

Silvia Tamayo Moreno
Diana Pérez Marín

**Evaluación del Agente
Conversacional Pedagógico
Dr. Roland para el nivel de
Educación Secundaria mediante el
método de medida de las
prestaciones**

Número 2017-12

Serie de Informes Técnicos DLSI1-URJC
ISSN 1988-8074
Grupo Docente de Lenguajes y Sistemas Informáticos I
Universidad Rey Juan Carlos

Índice

1	Introducción.....	5
1.1	Medida de las prestaciones, método de evaluación de la Interacción Persona Ordenador.....	5
2	Evaluación del agente Dr. Roland para Educación Secundaria mediante el método de medida de las prestaciones	7
3	Conclusiones.....	12
	Referencias	12

Evaluación del Agente Conversacional Pedagógico Dr. Roland para el nivel de Educación Secundaria mediante el método de medida de las prestaciones

Silvia Tamayo Moreno y Diana Pérez Marín

Dto. de CC Comp., Arq. Comp., LSI y EIO, Universidad Rey Juan Carlos,
C/ Tulipán s/n, 28933, Móstoles, Madrid
silviatamayomoreno@gmail.com, diana.perez@urjc.es

Abstract. In this paper the evaluation of the pedagogical conversational agent Dr. Roland is carried out for the level of Secondary Education following the performance measurement method. The logging method is one of the test methods of the techniques of evaluation of the Human Computer Interaction.

Keywords: Medida de las prestaciones, Agente Dr. Roland, Educación Secundaria, Interacción persona Ordenador.

1 Introducción

En primer lugar se va a hacer alusión a la interacción persona ordenador (IPO). Seguidamente a técnicas de evaluación de la IPO [1], centrandó la atención en los métodos de test. Concretamente en el método de medida de las prestaciones.

A continuación se realizará la evaluación del Agente Conversacional Pedagógico Dr. Roland [11, 12] para el nivel de Educación Secundaria usando el método de test de medida de las prestaciones.

1.1 Medida de las prestaciones, método de evaluación de la Interacción Persona Ordenador

Los sistemas interactivos son dispositivos hardware y software que actúan de interconexión entre las personas, permitiéndoles interactuar, y favoreciendo la realización de tareas y el alcance de los objetivos propuestos.

Con el desarrollo de los sistemas interactivos, y para tener una mejor comprensión de las relaciones persona-ordenador y mejorarla, surge la interacción persona ordenador (IPO). La IPO, conocida en la comunidad internacional como Human-Computer Interaction (HCI) o Computer-Human Interaction (CHI), según el grupo SIGCHI (Special Interest Group on Computer-Human Interaction, [2]) de ACM: “*Es la disciplina relacionada con el diseño, evaluación e implementación de sistemas informáticos interactivos para el uso de seres humanos, y con el estudio de los fenómenos más importantes con los que está relacionado*” [1].

Constantemente se presenta la situación en la que se diseñan productos que no se prueban con quienes van a ser los usuarios, sin prestar atención por tanto a la usabilidad de los productos y el software, como a las pruebas con usuarios. Los motivos suelen ser los costes y cambios que pueda suponer dicha evaluación. Se centran los esfuerzos y atención en la funcionalidad, y se tiene el pensamiento de que si quienes diseñan y programan, pueden usar el sistema y les gusta, es suficiente [3].

Evaluar algo implica hacer pruebas de funcionamiento y comprobar si cumple las expectativas, necesidades y requisitos. En el ámbito de los sistemas interactivos es clave y es importante que se realice durante todo el proceso de desarrollo. Los resultados de ello, obtenidos mediante la aplicación de una serie de técnicas, se pueden emplear para mejorar los sistemas. Lorés et al. [4] define la evaluación como “*La actividad que comprende un conjunto de metodologías y técnicas que analizan la usabilidad y/o la accesibilidad de un sistema interactivo en diferentes etapas del ciclo de vida del software*”.

Los prototipos se hacen para comprobar si el sistema reúne características que lo hagan usable para los usuarios finales. La evaluación del diseño conlleva comprobar si se adapta a lo esperado, permitiendo realizar las tareas, y comprobar el impacto en el usuario en aspectos como la facilidad de aprendizaje, manejabilidad, identificación de áreas que necesiten que sea recordada gran cantidad de información, etc. Con ello se identifican posibles problemas específicos del diseño, como todo aquello que cause resultados no esperados o confusos al usuario estando en su contexto de uso [5]. Según Dix [6] los objetivos de la evaluación son la comprobación de la funcionalidad del sistema, del efecto de la interfaz y la identificación de problemas específicos del sistema.

