el reconocedor de Voz. Para llevar a cabo el envío de los datos debe existir un enlace de comunicación entre el cliente móvil y el sistema de reconocimiento de voz. Esta comunicación es lograda a través del protocolo TCP (Ver sección 3.2.3 de este documento). Una vez que los comandos de voz son recibidos en el servidor del Reconocedor de Voz, estos son interpretados y traducidos a "modo texto". Cuando ya están convertidos en texto, el Servidor remoto los envía nuevamente al Cliente móvil, para que a su vez puedan ser enviados al Servidor de datos de PDLib, pasando previamente por el Movile Connection Middleware. Una vez que el servidor de datos obtiene el Comando de voz en modo texto, este lo reconoce y dependiendo de su significado, entonces provee los servicios propios de una biblioteca digital personal, almacenamiento de datos o el soporte de interoperabilidad que el cliente móvil necesita para que el usuario pueda realizar exitosamente la navegación hablada a través del dispositivo móvil. Cabe mencionar aquí que el Cliente móvil facilita la interacción entre los usuarios y la biblioteca digital con una interfaz de navegación parecida a la interfaz provista por los sistemas de archivos para acceder a las colecciones y documentos de la biblioteca.

Básicamente no existe una comunicación directa entre el Servidor de datos de PDLib y el Servidor del sistema de reconocimiento de voz. Toda la interacción se logra a través del dispositivo móvil que es también el medio por el cual el usuario navegará.

4.5 Diseño de las Interfaces de la Aplicación

En el diseño de la interfaz de voz, se utilizó un número finito de palabras, que previamente fueron entrenadas y posteriormente reconocidas en la navegación, por el reconocedor de voz; la mayoría de ellas fueron comandos que invocaban alguna acción dentro de la aplicación, y otras cuantas fueron usadas como ejemplo para el proceso de búsqueda de información que posee el sistema de bibliotecas digitales. El vocabulario que se utilizó fue un vocabulario completamente en inglés y dado que el tiempo de entrenamiento y reconocimiento para cada fonema definido, fue tardado, se limitó a un número pequeño de palabras y comandos permitidos, como se muestra en la Tabla 4.2.

COMANDOS DE VOZ					
Username	Logout	Full	And	Cancel	User
Password	Preferences	Text	Details	Stop	Import
Enter	My	Nested	Keyphrases	Hide	Move
Home	Favorites	Collection	Format	Java	Edit
Back	Browse	Document	Pdf	Send	Delete
Next	Search	View	Txt	Show	Tools
Pdlib	Look	As	Start	Download	New
Exit	For	List	Ok	Mail	Yes
No	Private	Public	Create	Name	Voice
PALABRAS DE BÚSQUEDA Y ATRIBUTOS					
Select	Author	Title	BabbleServer	Hardware	Two
I	Thought	You	Would	Find	This
Subject	To	L	zero	one	five
Guns	Michael	Eight	Six	from	useful

Tabla 4.2: Palabras entrenadas para la navegación hablada de PDLib

Dado que el sistema tiene la opción de que el usuario pueda navegar en "modo voz" y al mismo tiempo interactuar con la interfaz en "modo gráfico", para la parte de navegación, se utilizaron las mismas interfaces gráficas con que ya contaba la aplicación, haciendo algunas adaptaciones para el uso de comandos de voz, considerando como funcionalidad principal, en cada una de las pantallas desplegadas, el icono de "Voz" para activar el "Modo de Navegación Hablada".

La Figura 4.8 es la pantalla de *autenticación de usuario.* Únicamente los usuarios que cuentan con permisos de Acceso a PDLib pueden iniciar su sesión en la aplicación.

Figura 4.8: Pantalla de Autenticación del Usuario.

La Figura 4.9 es la pantalla de inicio para acceso al sistema, en ella se despliegan dos *opciones de navegación*, el usuario puede elegir si desea navegar a través de la interfaz gráfica o utilizando comandos de voz.

Figura 4.9: Pantalla que despliega las dos Opciones de Navegación.

La Figura 4.10 muestra la pantalla en donde se despliega la lista de los *comandos de voz permitidos* para navegar a través de la aplicación. Esta pantalla fue diseñada debido a que en el proceso de reconocimiento de voz, cada señal de voz es captada de forma diferente según el tipo de voz del hablante, y dependiendo también del canal de ruido que pueda existir al momento de transmitir la señal; por lo cual, antes de iniciar la navegación hablada, cada usuario debe pronunciar los comandos de la lista de comandos permitidos, y una vez realizada esta acción, la señal de voz, es captada por el reconocedor quien la debe registrar para que dicho usuario pueda iniciar su navegación.

