

IMAGINE: Interfacing Mobile Application with Voice Natural Language Interactivity

Organismo financiador: European Commission, IST-2000-29490

Grupos participantes en el proyecto:

Software AG España, S.A. (SAGE)

IT Innovation (ITI)

Institut für Angewandte Informationsforschung (IAI)

Royal National Institute for the Blind (RNIB)

Iniciativas Digital Media – Grupo La Vanguardia (IDM)

Telefónica Investigación y Desarrollo (TID)

Coordinador y Socio tecnológico

Socio tecnológico

Socio tecnológico

Usuario

Usuario

Socio tecnológico

Persona de contacto: Cristina Arana Ferrándiz. Software AG España, S.A. Ronda de la Luna, 4. Tres Cantos, E-28760 Madrid Tel: +34 91 8079400 - fax: +34 91 8079447 email: carana@softwareag.es

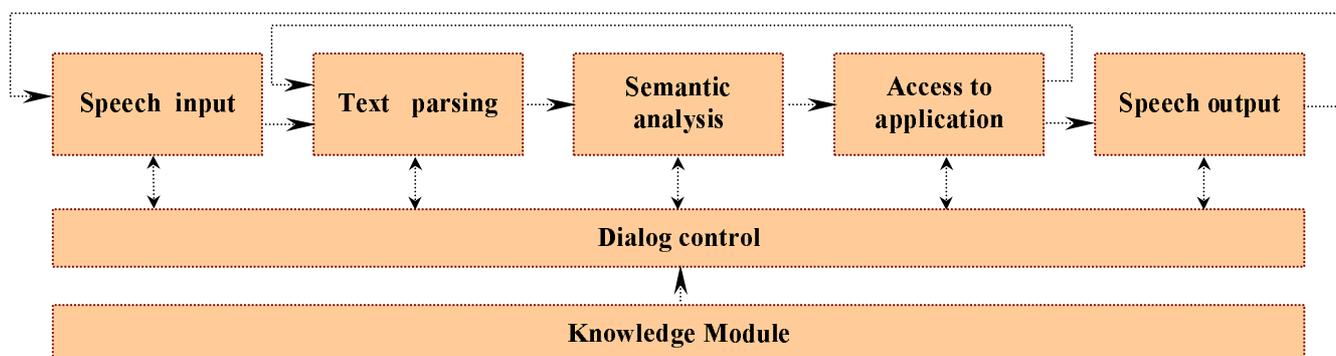
Resumen

IMAGINE es la evolución natural de MELISSA¹, proyecto financiado por la comunidad europea cuya función es proporcionar una interfaz en lenguaje natural a aplicaciones software. IMAGINE toma como punto de partida MELISSA, reforzando algunas características y añadiendo otras. Actualmente IMAGINE proporciona acceso a aplicaciones explotadas a través de Internet. Las principales características de IMAGINE son:

- Interfaz de lenguaje natural por voz y/o texto, sobre aplicaciones software previamente desarrolladas.
- La aplicación no debe modificar nada para ser accedida mediante IMAGINE.
- El protocolo de acceso a las aplicaciones es abierto: HTTP, RMI, Web services, ...
- La adaptación de IMAGINE a una aplicación se basa en información de tipo declarativo, describiendo las funcionalidades de la aplicación a acceder, y en la actualización de las gramáticas de los módulos lingüísticos.
- El gestor de diálogo tiene una iniciativa mixta, por el usuario o por IMAGINE.
- Multilingüalidad. El núcleo de IMAGINE no es dependiente del idioma. Las restricciones semánticas se obtienen de la modelización de la aplicación y se aplican independientemente del idioma.

Proceso completo de IMAGINE

La entrada se produce por voz o texto, al igual que la salida. El control del diálogo organiza todos los pasos e interacciones con el usuario utilizando la información de la aplicación contenida en el Knowledge Module. Cuando se han completado los datos para la ejecución de una funcionalidad, esta se ejecuta en Access to Application, dando el resultado al usuario e iniciándose de nuevo el ciclo.



¹ MELISSA: ESPRIT 22252, SEPLN 1999. <http://www.rtd.softwareag.es>

Arquitectura

La arquitectura de **IMAGINE** es una evolución de la de **MELISSA**:

- se ha adaptado para su uso en Internet,
- se ha añadido entrada y salida por voz,
- se ha incorporado un gestor de diálogo.

