

# Usando IRT y Agentes para Educación Distancia Adaptativa

**Gerardo Rossel**

grossel@computer.org

CAETI – Facultad de Tecnología Informática

Universidad Abierta Interamericana

Departamento de Computación – FCEyN

Universidad de Buenos Aires

## Resumen

En este trabajo presentamos una arquitectura que facilita la educación a distancia adaptativa. Para lograrlo se utilizan tests adaptativos para autoevaluación utilizando la *teoría de respuesta a ítems*. Los agentes inteligentes utilizarán los resultados de los tests para adaptar el contenido y material de acuerdo al alumno. De esta forma, logramos una arquitectura que permite integrar la tecnología de agentes para facilitar la tarea del tutor en la enseñanza a distancia.

**Palabras claves:** IRT, Agentes Inteligentes, Educación a Distancia, Hipermedia Adaptativa

## Introducción

La enseñanza a distancia ha tenido un enorme empuje a partir del desarrollo de las nuevas tecnologías de la información y la comunicación. Aunque en sus orígenes se rastrean a la enseñanza por correspondencia, con la rigidez y poca interacción que eso implica, esto ha cambiado rotundamente. Podemos según [28] distinguir cinco generaciones en la educación a distancia:

- Primera generación: por correspondencia
- Segunda generación: enseñanza multimedia (video, sonido, texto).
- Tercera generación: tele-aprendizaje
- Cuarta generación: aprendizaje flexible
- Quinta generación: aprendizaje flexible inteligente.

La mayor parte de la enseñanza a distancia se encuentra hoy día en la tercera generación. Si bien existen ambientes de cuarta y quinta generación, no es lo predominante hoy en día. Los ambientes de enseñanza a distancia actuales (ya sea de tercera, cuarta o quinta generación) están basados en aplicaciones Web que permiten la interacción entre estudiantes y docentes. Los sistemas educativos basados en Web permiten poner, además, contenido interactivo multimedia disponible fácilmente y crean nuevos canales de comunicación. A diferencia de otros sistemas de enseñanza asistida por computadora como los de segunda generación, los sistemas basados en Web logran la independencia del aula y la plataforma siendo actualmente la plataforma preferencial para la enseñanza a distancia. Sin embargo, la falta de contacto personal puede complicar un seguimiento personalizado por parte del tutor que permita adaptar contenidos y formas según la respuesta de los alumnos. Ese tipo de adaptación no está presente en la mayoría de los sistemas Web. Para lograrlo se necesita de sistemas de enseñanza adaptativos para Web (SEAW) [23] [24]. Estos sistemas, que tienen su origen en los sistemas hipermediales adaptativos y en los sistemas tutores inteligentes, permiten mejorar a los sistemas clásicos con el acceso personalizado a la información. La adaptabilidad propia de los sistemas hipermedia, es decir la posibilidad de navegación libre no es

suficiente ya que el alumno navega sin guía por la información presentada. La adaptación de los SEAW presentan dos facetas de adaptabilidad: la navegación y el contenido.

Presentamos la utilización de la teoría de respuesta a ítems (ITR *Item Theory Response*) [10] para obtener información acerca del alumno y de esa manera poder adaptar, mediante el uso de agentes inteligentes, la información a sus necesidades y conocimientos.

Este trabajo se organiza de la siguiente manera: primero describimos brevemente los fundamentos de la ITR y de los sistemas multiagentes. Luego mostramos como la *teoría de respuesta a ítems* puede ser utilizada en entornos de enseñanza a distancia para la autoevaluación e integrar los resultados de los tests con los agentes planificadores de contenido. Finalizamos con algunas conclusiones y descripción de trabajos en curso.

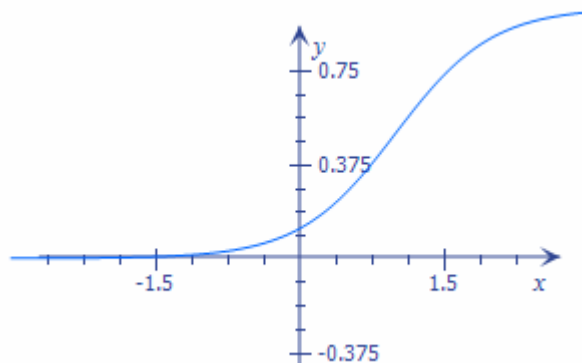
## Teoría de Respuesta a Ítems.

