

Diseño, Desarrollo y Evaluación de un Sistema de Decisión y Soporte: La Metodología MAKAVAS

Paolo Rosso¹

Resumen

Los sistemas de decisión y soporte son sistemas informáticos que ayudan en la toma de decisiones. La evaluación de un sistema de decisión y soporte, es decir la medición de su calidad, es una componente esencial de su diseño y desarrollo. La evaluación permite guiar el proceso de diseño y desarrollo del prototipo y no tiene que ser realizada al final del mismo. En este artículo describimos como diseñar, desarrollar y evaluar un sistema de decisión y soporte utilizando una metodología iterativa con prototipado. Con este fin, introducimos la metodología MAKAVAS, obtenida aplicando a la metodología de evaluación KAVAS el enfoque 'multi-facetas' de Adelman. Como test para la metodología, presentamos los resultados de la evaluación de un sistema de decisión y soporte para logopedas.

Palabras Claves: *sistema de decisión y soporte, metodología de evaluación iterativa con prototipado, enfoque multi-facetas.*

Abstract

Decision support systems refer to interactive computer-based systems which help decision-makers to make a decision. Evaluation of a decision support system, that is, the act of measuring its quality characteristics, is an essential component of its design and development if considered an integral part of it and not employed at the end of its completion. In fact, evaluation permits to keep the development process of the prototype on track. This paper describes how to design, develop and evaluate a decision support system using an iterative-prototyping. This is the aim we introduce the MAKAVAS evaluation methodology for, which was obtained enhancing the KAVAS methodology with the multifaceted approach introduced by Adelman. As benchmark, the evaluation of a decision support system for profiling disordered phonology in childhood was carried out following the principles of MAKAVAS.

Keywords: *decision and support system, methodology for iterative evaluation and prototyping, multifaceted approach.*

1 Introducción

Los Sistemas de Decisión y Soporte (SDS) representan un medio de soporte para tomar decisiones poniendo énfasis, a la vez, en el proceso de decisión y en el soporte de dicho proceso. La intención

¹Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, España, e-mail: proso@dsic.upv.es.

es diseñar SDS's para soportar la toma de decisiones llevadas a cabo por personas sin que estas sean remplazadas. La evaluación es un componente esencial en cuanto se considera parte integral del proceso de Diseño y Desarrollo (D&D) y no sólo al final de la implementación del SDS. Eso permite determinar que problemas se encuentran en cada etapa del desarrollo y puede ayudar a predecir posibles razones de fracaso e indicar remedios apropiados, ahorrando tiempo y esfuerzo. Por las razones que acabamos de mencionar, la evaluación tendría que empezar en el momento de la concepción del SDS. Para el equipo de diseño y desarrollo, las observaciones obtenidas por la evaluación continuada son generalmente mucho más importantes que las prestaciones obtenidas por el SDS, ya que indican la dirección hacia la cual el proceso de D&D tendrá que moverse. De esta manera, la evaluación indica al mismo tiempo deficiencias y tareas para investigaciones futuras.

Por si sola, la evaluación de las prestaciones de un sistema de decisión y soporte no es suficiente. Para la aceptabilidad del SDS, la evaluación de la interacción con el usuario es casi más importante que la validación de las respuestas del sistema. Por consiguiente, la tarea de la evaluación consiste en medir las prestaciones de un sistema y en verificar su aceptabilidad para guiar, por completo, el proceso de D&D del prototipo del sistema de decisión y soporte.

El objetivo de este artículo es describir, de una manera sencilla, como diseñar, desarrollar y evaluar un sistema de decisión y soporte a través de una metodología iterativa con prototipado. Con este fin, utilizamos como ejemplo la metodología KAVAS (Knowledge Acquisition Visualisation and Assessment System) [1], que extendemos aplicando a la evaluación el enfoque multi-facetado de Adelman [2]. La metodología KAVAS utiliza algunos conceptos formales del enfoque iterativo con prototipado de Boehm [3]. Al igual que otras muchas metodologías de evaluación, KAVAS utiliza el término evaluación sin distinguir entre las diferentes facetas de la propia evaluación. El método multi-facetado de Adelman enfoca la evaluación desde una perspectiva diferente: teniendo en cuenta que un SDS puede ser evaluado subjetivamente (obteniendo las opiniones de los usuarios acerca de sus puntos fuertes y débiles), técnicamente (mirando dentro de la 'caja negra': la atención es en el algoritmo y en la entrada / salida), y empíricamente (evaluando estrictamente las prestaciones del SDS). Este enfoque permite integrar los diferentes criterios usados al final de cada fase de la metodología KAVAS para evaluar el prototipo. La metodología de evaluación MAKAVAS (cuyo acrónimo significa Multifaceted Approach to KAVAS) que utilizamos ha sido obtenida aplicando el enfoque multi-facetado de Adelman a la metodología iterativa con prototipado KAVAS [4].