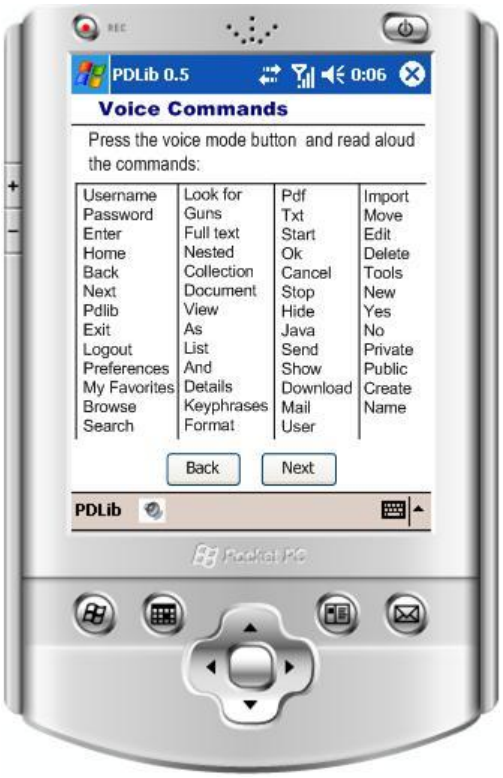

Figura 4.10: Pantalla que despliega la lista de Comandos permitidos para la navegación.

La Figura 4.11 muestra la pantalla principal del sistema y despliega como *ayuda* una lista de los comandos de voz que el usuario puede utilizar para navegar en esa pantalla. Así mismo aparece el icono de "voz" para activar o desactivar el modo de navegación hablada.

Figura 4.11: Despliegue de la lista de Comandos de la pantalla principal de la aplicación.

La Figura 4.12 muestra un ejemplo de *mensaje de Error* en caso de que uno de los comandos permitidos sea mal pronunciado por el usuario o bien, no exista en la lista de comandos permitidos.

Figura 4.12: Pantalla que despliega mensaje de Error cuando comando no existe.

La Figura 4.13 despliega la pantalla de *Browse* un usuario específico de PDLib, esto con la finalidad de poder ver su colección de documentos públicos.

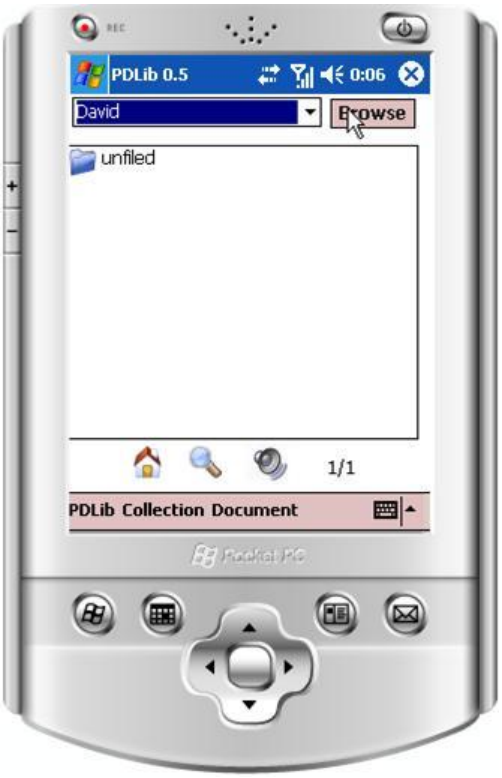

Figura 4.13: Pantalla de Búsqueda de un Usuario de PDLib.

La Figura 4.14 despliega la pantalla que muestra el archivo de *colección pública* de un usuario específico de PDLib y así mismo despliega la opción de *Search* (del menú de PDLib) de documentos del usuario.

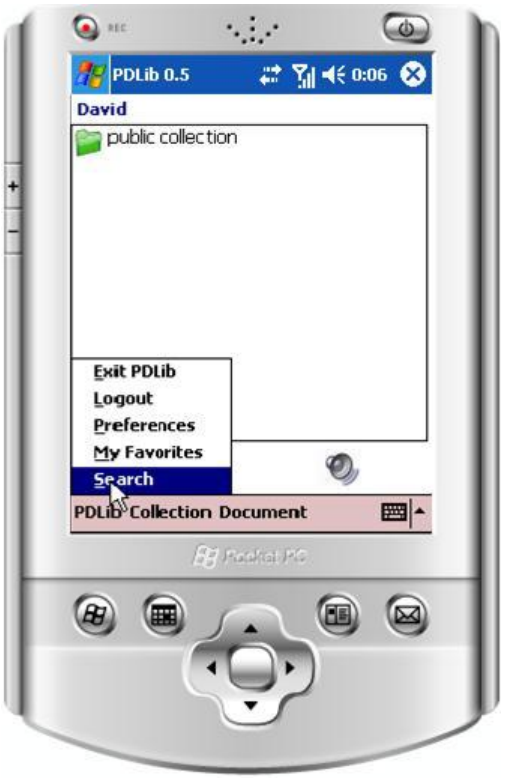

Figura 4.14: Pantalla donde se muestran las Colecciones de los Documentos de un usuario específico.

La Figura 4.15 muestra la pantalla en donde se especifica el documento que se está buscando.

Figura 4.15: Pantalla de especificación de la Búsqueda de un Documento.

La Figura 4.16 muestra la pantalla siguiente a la pantalla de especificación de la Búsqueda de un Documento. En esta pantalla se está realizando la acción de Search, cuando ya se han especificado las características del documento.

Figura 4.16: Pantalla donde se realiza la acción de Búsqueda de un Documento.