Los sistemas interactivos pueden ser evaluados en cualquier contexto que lo permita, donde se reúnan usuarios y evaluadores, desde espacios equipados como laboratorios, salas de reuniones o hasta en el entorno en el que los usuarios desarrollan sus tareas.

El abanico de métodos de para evaluar los diferentes aspectos es muy amplio, existiendo solapamiento entre algunos de ellos en cuanto a las actividades que desarrollan, no son independientes [7]. La elección de unos u otros depende de los costes y lo que se obtendrá con su uso.

En cuanto al tipo de técnica de comprobación utilizada, se distinguen:

- **Métodos de inspección:** Agrupa una serie de métodos que cuentan con expertos, llamados evaluadores, que explican el grado de usabilidad de un sistema. Se basan en la inspección de aspectos de la interfaz del sistema relacionados con la usabilidad ofrecida a los usuarios. Se considera opiniones, juicios o informes de los inspectores sobre elementos específicos de la interfaz como factor fundamental de la evaluación [8]. Permiten identificar, clasificar y contabilizar un gran número de errores potenciales de usabilidad a precio relativamente bajo, contribuyendo a ello el hecho de no usar usuarios [8, 9].

- **Métodos de indagación.** Consiste en alcanzar el conocimiento de una cosa mediante conjeturas y señales. Se llevan a cabo, hablando con los usuarios, observándolos, usando el sistema en tiempo real y observando las respuestas a preguntas, ya sea orales o escritas. La información sobre gustos del usuario, necesidades o identificación de requisitos es importante, especialmente en fases tempranas del proceso de desarrollo.
- **Métodos de test:** En ellos usuarios, normalmente representativos, usan el sistema o prototipo para tareas concretas. Los evaluadores usan los resultados para saber la forma en la que la interfaz de usuario da soporte a los usuarios con sus tareas.

Como parte de los métodos de test se encuentra el de medida de las prestaciones

Medida de las prestaciones

Este método se basa en tomar medidas sobre el rendimiento y otros tipos de aspectos subjetivos que afecten la usabilidad del sistema. Es importante la comprensión de lo que se pretende medir, se pueden seleccionar:

- Medidas de rendimiento, son medidas cualitativas, se puede contar las acciones, los comportamientos, número de personas, de errores, etc.
- Medidas subjetivas, es decir, percepciones de las personas, opiniones y juicios. Pueden ser cuantitativas o cualitativas.


Por ello, se necesita disponer de un sistema implementado o un prototipo que permita la evaluación de dichos aspectos. Los participantes del test son usuarios realizando tareas, se analiza la forma en la que usan el producto y se mide el tiempo que emplean. Es útil para ser realizado en laboratorios de usabilidad, aunque también puede hacerse en un entorno real. Es importante la selección de tareas que los usuarios tendrán que llevar a cabo, estas deberán demostrar problemas de usabilidad, siendo el criterio más importante para la selección que prueben problemas potenciales de usabilidad del producto. Deberán ser tareas sugeridas por la propia experiencia, ya que los desarrolladores tienen ideas sobre dónde encontrar problemas, qué partes del producto fueron más difíciles de diseñar y cuáles son los problemas que se tienen que probar [10].

2 Evaluación del agente Dr. Roland para Educación Secundaria mediante el método de medida de las prestaciones

Se ha llevado a cabo la técnica de test medida de las prestaciones [10] para el agente Dr. Roland para Educación Secundaria. El agente para dicho nivel puede verse, junto con un ejemplo de interacción entre el agente y el usuario en las figuras 1, 2 y 3.

Soy Dr. Roland, bienvenido
Colegio San Diego y San Vicente de Paúl

Ayuda



Dos amigos, Enrique y Juan tienen una baraja española de 40 naipes. En un momento dado, Enrique apuesta que es capaz de sacar el as de oros sin mirar. Juan acepta la apuesta y la dobla, diciendo que si Enrique saca el as de oros, en las cartas restantes, sacará el as de espadas. ¿Quién tiene más posibilidad de ganar?

Fig. 1. Ejemplo en el que el estudiante acaba comprendiendo el ejercicio pero no lo sabe resolver (1)

¿Sabes cual es el resultado?
no ▾

¿Entiendes qué es lo que te piden?
no ▾

¿Necesitas ayuda?
sí ▾

¿Qué frase crees que es la principal y que nos daría esta información?
¿Quién tiene más posibilidad de ganar?
.....

La respuesta es correcta
¿Entiendes qué es lo que te piden?
no ▾

¿Cuál o cuáles crees que son los verbos principales de la frase?
tiene, ganar
.....