2 La Metodología de Evaluación KAVAS

La evaluación de un SDS tiene los siguientes objetivos: guiar el diseño y el desarrollo del prototipo, cuantificar los diferentes aspectos de las prestaciones del sistema, y verificar su aceptabilidad. La metodología de evaluación KAVAS pone mucho énfasis en el ciclo de vida de diseño y desarrollo del sistema de decisión y soporte. El objetivo es proporcionar herramientas de validación y métodos consistentes y coherentes para guiar el desarrollo y la evaluación del SDS desde la concepción del prototipo (es decir, el diseño) a través de su desarrollo iterativo. En otras palabras, la intención es proporcionar un formato estructurado en el cual pueda tener lugar el desarrollo iterativo del prototipo. La técnica del prototipado sigue los principios básicos de la filosofía KADS [5]. La metodología está basada en un ciclo iterativo de cuatro fases. Las fases de la evaluación están representadas en la

Figura 1:

Fase I: primer desarrollo del prototipo de SDS o *exploración preliminar*.

Fase II: evaluación de la *validez* de la base de conocimiento del SDS.

Fase III: evaluación de la *funcionalidad* en circunstancias controladas (test de laboratorio).

Fase IV: evaluación del impacto en el escenario de trabajo (pruebas de consecuencias a largo plazo de aplicación del SDS).

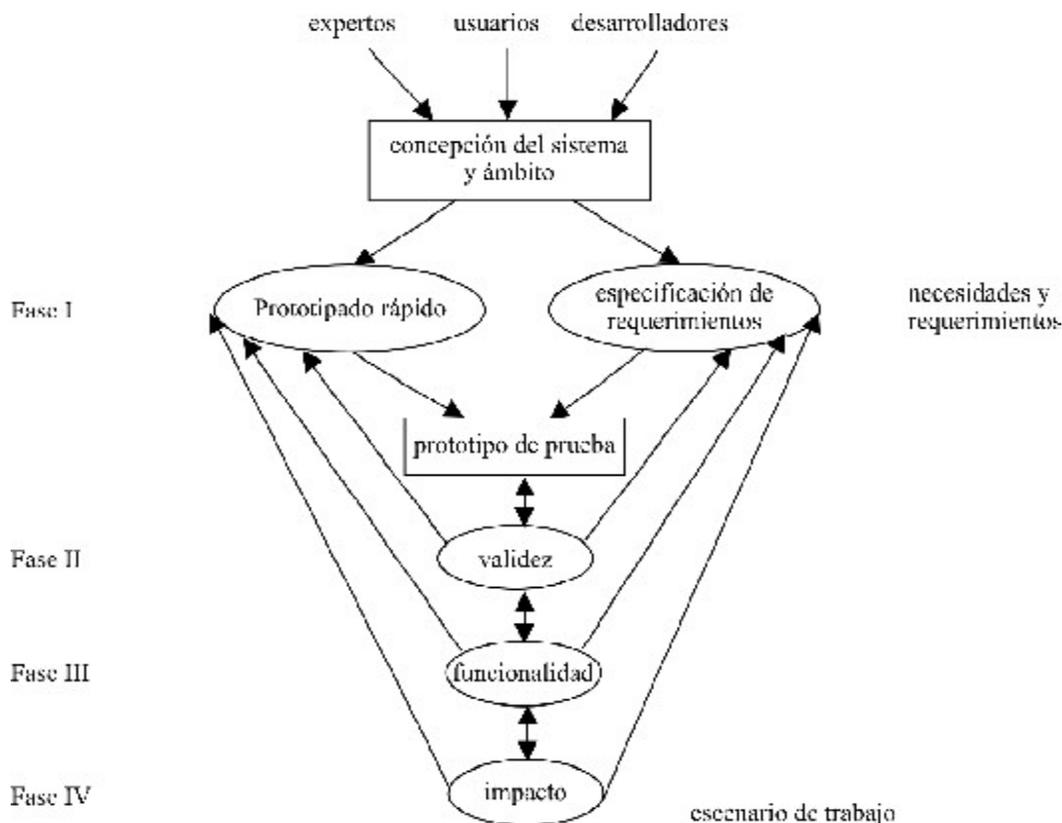


Figura 1: La metodología de evaluación KAVAS.

Cada fase de desarrollo del prototipo tiene que acabar con una sesión de evaluación, durante la cual, todas las partes responsables tienen que revisar los resultados y ponerse de acuerdo sobre los aspectos más importantes antes de pasar a la fase siguiente. Durante cada fase de evaluación, las deficiencias de todas las fases anteriores deben ser examinadas de nuevo. Cada punto débil encontrado durante una fase de la metodología, necesitará una nueva iteración de la fase de desarrollo correspondiente. El proceso de reiteración puede comenzar directamente en la fase de desarrollo anterior o en la fase preliminar de diseño, en la cual los requerimientos de los usuarios han sido previamente especificados. La primera fase es crucial en la planificación de la evaluación debido a que define metas y características:

Las necesidades para un SDS (los requerimientos de los potenciales usuarios en general).

El dominio para el que el SDS proporciona soporte (los límites del dominio deberían definirse de manera no equívoca).