La Figura 4.17 muestra la pantalla que despliega los datos generales del documento que se buscó.

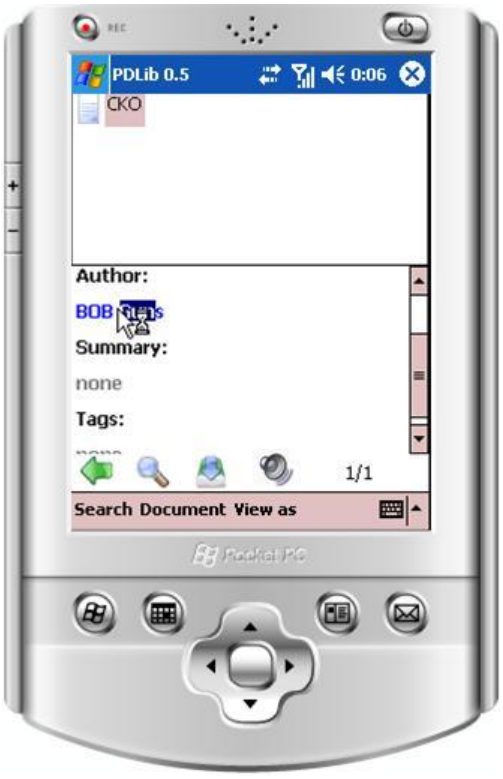

Figura 4.17: Pantalla con los Datos Generales del Documento que se buscó.

La Figura 4.18 despliega el contenido del documento que se buscó.

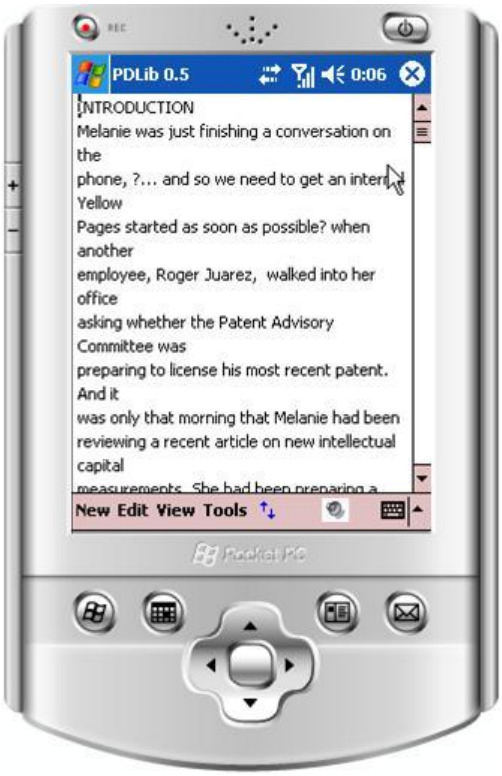

Figura 4.18: Pantalla que despliega el Contenido del Documento buscado.

La Figura 4.19 despliega la pantalla en donde se crea una nueva colección del usuario, siguiendo la opción *New* del menú *Collection*.

Figura 4.19: Pantalla para Crear una Nueva Colección del Usuario.

La Figura 4.20 muestra la pantalla en donde el usuario define las especificaciones para su nueva colección, como el nombre de la colección y si es una colección privada o pública.

Figura 4.20: Pantalla de especificaciones para la Nueva Colección Creada.

La Figura 4.21 muestra la pantalla en donde se realiza la acción *Create* de la nueva colección.

Figura 4.21: Pantalla donde se realiza la acción de Crear una Nueva Colección.

La Figura 4.22 despliega la pantalla para enviar un documento por mail. Siguiendo la opción *Send by Mail* del menú *Document*.

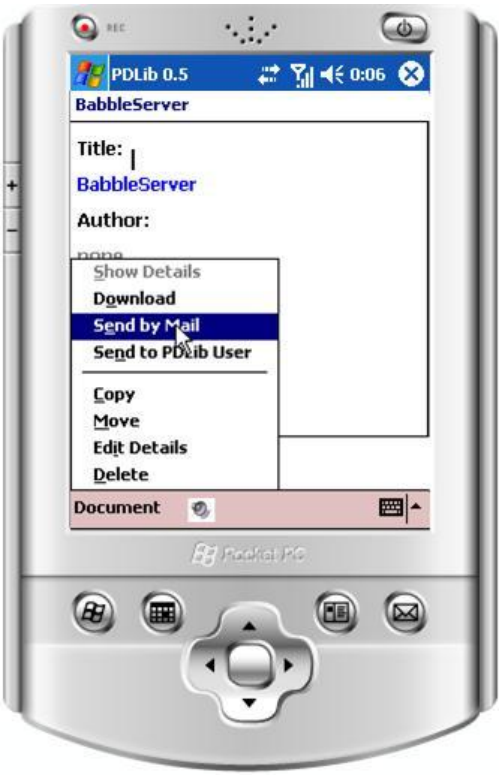

Figura 4.22: Pantalla para el Envío de un Documento por Mail.

La Figura 4.23 despliega la pantalla en donde se especifica el encabezado del mensaje (tema, destinatarios, mensaje) y donde también se especifica el tipo de formato en que será enviado el documento.