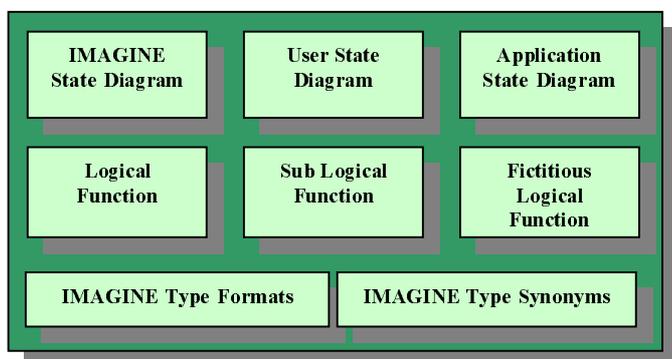
Diálogo

El diálogo es el núcleo de **IMAGINE**. Todos los componentes del sistema actúan como recursos del diálogo. La ejecución de **IMAGINE** se puede resumir en una secuencia de interacciones alternativas entre el usuario y el sistema. Es posible romper esta secuencia si el usuario adelanta intervenciones antes de que **IMAGINE** las espere (barge in). Esta función es soportada pero no incluida en los prototipos.

La secuencia del diálogo se inicia con un mensaje de bienvenida de **IMAGINE** adecuado a la aplicación que se esté accediendo. A partir de aquí es el usuario, en su propio lenguaje, quién demanda una funcionalidad concreta. No hay un sistema de menús que dirija al usuario. En su lugar, los mensajes adecuados, y la ayuda, proporcionan la información suficiente.

El diálogo basa su gestión en las interacciones previas del usuario y el modelo de la aplicación. Este modelo se compone de:

- Taxonomía con todos los tipos de datos involucrados en la aplicación.
- Funcionalidad de la aplicación, estructurada en Funciones Lógicas de varios tipos.
- Mapas de estado de la funcionalidad desde diferentes puntos de vista: del usuario, de **IMAGINE** y de la propia aplicación.



Herramientas de voz

Se utilizan dos sistemas de voz (in/out) uno para español y otro para inglés. El español es un sistema propietario de TID basado en una aproximación estadística en el que se ha

desarrollado un modelo de lenguaje para un prototipo español. Para un prototipo inglés, se ha utilizado tecnología de Nuance® con una aproximación gramatical. El diálogo ordena las gramáticas por probabilidad según el contexto de ejecución, lo que aumenta el rendimiento del módulo de voz. Ambos componentes además permiten acceso telefónico.

Analizador lingüístico

El analizador lingüístico distingue un pre-proceso y un proceso lingüístico propiamente dicho. La primera fase realiza un análisis de las propiedades más superficiales (formato del texto, delimitación de los componentes, aspectos morfológicos), mientras que en la segunda se tratan los aspectos más profundos (análisis morfosintáctico, estructura de constituyentes, representación semántica), produciendo un completo análisis de la frase propuesta por el usuario. Es de destacar que la versión española se basa en el uso de ALEP.

Analizador semántico

Una vez que el analizador sintáctico semántico ha generado su estructura de salida independiente del idioma, es el analizador semántico el que traduce esta información a conceptos relacionados con la aplicación. Obtiene el sentido y objetivo de la frase en términos de las funciones que se pueden ejecutar. También localiza conceptos concretos, dependiendo del contexto del diálogo, en las respuestas del usuario a preguntas detalladas hechas por **IMAGINE**.

Front End

El denominado Front End es el módulo encargado de acceder efectivamente a la aplicación. Se compone de un módulo común y diferentes adaptadores para los diferentes protocolos. Actualmente realizan accesos por Http/html y Http/wml. Fácilmente se le pueden añadir adaptadores para RMI, Web services, ...

Adaptador de Respuesta

Dado que las aplicaciones accedidas tienen su propio interfaz, habitualmente gráfico, se necesita el adaptador de respuestas para obtener una salida en formato textual representativa de la respuesta de la aplicación. Esta respuesta puede ser enviada al usuario en forma escrita o a través del módulo TTS (Text To Speech).