Qué tipo de usuario utilizará el SDS (el tipo de usuario determina cómo deben plantearse las cuestiones).

El papel que se quiere que juegue el SDS (en otras palabras, como va a influenciar a las decisiones en la salida del SDS).

Los *criterios de prestaciones mínimas*, respecto a los cuales los resultados de la evaluación deben ser

comparados, también tienen que ser definidos. Estos criterios corresponden a las razones potenciales de fracaso (p.r.f.). El primer proceso de evaluación es valorar de que manera los requerimientos satisfacen las necesidades del usuario. Esta fase, que corresponde a la concepción del SDS, viene seguida por la especificación de los requerimientos y de los ciclos rápidos de prototipado que llegan al desarrollo del prototipo de prueba del sistema. Hay dos extremos de desarrollo de un primer prototipo:

Un prototipado rápido con ciclos de desarrollo en un ambiente protegido de laboratorio.
La especificación formal de la técnica de desarrollo de los requerimientos.

En general, el desarrollo del prototipo de pruebas se lleva a cabo a través de una combinación de especificación de requerimientos y ciclos rápidos de prototipado y pruebas. Un análisis detallado de la especificación de los requerimientos es más lento que el prototipado rápido pero, siendo más metódica, suele llevar a un prototipo de trabajo satisfactorio en un menor tiempo total. El prototipado rápido es particularmente útil cuando no es posible lograr una definición completa de los requerimientos del usuario sin antes desarrollar un primer prototipo. Este método puede resultar más útil en un dominio donde desarrolladores del SDS y usuarios pueden ser conscientes de las posibilidades futuras del sistema. Muchas veces estos conocimientos pueden ser adquiridos después de pruebas y errores con uno o más prototipos. Algunas de las especificaciones de los requerimientos pueden ser definidas antes, mientras otras hay que definir las durante el ciclo de prototipado rápido. Antes de proceder al proceso de desarrollo a través de los estados de evaluación y desarrollo que llevan al prototipo de pruebas, deben ser elegidos el sistema gestor de base de datos y el mecanismo de inferencia del SDS.

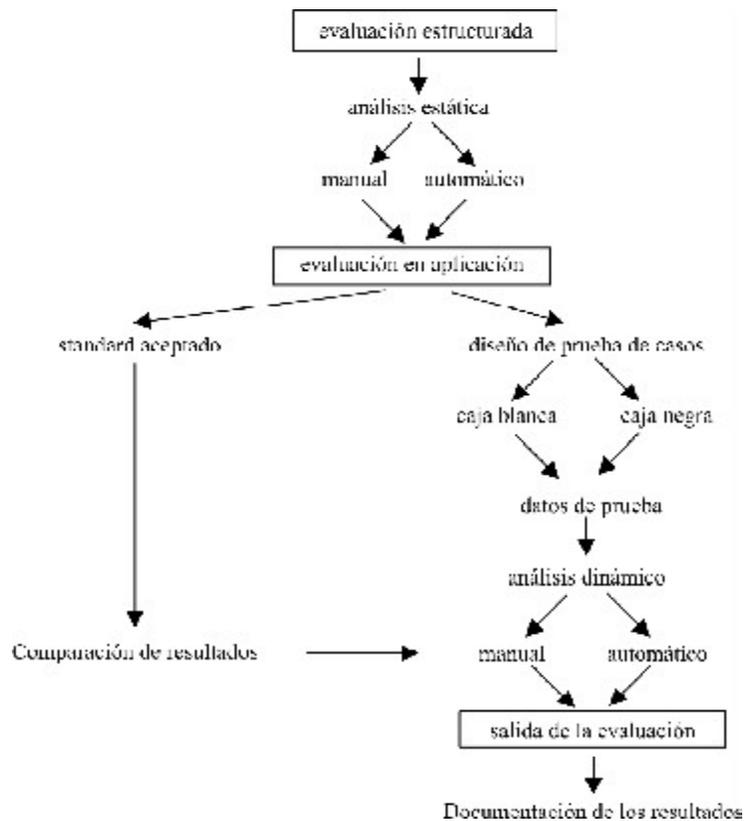


Figura 2. Evaluación de la validez del SDS.

Una vez elegido un prototipo del SDS que responda bien con los ejemplos que se le han presentado, es apropiado avanzar a una evaluación más estructurada en el ambiente del laboratorio. El enfoque puede ser en anticipación o en predicción de las razones de fracaso de cada fase. La segunda fase de la metodología de evaluación implica examinar la validez del SDS. Eso se alcanza con una inspección de la componente base de datos, y verificando que la salida del SDS sea la misma que el estándar de exactitud definido (véase la Figura 2).