La respuesta es correcta
¿Entiendes qué es lo que te piden?
no ▾

¿Crees que hay algún o algunos verbos que te dicen de forma más concreta cómo tienes que dar el resultado? Escríbelo/s. Si crees que no hay, no escribas nada
tiene, ganar
.....

La respuesta es correcta
¿Entiendes qué es lo que te piden?
no ▾

Fig. 2. Ejemplo en el que el estudiante acaba comprendiendo el ejercicio pero no lo sabe resolver (2)

¿Cuál o cuáles crees que son las palabras más importantes del ejercicio?
 más probabilidad

La respuesta es correcta
 ¿Entiendes qué es lo que te piden?
 no

¿Crees que puede haber algun pronombre importante? Introdúcelos, si crees que no hay, no introduzcas nada

La respuesta no es correcta ¿Quieres ver o volver a introducir el o los pronombres?
 ver

La respuesta correcta es: quién,quién
 ¿Entiendes qué es lo que te piden?
 si

¿Sabes resolver el ejercicio?
 no

En cualquier caso, para tener la información, escribe qué es lo que no entiendes
 entiendo que tengo que calcular quién tiene más probabilidad de ganar pero no sé resolverlo porque no se calcular probabilidades

Habla con tu profesor sobre este ejercicio
 ¿Desea realizar otro ejercicio?

Fig. 3. Ejemplo en el que el estudiante acaba comprendiendo el ejercicio pero no lo sabe resolver (3)

En este apartado se va a hacer alusión a la medida de las prestaciones, obtenidas tras el uso del agente Dr. Roland en las aulas.

La Tabla 1 recoge una serie de medidas para la experiencia de uso de Dr. Roland en Secundaria. Recoge números de ejercicios realizados, respondidos correctamente, tiempos de uso, tiempos medios de ejercicios y una serie de porcentajes. El dispositivo de uso han sido ordenadores. El tiempo medio de uso por sesión es de 35 minutos. El número de ejercicios respondidos correctamente por experiencia (colegio) es 72, por experiencia y por sesión es 36, el porcentaje de ejercicios respondidos correctamente sobre el total de ejercicios hechos por experiencia es del 31,72%, mientras que por sesión coincide. El tiempo medio de uso por ejercicio por experiencia (minutos) es de 3,24.

Tabla 1. Ev. - Medidas de las prestaciones en Secundaria: medidas por experiencia

Medidas	Experiencia Sec-1 (2 sesiones)
Número ejercicios totales hechos en cada experiencia	227 totales entre 19 niños
Número ejercicios medios hechos en cada experiencia por sesión	113,5 entre 19 niños
Tiempo total de uso por experiencia (minutos)	70 minutos cada niño
Tiempo medio de uso por experiencia por sesión (minutos)	35
Dispositivo usado	Ordenadores
Número Ejercicios respondidos correctamente por experiencia	72
Número de ejercicios respondidos correctamente por experiencia por sesión	36
Porcentaje de ejercicios respondidos correctamente sobre el total hechos por experiencia	31,72%
Porcentaje de ejercicios respondidos correctamente sobre el total hechos por experiencia por sesión	31,72%
Tiempo medio de uso por ejercicio por experiencia(minutos)	3,24
Tiempo medio de uso por ejercicio por experiencia por sesión(minutos)	3,24

Se recoge una serie de medidas por sesiones de cada una de las experiencias en la Tabla 2, entendiendo como experiencia aquello que se mencionó previamente, y como sesión cada uno de los días en los que se ha puesto en práctica el agente. Una experiencia (Sec-1) y dos sesiones en este caso. Recoge los ejercicios realizados, tiempos de uso, dificultades en la interacción, porcentajes de ejercicios resueltos de voz o por teclado, uso de reintentos en los ejercicios, uso de la ayuda, ejercicios respondidos correctamente o entendimiento de los ejercicios.

Añadir que se ha considerado que un ejercicio se ha resuelto correctamente, cuando así ha sido en una misma realización del ejercicio, independientemente del número de reintentos en la misma. A diferencia de lo que sucede para Infantil y Primaria y que se comentará en sus apartados correspondientes de medidas de las prestaciones.

El tiempo total de uso por sesión fue de 10 minutos la primera y 60 minutos la segunda, todos los ejercicios se realizaron por teclado. El porcentaje de ejercicios en los que no se entendía lo que preguntaba fue del 28% y del 13,56% en cada una de las sesiones, porcentaje que coincide con los ejercicios en los que se ha hecho uso de la ayuda. Lo que indica que los niños han hecho uso de la misma cuando no entendía qué se les pedía. El porcentaje de ejercicios respondidos correctamente en la primera sesión ha sido de 23,95% y en la segunda de 33,94%.