Figura 4.23: Pantalla que especifica los datos del destinatario del Mail que se envía del Documento.

La Figura 4.24 despliega la pantalla en donde se muestra un mensaje especificando que en documento fue enviado.

Figura 4.24: Pantalla que despliega el Mensaje de Envío del Documento.

La Figura 4.25 muestra la pantalla para finalizar la sesión del usuario en la aplicación.

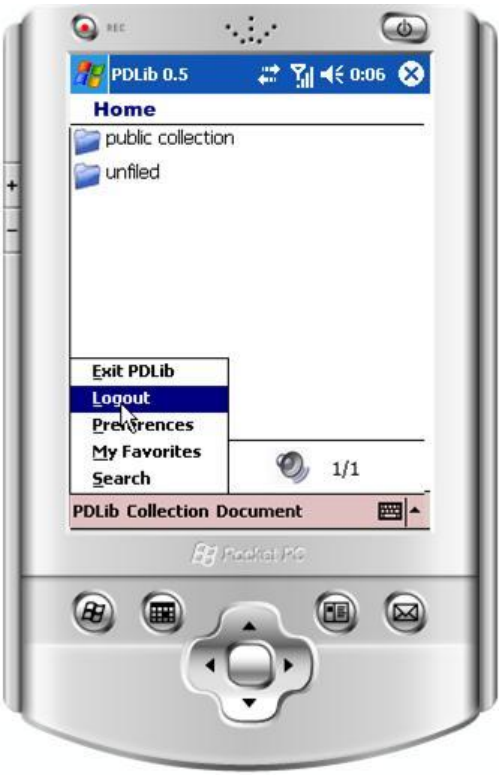

Figura 4.25: Pantalla que indica que el usuario cerrará su Sesión en la Aplicación.

Capítulo 5

Conclusiones y Trabajo Futuro

Este capítulo se encuentra dividido en dos secciones. En la primera sección se definen las conclusiones obtenidas de esta investigación, en donde se presentan así mismo, las contribuciones y ventajas derivadas de este proyecto. En la segunda sección, se sugiere el trabajo futuro que podría ser considerado como mejora de este tipo de aplicaciones.

6.1 Conclusiones

Como conclusión, específicamente se puede decir que el desarrollo de un sistema de navegación hablada puede resultar fácil de implementar si los elementos que lo componen son previamente diseñados pensando en una forma factible de adaptación entre ellos. Sin embargo la tarea de adaptar un sistema de reconocimiento de voz para cómputo móvil, cuando cada una de las partes involucradas es desarrollada por separado, en la mayoría de los casos resulta compleja, pues aunque el diseño de las interfaces, del lado del dispositivo móvil, es sencillo y práctico, la parte más compleja se centra en el hecho de poder desarrollar un buen sistema de reconocimiento de voz que, además de poseer elementos de calidad para el entrenamiento y el reconocimiento de los comandos, debe también contar con la característica de poder interactuar entre las diversas plataformas afines, a través de un enlace de comunicación adecuado entre la tecnología móvil y el sistema reconocedor de voz. Lo cual nunca es un trabajo sencillo de realizar si se parte del hecho en el que el reconocimiento de fonemas es un problema difícil de atacar, debido a las características propias de la señal de voz.

En este proyecto se detectó que muchos de los reconocedores de voz no tienen las características de software apropiadas para poder desarrollarse en un ambiente móvil, fue por eso que el desarrollo de esta aplicación tuvo muchas limitantes para poder integrar el mejor sistema de reconocimiento de voz para este ambiente de tecnología móvil y las bibliotecas digitales. Otro punto importante es que los comandos de voz que se utilizaron en el entrenamiento, tienen la limitante en cuanto a la forma en cómo cada usuario hace la emisión de tales comandos, por lo que cada vez que un nuevo usuario acceda a la aplicación, antes deberá “pronunciarlos”, para que el reconocedor pueda identificar así la voz del hablante y de esta forma iniciar la navegación hablada. Esto resulta en una tarea tediosa y quizá un tanto impráctica si lo que se pretende es que cualquier usuario pueda

hacer uso de la aplicación sin necesidad de invertir tiempo en un "reconocimiento" previo de su voz, cada vez que intente acceder al sistema.

El reconocimiento de todo el conjunto de fonemas de cualquier lenguaje hablado conlleva tiempo extensivo de investigación y de desarrollo. Por este motivo, en este proyecto se limitó a un grupo reducido de comandos de voz, por lo que el usuario solo podrá hacer uso de tales comandos para la navegación.

En relación al desarrollo de las interfaces, esta tarea fue muy fácil de llevar a cabo, debido a que en términos de diseño visual no se requirió de utilizar gran "creatividad" principalmente porque al ser un sistema de navegación hablada, el usuario no requiere más que de valerse de su propia voz para poder navegar. Sin embargo, el sistema cuenta también con la opción de que el usuario pueda alternar los dos modos de navegación al mismo tiempo (modo voz y modo texto), por lo que la interfaz gráfica implementada, se diseñó de una forma muy sencilla, debido a la limitada capacidad de almacenamiento y al reducido espacio en pantalla, con que cuentan los dispositivos móviles.