Para validar la corrección de la base de conocimientos del SDS, pueden usarse dos clases de métodos de evaluación técnica: *métodos estáticos* y *dinámicos de pruebas*. Los métodos de pruebas estáticos vienen realizados por los ingenieros de los conocimientos, cuya tarea es la verificación directa de los aspectos estructurales de la componente de base de datos. El análisis dinámico de la base de los conocimientos, a través de métodos dinámicos de pruebas, implica la ejecución del SDS y viene realizado después del análisis estático. El análisis dinámico de una base de datos involucra la generación de datos de pruebas que pueden desarrollarse teniendo en cuenta, o no, la estructura interna de un SDS (técnicas de la caja blanca y de la caja negra).

Para poder comparar los resultados, es muy importante definir *la tasa aceptable de error*. Además, se debe determinar cuáles son las *respuestas aceptables*, ya que puede no existir una única respuesta correcta porque un cierto número de respuestas pueden considerarse como aceptables. Cuando ni el SDS y ni los expertos pueden proporcionar respuestas perfectas, la salida del SDS debe tener en consideración la tasa aceptable de error definida. Para estimar la precisión del SDS es importante comparar la salida que se obtiene, como resultado del procesamiento de casos de prueba, con alguna forma de standard comúnmente aceptado (agreed truth o gold standard).

La tercera fase de la metodología de evaluación se refiere a la evaluación de la funcionalidad del prototipo de prueba, e implica la presentación del SDS a usuarios y expertos con el fin de proporcionar una retro-alimentación. La funcionalidad de un SDS no debe ser confundida con su usabilidad o aceptabilidad, que es sólo una parte de lo que se entiende por funcionalidad. La funcionalidad se divide en interna y externa: la funcionalidad interna se refiere a la interacción entre usuario y SDS, mientras la funcionalidad externa tiene que ver con asuntos más generales de transferibilidad y mantenimiento del SDS. La última fase se refiere al impacto de un SDS validado y a las consecuencias a largo plazo debidas a su aplicación en el contexto de trabajo.

3 Adelman y el Enfoque Multi-facetado para la Evaluación

Patrocinadores, usuarios y desarrolladores tienen unos conceptos diferentes de cómo tiene que ser un SDS para que pueda tener éxito: a veces tienen ideas diferentes de cómo evaluar los criterios y hasta pueden diverger en la definición de los mismos. Diferentes criterios de evaluación requieren diferentes métodos de evaluación. El evaluador tiene que conocer cómo implementar los diferentes métodos para proporcionar la información necesaria sobre los diferentes criterios. La labor del evaluador es seleccionar los métodos que son más apropiados, con el objetivo de guiar a los miembros del equipo de D&D y a los patrocinadores en la tarea de evaluar las decisiones inherentes al desarrollo del SDS. Por lo tanto, para una buena evaluación los evaluadores necesitan poder utilizar diferentes métodos para contestar a las distintas preguntas. Esto requiere aplicar un *enfoque multi-facetado* al proceso de evaluación. El método de evaluación multi-facetado está compuesto por tres facetas explícitas: subjetiva, técnica y empírica. Integrando los métodos de evaluación en el proceso de diseño y desarrollo, el equipo de D&D es capaz de realizar correcciones durante el proceso, incrementando así la probabilidad de que un SDS se utilice y, al mismo tiempo, disminuyendo el tiempo y el coste de desarrollo.

La faceta *subjetiva* de la evaluación, permite obtener la opinión de los usuarios sobre los puntos fuertes y débiles del SDS. Ésta se centra en la evaluación del SDS desde la perspectiva de los usuarios potenciales. La meta es determinar si, en general, a los usuarios le gusta el SDS, cuáles consideran que son sus puntos fuertes y débiles relativos, así como cuáles son los cambios que sugieren para mejorar el sistema. Muchas de las sugerencias generalmente mejoran la componente del SDS de interfaz hacia el usuario, especialmente en términos de facilidad de manejo. El evaluador debe conocer cómo utilizar los métodos de evaluación subjetiva para valorar las opiniones de los usuarios. Los tipos de cuestionarios que pueden ser utilizados pueden ser: tipo test, de respuestas cortas, y entrevistas y discusiones de tipo mesa redonda. Estos cuestionarios son importantes porque dan a los usuarios la oportunidad de indicar los aspectos favorables y desfavorables sobre el SDS y sobre como mejorar el sistema. Por otro lado, los cuestionarios tienen la ventaja de ser anónimos.

La faceta de la evaluación técnica mira dentro de la ‘caja-negra’ y se centra en evaluar el SDS desde la perspectiva interna (algorítmica y/o heurística) y externa (entrada / salida). El objetivo es elegir el mejor mecanismo de inferencia para el SDS, evaluar la corrección de su base de datos (es decir, si es consistente de un punto de vista lógico, funcionalmente completa, y correcta en sus predicciones), si los requerimientos han sido satisfechos, y si el SDS puede ser eficientemente integrado con otro software y hardware. La aplicación de los métodos de evaluación técnica debería realizarse lo antes posible dentro del proceso de D&D del prototipo, ya que en estos momentos es cuando los errores son más fáciles y económicos de corregir. El coste total de los esfuerzos de desarrollo es menor si estos errores vienen detectados al principio del proceso de D&D. Los resultados de la evaluación técnica son utilizados con frecuencia para predecir los resultados de la evaluación subjetiva y empírica. La corrección en la construcción del SDS determina la calidad de las tareas realizadas por los usuarios y en consecuencia, lo que estos piensan del sistema. Los evaluadores de esta fase son los desarrolladores, es decir, los ingenieros de software (que son también responsables de recomendar si la inversión y el esfuerzo deberían continuar).