Tabla 2. Medidas de las prestaciones en Secundaria: medidas por experiencia

Medidas	Experiencia Sec-1: Sesión 1	Experiencia Sec-1: Sesión 2
Número ejercicios hechos en cada sesión	50 ejercicios totales 19 niños 2,63	177 ejercicios totales 19 niños 9,31
Tiempo total de uso por sesión	10	60
Número Ejercicios respondidos correctamente	12 ejercicios totales 19 niños 0,63	60 ejercicios totales 19 niños 3,16
Porcentaje de ejercicios respondidos correctamente sobre el total hechos	23,95%	33,94%
Tiempo medio de uso por ejercicio por sesión (minutos)	3,8	6,44
Porcentaje de ejercicios en los que se ha hecho uso de la ayuda del ejercicio	En 14 ejercicios se ha hecho uso de la ayuda, sobre el total de 50 (28%)	En 24 ejercicios sobre un total de 177 (13,56%)
Porcentaje de ejercicios en los que se ha hecho uso de reintento en la respuesta	En 27 de 39 que ha intentado resolver de un total de 50 que supone un: 69,23% de los que intentó resolver y un 54% del total	En 57 de 117 que intentaron resolver de un total de 177 que supone un 48,72% de los que intentó resolver y un 32,2% del total
Dificultad en la interacción	Al inicio al introducir las respuestas	No
Porcentaje de ejercicios en los que no entendían lo que se preguntaba	En 14 ejercicios sobre el total de 50 (28%)	En 24 ejercicios sobre un total de 177 (13,56%)
Porcentaje de ejercicios resueltos de voz	No aplica	No aplica
Porcentaje de ejercicios resueltos por teclado	100%	100%

3 Conclusiones

El método de test de medidas de las prestaciones de la Interacción Persona Ordenador fue usado para la evaluación de Agente Dr. Roland para el nivel de Educación Secundaria. Tras la evaluación puede concluirse, que el agente fue integrado en el aula y usado por los alumnos. De ello puede extraerse, que los estudiantes fueron capaces de interactuar con el agente, siendo el algoritmo RUANLP el que guía la interacción y el diálogo entre estudiantes y el agente.

Referencias

1. Hewett, T., Baecker, R., Card, S., Carey, T., Gasen, J., Mantei, M., Perlman, G., Strong, G. & Verplank, W. (1992). ACM SIGCHI Curricula for Human-Computer Interaction. Technical Report. ACM, New York, NY, USA.
2. SIGCHI Special Interest Group in Computer Human Interaction. Última fecha de consulta: 20-04-2017.
<http://www.sigchi.org/>
3. Granollers i Saltiveri, T., Lorés Vidal, J., & Cañas Delgado, J. J. (2012). Diseño de sistemas interactivos centrados en el usuario.
4. Lorés, J., Granollers, T. y Lana, S. (2002). Introducción a la interacción persona-ordenador. En J. Lorés (ed.) La interacción persona-ordenador. Lérida, 2001-2002, 20-40.
5. Saltiveri, G. (2007). MPIu+ a. Una metodología que integra la ingeniería del software, la interacción persona-ordenador y la accesibilidad en el contexto de equipos de desarrollo multidisciplinares. Universitat de Lleida.
6. Dix, A. (1993). Human computer interaction. Prentice Hall, Englewood Cliffs, NJ.
7. Rodeiro, J. (2001). Representación y Análisis de la componente visual de la interfaz de usuario. Tesis doctoral. Universidad de Vigo.
8. Nielsen (1994), "Heuristic evaluation", Usability Inspection Methods, John Wiley & Sons, New York, NY.
9. Bias, R. & Mayhew D. (1991). Cost-justifying usability. IEEE Software.
10. Hassan Montero, Y. & Ortega Santamaría, S. (2009). "Informe APEI sobre usabilidad".
11. Tamayo-Moreno (2012). Propuesta de Desarrollo Centrado en el Usuario de un Agente Conversacional Pedagógico para la Comprensión Lectora de Ejercicios de Matemáticas a nivel escolar. Máster Universitario en Investigación en Sistemas Hardware y Software Avanzados de la Escuela Técnica Superior de Ingeniería Informática de la Universidad Rey Juan Carlos.
12. Tamayo-Moreno, S & Pérez-Marín, D. (2014). Diseño y gestión de diálogos educativos en agentes pedagógicos conversacionales para uso escolar. Aplicaciones multidisciplinares de sistemas de diálogo. López-Cózar Delgado, Ramón (Ed.), Coca.