En relación a la parte del cliente (dispositivo móvil), es importante recalcar que dichos dispositivos cuentan con diversas restricciones debido a su principal característica: la movilidad, que también se convierte en su principal desventaja, ya que para logar tal movilidad, estos dispositivos son considerados como "dispositivos con capacidades limitadas" si se compara con dispositivos de cómputo tradicionales. Esta analogía se debe a que cuentan con métodos deficientes de entrada de información (por ejemplo, teclados pequeños si existen, mal reconocimiento de escritura y de voz); poca cantidad de recursos informáticos como memoria, periféricos y velocidad de procesamiento; pantallas demasiado pequeñas en las cuales la información no se puede desplegar de manera adecuada, baja calidad en sus interfaces de red inalámbricas (poco ancho de banda, desconexiones frecuentes, etc.) y por último, cuentan con un consumo de energía bastante alto y la capacidad de las baterías es sumamente limitado. Considerando todas estas limitantes, el proyecto PDLib montó la infraestructura apropiada para que la navegación a través del dispositivo se realice exitosamente, esto por el lado de la navegación en modo gráfico, sin embargo, haciendo un particular enfoque en la interfaz de voz implementada, estas limitantes están también presentes sobre todo cuando se trata de emitir los comandos de voz y que la señal sea transferida lo más clara posible para poder ser reconocida en el proceso de reconocimiento.

Las contribuciones principales en este proyecto fueron las siguientes:

- Se implementó un proceso de navegación hablada para aplicaciones móviles.

- Se definió un canal de comunicación apto, entre un sistema cliente (dispositivo móvil con interfaz de voz) y un sistema servidor (reconocedor de voz), independientemente de las plataformas sobre las que se montaron las aplicaciones de los diferentes sistemas; de esta forma cualquier dispositivo móvil de tipo PDA, puede interactuar con el reconocedor de

voz, sin importar el tipo de Sistema Operativo a utilizar, así como el lenguaje de programación utilizado para el diseño de las interfaces de navegación hablada.

- Se adaptó el sistema de reconocimiento de voz de tal forma que dicha adaptación no afectara de ningún modo la infraestructura planteada de PDLib. Con esto se logra que cualquier dispositivo móvil con la tecnología implementada del sistema de navegación de voz, pueda adaptarse de forma práctica a cualquier sistema de bibliotecas digitales en donde se permita la interacción de dispositivos móviles de tipo PDAs.

- Se montaron las interfaces de voz en adecuación con las interfaces gráficas ya existentes de PDLib. Esto facilita la navegación del usuario pudiendo interactuar a la par con ambas interfaces.

- El proceso de reconocimiento de voz se ejecuta por separado desde un servidor remoto, con esto la funcionalidad de los comandos de voz se concentra toda en dicho servidor, haciendo totalmente independiente la funcionalidad del dispositivo móvil, pues éste sólo se utiliza como medio para que el usuario navegue por las interfaces del sistema, logrando una interacción entre PDLib y la navegación hablada.

- Se creó un sistema cliente-servidor, que proporcione la facilidad para que el dispositivo móvil pueda interactuar a distancia y en cualquier tiempo.

Con las contribuciones antes mencionadas se pueden identificar las siguientes ventajas:

1. Una muy importante es la relacionada a las personas discapacitadas (invidentes, por ejemplo); estas personas pueden encontrar una gran ayuda en un software de navegación hablada, que permite transformar las órdenes de viva voz, para la búsqueda de documentos en una biblioteca digital. Así por ejemplo, pueden mediante órdenes dadas con la voz, encontrar información que sea de su interés, además de subir y bajar archivos de diverso contenido. Esto permite tener algo que se conoce como “hands-free computing”, es decir poder interactuar con la computadora sin necesidad de usar el teclado y/o ratón.

2. Permite acceso remoto. Otra de las ventajas del sistema es que está basado en dispositivos móviles, lo que permite acceder, directamente, mediante la captura de la voz, desde cualquier punto en donde se encuentre el usuario con conexión inalámbrica.

3. El sistema puede entender el lenguaje natural del usuario, es decir, entiende referencias que el usuario hace de algún pronombre. También puede manejar

comandos y pedir especificaciones cuando los comandos sean ambiguos o pedir información necesaria cuando así sea requerido.

4. Hace esta comunicación más rápida, y más agradable para el usuario, ya que al ser el habla la forma natural de comunicarse no se necesita ninguna habilidad especial, como en el caso de las interfaces gráficas en donde el usuario debe conocer más a detalle la orientación de los diferentes menús a utilizar para poder navegar a través de la aplicación.

5. Por ser el cliente móvil un dispositivo pequeño y portátil, permite el tener las manos libres para utilizarlas en alguna otra actividad, a la vez que se van dando órdenes por medio de la voz.