La faceta empírica valida las prestaciones y el impacto del SDS. Su objetivo es obtener mediciones objetivas de los resultados del sistema. Un usuario puede discrepar con las recomendaciones del SDS. Desde una perspectiva iterativa con prototipado, es conveniente que los experimentos se realicen durante todo el desarrollo del prototipo (lo más pronto posible y no cuando el prototipo ha sido desarrollado) para poder medir objetivamente los resultados y obtener una retroalimentación vital, que es necesaria para el desarrollo iterativo del SDS. Tres son los tipos de evaluación empírica que pueden ser utilizados para validar los resultados: experimentos, quasi-experimentos, y estudio de casos. Los dos requerimientos necesarios para utilizar experimentos son la talla de las muestras y la aleatoriedad. En muchos casos estos requerimientos no son posibles y en estas situaciones, el diseño de quasi-experimentos y estudio de casos es el más apropiado. Se ha demostrado que estos métodos empíricos proporcionan una retroalimentación válida para el desarrollo del prototipo. Los experimentos son más viables durante las primeras fases del desarrollo y permiten demostrar desde un punto de vista empírico los beneficios del SDS. Por el contrario, los quasi-experimentos y el estudio de casos resultan más viables después de que el SDS haya sido transferido a su escenario de trabajo. De todas maneras, hay que decir que los experimentos en escenario de trabajo son altamente informativos si los factores de coste permiten su implementación.

4 La Metodología MAKAVAS

Todos los métodos de evaluación tienen sus puntos fuertes y sus puntos débiles, y el uso de más métodos incrementa la probabilidad de que los esfuerzos de la evaluación permitan una correcta validación del SDS. Por ejemplo, la funcionalidad interna de un SDS podría ser medida a través de

métodos de evaluación subjetiva (cuestionarios, entrevistas etc.) o empírica (número de eventos ocurridos). Para que una evaluación pueda ser efectiva, el evaluador debe decidir con antelación qué debe ser validado. Primero tienen que ser identificadas las *medidas de eficacia* (los atributos medibles) diseñados para contestar a las preguntas del evaluador. Estas cuestiones dependen de quien necesita la información (un miembro del equipo de diseño y desarrollo o uno de los patrocinadores), del tipo de la información que se requiere, de la fase del proceso de D&D etc. Las medidas de eficacia resultantes pueden ser variables objetivas o subjetivas dependiendo del método de evaluación elegido.

Los atributos pueden ser cuestiones sobre la interfaz gráfica (como por ejemplo el tiempo de respuesta a una consulta) y otros aspectos del sistema. Las cuestiones pueden ser contestadas subjetivamente (en el caso en que se utilice un método subjetivo) o ‘contestadas’ objetivamente (con métodos técnicos o empíricos). Las cuestiones vienen organizadas en una jerarquía con categorías en los niveles superiores (el mecanismo de razonamiento del sistema, su base de datos, su interfaz gráfica etc.) que son divididas sucesivamente hasta que sea fácil identificar claramente atributos medibles (la facilidad de uso, la habilidad de modificar respuestas etc.). Si se asume que cada nodo terminal de la jerarquía (es decir, cada atributo medible) puede ser medido en términos de ‘medidas de eficacias’, la tarea de evaluar un SDS corresponde a asignar una puntuación a los nodos del nivel inferior de la jerarquía de evaluación. Además, si se da un peso a los *atributos* individuales y a las *categorías de eficacia* más genéricas, subiendo en la jerarquía, se obtiene un *proceso de evaluación explícito* para poder validar cada aspecto del prototipo del sistema de decisión y soporte y así encontrar sus puntos débiles. Es muy importante definir exactamente lo que se quiere evaluar a través de las técnicas de evaluación, por eso se pide estructurar jerárquicamente los atributos medibles (*jerarquía multi-atributos*). El evaluador del sistema tiene que definir los criterios de evaluación junto a los usuarios y a quién subvenciona el desarrollo del prototipo y, sucesivamente, ‘medir’ el prototipo en cada atributo del nivel inferior de la jerarquía. Por último, el evaluador necesita combinar las puntuaciones en un resultado total que mida la *utilidad del sistema*: a todas las medidas de eficacia de la jerarquía (es decir, sus atributos), desde las del nivel inferior a las del superior, se asigna un peso de importancia relativo para poder obtener una valoración en su totalidad de la utilidad del sistema.

El énfasis se pone en la percepción de la utilidad de todo el SDS para el usuario y no solamente en el análisis de los resultados proporcionados. Un SDS podría proporcionar resultados correctos y al mismo tiempo tener una utilidad baja debido a que no ha sido bien adaptado a las características del personal que lo debería usar. El enfoque multi-atributos que hemos analizado está basado en una técnica de *descomposición y reintegración* (divide y vencerás). Desde un punto de vista conceptual el SDS viene descompuesto en criterios de evaluación que permitirán su medición. Esta descomposición se realiza típicamente a través de la creación de una jerarquía de categorías genéricas (como las descritas en el párrafo anterior: el mecanismo de razonamiento del sistema, su base de datos, su interfaz gráfica etc.) y una descomposición posterior hasta un nivel que permita definir y obtener puntuaciones precisas, fiables y válidas de los atributos del sistemas de decisión y soporte (es decir, de características fácilmente medibles). Hay que decir que el enfoque de Adelman recomienda incorporar diferentes tipos de medición para obtener puntuaciones objetivas y subjetivas. Para la reintegración de las puntuaciones que se refieren a los diferentes atributos en un único valor, se utilizan ‘funciones de utilidad’ que se basan en los pesos de importancia relativa asociados a cada atributo. La obtención de un único valor final que pueda, de una cierta manera, expresar la evaluación del SDS, se obtiene a través de un algoritmo en tres pasos. En el caso de métodos subjetivos, por ejemplo, los tres pasos son los siguientes:

- Primero, los atributos del nivel inferior deben ser evaluados por los usuarios y los patrocinadores del SDS (los cuestionarios deben reflejar la intención de descomponer el SDS

en sus partes componentes de manera que se pueda preguntar sobre los atributos de cada una de estas partes).

- Segundo, la puntuación de cada atributo en el caso de métodos de evaluación técnica o empírica, debe ser convertida en un 'valor de utilidad'.
- Finalmente, a la puntuación obtenida por cada atributo se le da el peso correspondiente en la jerarquía, teniendo en cuenta el peso de los niveles superiores.

El algoritmo nos permite combinar los *valores de utilidad* de los atributos del nivel inferior en puntuaciones más globales correspondiente a los niveles superiores de la jerarquía. La salida final del algoritmo de reintegración consta de un único valor, que expresa la evaluación del SDS en función de su utilidad. La suma de los pesos de cada nivel de la jerarquía, debería ser la misma y se recomienda que sea igual a uno. De esta manera, si multiplicamos los pesos a lo largo de cada rama, desde el tope hacia los atributos de la base, se obtiene un peso acumulado en cada atributo que indica la importancia relativa del mismo.

La escala de medición para un atributo depende del atributo mismo y de la manera en que se decida medirlo. Por ejemplo, una escala podría ser en unidades objetivas (minutos etc.) o unidades subjetivas. Generalmente un cuestionario que mide la utilidad de los atributos de un SDS, o *cuestionario de utilidad*, usa la escala Likert, desde 0 puntos (total discrepancia) a 10 puntos (total acuerdo). Es también posible usar una escala de 0 a 100, o con números negativos para representar desacuerdo (el punto 0 representa 'ni acuerdo ni desacuerdo'). La utilidad para un atributo i de la jerarquía de evaluación puede ser representada algebraicamente con la siguiente ecuación [2]:

$$U(i) = w_1 u(x_{i1}) + \dots + w_j u(x_{ij})$$

Donde: hay j sub-atributos; w_j es el peso relativo acumulado del atributo j ; y $u(x_{ij})$ es el valor de utilidad para la alternativa j del atributo i .

La regla de combinación de la ecuación es compensatoria, debido a que una puntuación alta obtenida para ciertos atributos puede compensar bajos valores obtenidos para otros atributos, sin perjudicar la puntuación de la medición total de eficacia del SDS. A veces puede resultar más apropiado usar una regla no-compensatoria para asegurar que el sistema de decisión y soporte obtienen una baja puntuación en la eficacia total si no consigue alcanzar un nivel mínimo de prestación en un atributo crítico del nivel básico de su jerarquía multiatributos. Una regla no-compensatoria puede ser obtenida aritméticamente de la siguiente manera: indicando con un cero si el SDS no supera un umbral de evaluación en uno o más atributos críticos, y con un uno en caso contrario; usando una ecuación multiplicativa y no sumativa para obtener una medida global del resultado de la evaluación, ya que un cero obtenido en un aspecto del SDS que se considere importante (atributo crítico) provoca una puntuación de cero para el SDS. Con respecto a la importancia de los pesos relativos, hay que decir que los pesos asignados a los atributos de la jerarquía de evaluación por los usuarios o los miembros patrocinadores del proyecto, determinan los aspectos del SDS más importantes y ayudan al evaluador a concentrarse en el impacto de posibles errores desde el punto de vista de la utilidad potencial del sistema. No sería eficaz gastar tiempo y recursos mejorando un aspecto del SDS si es de pequeña utilidad para el usuario.