6. Permite movilidad, ya que la voz se puede enviar a distancia y ser recogida por un micrófono.

Para finalizar, la conclusión general a la que se llega es que los sistemas de reconocimiento automático de voz o habla, frente a otros sistemas de interacción hombre-máquina como teclados, paneles, etc., proporcionan una mayor naturalidad, así como un amplio rango de utilización por parte de diferentes tipos de usuarios en distintos entornos de operación. No obstante, a pesar de los grandes avances realizados, se está todavía muy lejos de un sistema de reconocimiento automático de voz universal que funcione bien en cualquier aplicación a la que sea destinado. En general, el diseño y las características de los actuales sistemas de reconocimiento automático de voz para cómputo móvil dependen fuertemente de la aplicación a la que van a ser destinados y a las condiciones de funcionamiento.

6.2 Trabajo Futuro

Dado que el área de la tecnología móvil va día a día en aumento, existen diversas propuestas de mejora para darle seguimiento a la interacción con la navegación hablada que se sugiere en este proyecto.

- *Optimización en el reconocimiento de los comandos de voz.* Este punto se refiere al aspecto de mejoramiento en el área de procesamiento para el reconocimiento de los fonemas, con relación a la distorsión que se produce cuando es transferida la señal de voz, lo que la vuelve en ocasiones un tanto "extraña" para poder ser interpretada.

- *Independencia del usuario final.* El grado de dependencia del usuario final, como locutor define si el sistema incorpora patrones de unidades lingüísticas adaptados a un locutor determinado, y, por tanto, sólo funcionará correctamente para él, o si los patrones pretenden ser válidos para cualquier

hablante. En el primer caso se habla de reconocimiento dependiente del locutor, mientras que en el segundo de reconocimiento independiente del locutor. A parte de las actividades específicas que se desarrollan para sistemas dependientes e independientes del locutor, existe un importante número de esfuerzos dirigidos a conseguir la adaptación de un reconocedor a un locutor específico con la menor cantidad de voz posible, por tanto, una de las limitantes presentadas en esta aplicación se presenta en el aspecto de que, al ser un sistema de reconocimiento dependiente del locutor, éste, tiene que emitir antes los comandos permitidos de voz para poder iniciar con la navegación. Más específicamente, para que el usuario inicie la navegación hablada, su voz debe ser identificada previamente por el sistema reconocedor. Esto sería más óptimo si se pudiera omitir este aspecto en el que el usuario pudiera iniciar la navegación sin la necesidad previa de emitir los comandos de voz permitidos.

- *Incremento y flexibilidad del vocabulario.* Las prestaciones de un reconocedor dependen fuertemente del tamaño y grado de dificultad del vocabulario. Es decir, del número de palabras que el sistema es capaz de reconocer, y de la mayor o menor dificultad de su reconocimiento en base a las relaciones de similitud fonética entre palabras. Otra importante dimensión, en relación con el vocabulario, es la que afecta a la distinción entre vocabularios fijos y flexibles. Una determinada aplicación, cuando esté reconociendo, siempre actuará sobre un vocabulario fijo. Pero en muchos casos ese vocabulario deberá variarse o actualizarse para eliminar y/o dar cabida a nuevas palabras. Tradicionalmente, una variación del vocabulario suponía comenzar un largo y costoso proceso de recogida de una nueva base de datos y re-entrenamiento de los patrones del sistema. En la actualidad hay diversas aproximaciones para conseguir un sistema con vocabulario flexible, que no necesite re-entrenarse para cada nuevo vocabulario. Esto sería lo ideal para aplicaciones robustas de navegación hablada en un ambiente de tecnología móvil, principalmente en aquellas que como, las aplicaciones de bibliotecas digitales, requieran de grandes cantidades de palabras a utilizar.

- *Realizar pruebas de aceptación.* El objetivo de las pruebas de aceptación consiste en validar que el sistema de navegación cumpla con el funcionamiento esperado y permitir al usuario de dicho sistema que determine su aceptación, desde el punto de vista de su funcionalidad y rendimiento, por lo tanto es necesario que como mejora de este trabajo, se lleven a cabo este tipo de pruebas entre un grupo de usuarios para determinar la eficiencia en la navegación hablada que se implementó.

- *Adaptación de un módulo de navegación hablada en todo el entorno de PDLib.* Esto se refiere a la implementación de un servicio de navegación de voz, tanto para los clientes móviles como para los clientes fijos de PDLib.

Bibliografía

Adam, N. R., Halem, M., Holowczak, R., Lal N. & Yesha, Y. (1996). Digital Library Task Force. *In IEEE Computer, 29(8)*. 89–91.

Alvarez, F., García, R., Garza, D.A., Lavariega, J. C., Gómez, L.G. & Sordia, M. (2005). *Universal Access Architecture for Digital Libraries*. *Proceedings of the 2005 conference of the Centre for Advanced Studies on Collaborative research, Toronto, Ontario, Canada,*12-28.

Alvarez, J.C., Costa, L. & Del Ser L. (2006). Movilok. (n.d.). *Interactividad Móvil.* Retrieved June 5, 2006, from http://www.madrimasd.org/revista/revista35/innovaciones/innovacion1.asp

Arms, W.Y. (1995). Key concepts in the architecture of the digital library. *D-lib Magazine*. Retrieved September 18, 2008 from http://www.dlib.org/dlib/July95/07arms.html

Arons, B., Back, M. Gaver, W., Hindus, D., Mynatt, E. & Stifelman, L. (1995). *Designing auditory interactions for PDAs.* Paper presented at the 8th annual ACM symposium on User interface and software technology, Pittsburgh, Pennsylvania, United Stat,143-146.