Este enfoque multiatributo no debe considerarse solo una herramienta de evaluación subjetiva para 'medir', a través del uso de un cuestionario, las opiniones de los usuarios con respecto al prototipo. Esto puede ser usado como un marco dentro del cual integrar los tres diferentes criterios de evaluación: subjetivo, técnico y empírico. Los pasos a seguir son:

- Identificar el/los criterio(s) de evaluación que se quiere(n) usar subjetiva u objetivamente.
- Medir las prestaciones del SDS con respecto a los criterios utilizados.
- Desarrollar funciones de utilidad para convertir 'puntuaciones de prestación' a una escala de valores común.
- Determinar la importancia, en términos de pesos relativos, para cada criterio utilizado.
- Finalmente, combinar las 'puntuaciones de utilidad' para cada criterio de evaluación utilizado, teniendo en cuenta los pesos relativos correspondientes. Con esto se obtiene una única puntuación que indica la utilidad global del sistema evaluado.

A lo largo del proceso de diseño y desarrollo del prototipo, y con el fin de evaluar el progreso durante las distintas fases, pueden usarse diferentes métodos de evaluación. Con respecto a la integración del enfoque multifacetas en la metodología de evaluación KAVAS, consideramos que la evaluación subjetiva es apropiada durante todo el proceso de diseño y desarrollo: al principio para poder definir los requisitos del SDS y al final de las fases segunda y tercera para, por ejemplo, obtener las opiniones de los usuarios potenciales y patrocinadores sobre aspectos del prototipo de prueba que se ha desarrollado. Los métodos de evaluación técnica son apropiados para verificar si los requerimientos del SDS se están satisfaciendo (final fase I), comprobar la funcionalidad del prototipo (fase III), y sobre todo seleccionar el mecanismo de razonamiento del SDS y validar la base de conocimientos (fase II). Los métodos de evaluación empírica se aplican al final de la tercera fase y tras la cuarta, ya que validan las prestaciones del SDS, así como su impacto.

5 Test de la Evaluación de un Sistema de Decisión y Soporte para Logopedas

Para probar la metodología MAKAVAS se ha diseñado, desarrollado y evaluado un sistema de ayuda a los logopedas en el análisis de las disfunciones del habla en niños y jóvenes con síndrome de Down. El sistema de decisión y soporte se ilustra en la Figura 3.

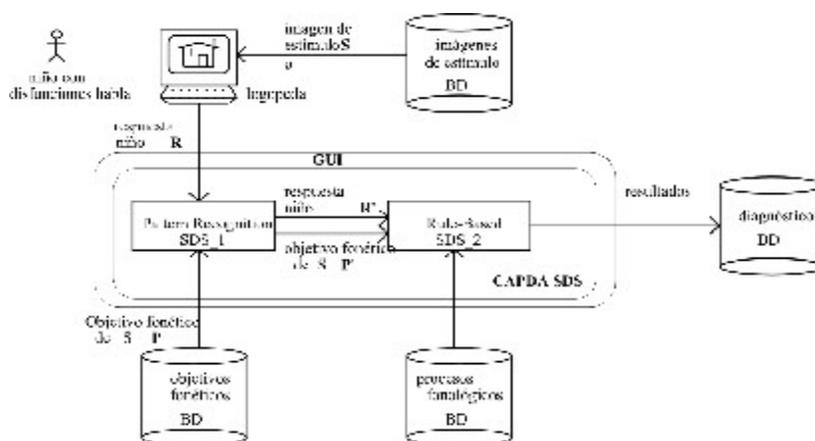


Figura 3.

Sistema de decisión y soporte CAPDA (Computer Assisted Phonological Data System) [4]. El logopeda muestra las imágenes de estímulo al niño e introduce las respuestas. Al final de la sesión de terapia los resultados se memorizan en la base de datos.

El prototipo ha sido instalado en la Escuela Clinical Speech & Language Studies del Trinity College Dublin (Irlanda). Con respecto a la aceptabilidad de las interpretaciones de las disfunciones fonológicas proporcionadas por el SDS, el diagnóstico ha sido clasificado, de acuerdo con los logopedas del centro irlandés, en cuatro categorías:

- *Apropiado*: el logopeda está de acuerdo con la salida del diagnóstico fonológico del SDS.
- *Aceptable*: el logopeda no está totalmente de acuerdo con el diagnóstico fonológico proporcionado por el SDS y su interpretación de los resultados diverge ligeramente.
- *Inapropiado*: el experto discrepa con el diagnóstico del SDS.
- *No procesado*: el prototipo del SDS no presenta ninguna interpretación; interpretaciones no resueltas indican áreas incompletas del dominio del conocimiento fonológico (respuestas ininteligibles de niños).

El prototipo ha sido validado inicialmente con 121 casos de análisis fonológico procesados previamente por logopedas. Para minimizar tipos de clasificación no deseados y efectos colaterales, como la reducción de la corrección del SDS, se han estudiados los diagnósticos obtenidos. Los resultados se muestran en la Tabla 1.

<i>clasificación</i>	<i># casos</i>	<i>% casos</i>	<i># casos</i>	<i>% casos</i>
apropiados	72	59.5	150	81.5
aceptables	25	20.6	23	12.5
inapropiados	10	8.3	0	0
no procesados	14	11.6	11	6

Tabla 1. Resultados obtenidos antes (primeras dos columnas) y después de la extensión y refinamiento del prototipo.

En general, la clasificación de los resultados obtenidos depende del nivel de conocimientos del evaluador y del papel desempeñado por el prototipo. En nuestro caso, los casos no procesados han sido considerados como tolerables ya que el SDS ha sido diseñado y desarrollado con el objetivo de ayudar a procesar los datos más rápidamente y proporcionar un posible diagnóstico que el logopeda debe complementar en última instancia. Después de un cierto número de refinamientos y extensiones del número de reglas fonológicas de la base del conocimiento, se obtuvo una tasa de error aceptable. Los resultados de los 184 casi-experimentos de datos fonológicos presentados al prototipo durante las sesiones de evaluación empírica, están representados en la Tabla 1. Se puede apreciar que la tasa de los diagnósticos apropiados se ha incrementado un 20%. Si se incluyen también aquellos casos que se consideran aceptables, podemos decir que el prototipo obtuvo una interpretación acertada de las disfunciones fonológicas en el 94% de los casos. Los diagnósticos aceptables pueden ser considerados como acertados debido al hecho de que, con frecuencia, en la logopedia puede no existir una ‘mejor’ interpretación, y puede haber diferentes análisis de los resultados. Por lo tanto, hay casos para los cuales puede no existir una única interpretación correcta y debe acordarse previamente cuáles son los diagnósticos que pueden ser considerados aceptables. Debemos decir que 184 casos de pruebas (aunque diseñadas ad hoc para validar todas las reglas fonológicas), no son suficientes para proporcionar una evaluación completa del prototipo, ya que debería ser evaluado diseñando un número mayor de pruebas para asegurar el diagnóstico de todas las posibles disfunciones fonológicas. De todas formas, el resultado de esta primera evaluación ha indicado que el SDS ha operado a un nivel más que satisfactorio de corrección.

6 Conclusiones

En este artículo hemos descrito una metodología iterativa con prototipado que permite la evaluación durante las diferentes fases del diseño y desarrollo de un sistema de decisión y soporte. A través de la integración de los dos enfoques en la evaluación, el enfoque iterativo con prototipado de la metodología KAVAS y el enfoque ‘multi-facetas’ de Adelman, la metodología MAKAVAS permite realizar una evaluación exhaustiva de un SDS. Esta metodología formal proporciona un formato estructurado en el cual puede tener lugar el desarrollo, satisfaciendo los intereses de todas las partes involucradas, y proporcionando un criterio de medida a utilizar en la evaluación del SDS, desde su concepción (es decir, desde su diseño) hasta su desarrollo y utilización. Con la integración de diferentes métodos para la evaluación, durante el proceso de diseño y desarrollo, la probabilidad de una implementación con éxito del SDS aumenta: el equipo de diseño y desarrollo es capaz de realizar correcciones durante el proceso mismo, lo que incrementa la probabilidad de que el SDS sea de utilidad, y decremente tiempo y costes de desarrollo. Los resultados de una metodología exhaustiva dan medidas de éxito y fracaso de un SDS junto con las causas que los provocan. El uso de una metodología formal para la evaluación debería facilitar el diseño y desarrollo de un SDS y, consecuentemente, su aceptabilidad.

Agradecimientos

Deseo expresar mi reconocimiento a Prof. Jane Grimson y a todo el grupo de Knowledge Data Engineering del departamento de informática de la universidad irlandesa de Trinity College Dublin, donde mi trabajo de investigación tuvo lugar, así como a Emilio y José Luis para ayudarme con mi ‘itañol’...

Referencias

- [1] K. Clarke. Evaluation and Refinement of Medical Knowledge Base Systems. Ph.D. Thesis, Trinity College, University of Dublin, Irlanda, Octubre 1995.
- [2] L. Adelman. Evaluating Decision Support and Expert Systems. Systems Engineering Series, 1992.
- [3] A. Rector et al. Validation and Verification Requirements and Operational Criteria. Technical Report. AIM / Eurodiabeta, 1990.
- [4] P. Rosso. Decision Support in Speech and Language Therapy: Managing Phonologically Disordered Speech in Childhood. Ph.D. Thesis, Trinity College, University of Dublin, Ireland, Mayo 1999.
- [5] F. R. Hickman et al. Analysis for Knowledge-Based Systems: A Practical Guide to the KADS Methodology. Ellis Horwood, 1989.
- [6] P. Rosso P. y J. Grimson . The MAKAVAS Evaluation Methodology, In Proceedings VIII Conferencia de la Asociación Española para la Inteligencia Artificial (AEPIA), Murcia, volumen 1, páginas 35-43, 1999.
- [7] P. Rosso et al. Managing Intelligibility in Children or Young Adults with Down’s Syndrome or Speech Disorders. In Proceedings EUROSPEECH ’95, Madrid, volumen 3, páginas 1875-1878, 1995.