Avances en software de reconocimiento de voz. (2006). Retrieved December 08, 2008, from http://www.channelplanet.com/?idcategoria=15882

Bengio, Y. (1996). *Neural Networks for Speech and Sequence Recognition.* International Thomson Publishing.

Boyd, L.H., Boyd, W.L., & Vander-heiden, G.C. (1990). The graphical user interface: Crisis, danger and opportunity. *Journal of Visual Impairment and Blindness,* 496-502.

Bullinger, H. (1999). *Human-computer Interaction. Vol. 2, Communication, Cooperation and Application Design:* Proceedings of HCI International '99 (The 8th International Conference On Human-Computer Interaction). Publication: Mahwah, N.J.; London Lawrence Erlbaum Associates, Inc., Munich, Germany.

Buxton,W.(1986). Human interface design and the handicapped user. *In CH1'86 Conference Proceedings,* 291–297.

Carey, J. M. (1997*). Human Factors in Information Systems. Relationship between User Interface Design & Human Performance*. Publication: Norwood, N.J. Intellect Books.

Carrillo, R. (Ed.). (2007*). Diseño y Manipulación de Modelos Ocultos de Markov, utilizando herramientas HTK.Un Tutorial.* Retrieved October 26, 2008, from http://www.scielo.cl/pdf/ingeniare/v15n1/Art03.pdf

Casanovas, J. (2005). Interfaces de Voz IVR. Retrieved September 23, 2008, from http://www.alzado.org/articulo.php?id_art=431&s=1

Chepesuik, R. (1997). The future is here: America's libraries go digital. *American Libraries*, 2(1), 47-49.

Cleveland, G. (1998). Digital Libraries: Definitions, Issues and Challenges. Retrieved September 18, 2008 from http://www.ifla.org/VI/5/op/udtop8/udtop8.htm

Constantine, S. (2001). *User Interfaces for All: Concepts, Methods, and Tools.* Publication: Mahwah, N.J. Lawrence Erlbaum Associates.

Coulouris, G., Dollimore, J. & Kindberg, T. (Eds.). (2001). *Distributed Systems. Concepts and Design.* (3rd ed.). England: Addison Wesley.

D-Lib Magazine. (2008). Retrieved August 1, 2008, from http://www.dlib.org/

Defense Advanced Research Agency. (2008). Retrieved September 3, 2008 from http://www.darpa.mil/

Edwards, W. K. & Mynatt E. D. (1994). *An architecture for transforming graphical interfaces. Proceedings of the 7th annual ACM symposium on User interface software and technology, Marina del Rey, California,United States*, 39-47.

Escoffié, M. A. (2006). *Un Enfoque de Adaptación de Contenido Basado en la Abstracción Automática de Keyprhases para Bibliotecas Digitales en Ambientes Móviles.* Tesis de maestría, ITESM, Campus Monterrey.

Flores, E. (2006). *Diseño e Implementación de un Modelo Punto a Punto para la Interoperabilidad entre Servidores de una Biblioteca Digital.* Tesis de maestría, ITESM, Campus Monterrey.

Fox, E. & Sornil, O. (1999). Digital Libraries. In *Modern Information Retrieval* (Eds. R. Baeza_Yates & B. Ribeiro). pp. 415-432. Addison-Wesley-Longman Publishing co.

Galitz, W. O. (1997). *The Essential Guide to User Interface Design: An introduction to GUI design principles and techniques*. New York: John Wiley.

García, R. (2004). *Técnicas de adaptación a la conexión para clientes móviles que accesan servicios de biblioteca digital.* Tesis de maestría, ITESM, Campus Monterrey.

Garza, D.A., Gómez, L.G., Lavariega, J. C. & Sordia, M. (2003). Information retrieval and administration of distributed documents in internet. *In Witold Abra, editor, Knowledge Based Information Retrieval and Filtering from the Web.*

Garza, D.A., Gómez, L.G., Lavariega, J. C., Nolazco, J.A. & Sordia, M. (2004). PDLib: The personal digital library project. *Technical report, ITESM Monterrey.*

Graham, P.S. (1995a). Requirements for the digital research library. Retrieved September 18, 2008 from http://aultnis.rutgers.edu/texts/DRC.html

Greg, G. (1999). Mobile Devices Present Integration Challenges, IT Professional, *IEEE Computer, 3(1).* 11-15.

Harkey, D., Appajodu S. & Larkin M. (2002). *Wireless Java Programming for Enterprise Applications.* Editorial Wiley Publishing. 153-154.

Jelinek, F. (1997).*Statistical Methods for Speech Recognition*. MIT Press.

Jochems, W., Koper, R. & Merriënboer, J. (2004). *Integrated E-learning: Implications for Pedagogy, Technology and Organization. Open and Flexible Learning Series.* Publication: London, New York Taylor & Francis.

Johnson, J. (2000). *GUI Bloopers: Don'ts and do's for software developers and web designers.* Morgan Kaufmann, San Francisco, CA.

HTK Hidden Markov Model Toolkit. (2007). Retrieved October, 27, 2008, from http://htk.eng.cam.ac.uk/

Horton, S. & Lynch, P. J. (1999). *Web Style Guide: Basic design principles for creating web sites.* Yale University Press, New Haven, CN and London.

Huang, X., Acero, A. & Hon, H. (2001). *Spoken Languaje Processing, A Guide to Theory, Algorithm, and System Development.* Prentice Hall PTR.

Lynch, CA. & Garcia-Molina, H. (1995). Interoperability, scaling, and the digital libraries research agenda: a report on the May 18-19, 1995 IITA Digital Libraries Workshop. Retrieved September 18, 2008 from http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/main.html

Mandriva. (2007). Retrieved October 26, 2008, from http://www.mandriva.com/es

Microsoft. NET Compact Framework. (2008). Retrieved Agost 6, 2008, from http://msdn.microsoft.com/en-us/library/f44bbwa1.aspx

Microsoft. NET Compact Framework. (2008). Retrieved Agost 6, 2008, from http://msdn.microsoft.com/en-us/library/ms950416.aspx

Microsoft Windows CE. (2008). Retrieved Agost 6, 2008, from http://msdn.microsoft.com/en-us/library/ms905511.aspx

Mobile Magnifier para Pocket PCs V2.0. (2008). Retrieved December 08, 2008, from http://www.codefactory.es/es/products.asp?id=98

National Advisory and Space Administration. (2008). Retrieved September 9, 2008 from http://www.nasa.gov/

National Science Foundation. (2008). Retrieved February 08, 2008, from http://www.nsf.gov/

OAI Initiative (2008). Retrieved February 15, 2008, from http://www.openarchives.org/

OWASYS 22C y 112C. (2003). Retrieved December 08, 2008, from http://www.owasys.com

PDLib: The Personal Digital Library Project Webpage, ITESM-Campus Monterrey. (2006) Retrieved September 10, 2006, from http://copernico.mty.itesm.mx/pdlib

Peña, M.O. (2004). *Agrupamiento de la Predicción de la Señal de Voz en una Tarea de Reconocimiento Automático de Fonemas.* Tesis de maestría, ITESM, Campus Monterrey.

Reconocimiento de Voz. (2008). Retrieved October 30, 2008, from: http://www.pas.deusto.es/recursos.htm

Shneiderman, B. (1992). *Designing the User Interface: Strategies for effective human-computer interaction*. (2dn edn), Readin, MA:Addison Wesley.

Sharp, H., Rogers, Y. & Preece, J. (2007). *Interaction design: beyond human-computer interaction*. New York, NY: John Wiley & Sons.

Speaker at the SVOX Forum 2008 - The Ultimate User Experience. (2008). Retrieved December 08, 2008, from http://www.svox.com/

SpeechWorks Releases ETI-Eloquence Small Footprint -SF- Text-to-Speech Engine for Handsets and other Mobile Devices. (2003). Retrieved December 08, 2008, from http://findarticles.com/p/articles/mi_m0EIN/is_/ai_100455671

Steen, M. & Tanenbaum, A.S. (Eds.).(2002).*Distributed Systems.Principles and Paradigms.*New Jersey: Prentice Hall.

Sun Microsystems. The java platform for consumer and embedded devices. (2008). Retrieved February 29, 2008, from http://java.sun.com/j2me/docs/

Sun Developer Network (SDN). (2008). Retrieved October 26, 2008, from http://java.sun.com/javase/index.jsp

TALKS for Series and Nokia Communicator. (2008). Retrieved December 08, 2008, from http://www.nuance.com/talks

Thackara, J. (2001). The design challenge of pervasive computing. *Interactions,* 47-52.

The Association for Computing Machinery. The ACM portal. (2008). Retrieved September 5, 2008, from http://portal.acm.org/dl.cfm

The Owasys 22C. (2007). Retrieved December 08, 2008, from http://www.screenlessphone.com/

Tidwell, J. (2005). *Designing Interfaces.* Editorial O'Reilly Media , Inc. 21-22.

Tufts University. The Perseus Digital Library. (1998).Retrieved September 5, 2008, from http://www.perseus.tufts.edu/

Van der Harstt, G. & Maijers, R. (1999). *Effectief GUI ontwerp Een praktische ontwerpaanpak voor browser en Windows Interfaces (Effective GUI design: a practical design approach to browser - and windows Interfaces) Academic Service.* Schoonhoven, Netherlands.

Witten, I. (2003). *The new zealand digital library project*. Department of Computer Science, University of Waikato, New Zealand.

Winograd, T. (1997). From computing machinery to interaction design. In P. Dening and R. Metcalfe (eds), *Beyond Calculation: The Next Fifty Years of Computing. Springer-Verlag,* pp. 149-162. Amsterdan.

Young, S., Kershaw, D., Odell, J., Ollason, D., Valtchev, V. & Woodland, P. (2000). *The HTK Book.* Retrieved October 27, 2008, from http://nesl.ee.ucla.edu/projects/ibadge/docs/ASR/htk/htkbook.pdf