Los exámenes tradicionales como los realizados en papel y lápiz no cuentan con la flexibilidad que proveen los test adaptativos basados en software. Si los ítems de los exámenes son fijos, no pueden evaluarse fácilmente las diversas habilidades de los alumnos. Para algunos alumnos, los ítems del test pueden ser muy fáciles y para otros dificultosos. Por ello, este tipo de test provee poca información acerca de las habilidades particulares de los alumnos.

Los test adaptativos por computadora permiten ir evaluando interactivamente la habilidad del alumno para asignar los próximos ítems según el nivel de habilidad evaluado hasta el momento. De esta forma, y dentro de ciertos límites concernientes por ejemplo a la necesidad de acreditación, los alumnos no tendrán ítems muy fáciles o muy difíciles para ellos. Algo similar a lo que realiza un profesor de manera informal durante un examen oral.

La teoría subyacente a la evaluación interactiva de los alumnos durante el test es la teoría de respuesta a ítems [11] [10]. Puede decirse que es uno de los principales logros de la medición psicológica en el siglo pasado [27]. La IRT es un framework matemático que permite describir a los alumnos que están siendo examinados mediante un conjunto de calificaciones (*scores*) predictivas de habilidad a través de modelos matemáticos que involucran estadísticas sobre los ítems, habilidades del alumno y desempeño sobre los ítems del test tomado [12][13].

Se asume que cada alumno que se encuentra realizando un test posee una habilidad subyacente. A dicha habilidad puede asignársele un valor numérico. Este valor numérico o *score* posiciona al alumno en una escala de habilidad. Notamos con la letra theta,  $\theta$  a dicho valor numérico. Para un ítem dado hay una probabilidad  $P(\theta)$  de que un alumno con el nivel de habilidad  $\theta$  responda correctamente. La probabilidad de responder correctamente es menor para alumnos con un nivel de habilidad más bajo y es mayor para alumnos con niveles de habilidad altos. La siguiente curva describe  $P(\theta)$  como función de la habilidad del alumno [11].



La curva anterior es lo que se denomina curva característica del ítem y es parte esencial de la teoría de respuesta a ítems. Estas curvas están dentro de la familia de las curvas acumulativas de la función logística [11].

Las curvas tienen dos propiedades técnicas fundamentales: la dificultad del ítem y la discriminación. La dificultad es un índice de ubicación del ítem, mientras que la discriminación determina que tan bien el ítem discrimina entre los que están en un nivel inferior de habilidad (según la dificultad del ítem) y en un nivel mayor. La discriminación está determinada por lo plano que sea la curva, mientras más plana menos discrimina. La dificultad de un ítem puede ser medida por la probabilidad  $P(\theta) = 0.5$ . El valor del punto con dicha probabilidad indica la dificultad del ítem.

De lo anterior, se deduce que el aspecto central es el modelo matemático que describa la curva. Este modelo es el que le da rigurosidad al concepto intuitivo. Se han desarrollado diferentes modelos como el modelo logístico de dos parámetros, el modelo de Rasch y el modelo logístico de tres parámetros. A continuación, mostraremos las ecuaciones de los dos primeros y describiremos más detalladamente el tercero que es el que utilizamos.

### Modelo logístico de dos parámetros y modelo de Rasch

La siguiente ecuación es la correspondiente al modelo logístico de dos parámetros:

$$P(\theta) = \frac{1}{1 + e^{-a(\theta-b)}}$$

Donde  $a$  es el parámetro de discriminación y  $b$  es el parámetro de dificultad y naturalmente  $\theta$  es el nivel de habilidad. El parámetro  $b$  es el punto de la escala de habilidad al cual la probabilidad de respuesta correcta del ítem es  $0.5$ . En general se utiliza un rango entre  $-3$  y  $3$  para los valores de  $b$ . El parámetro  $a$  se considera la pendiente en  $b$  (en realidad sería la pendiente por  $4$ ).

El modelo de Rasch es muchas veces llamado modelo logístico de un parámetro debido a que el valor del parámetro de discriminación es fijo con valor de  $1$ . La ecuación es:

$$P(\theta) = \frac{1}{1 + e^{-(\theta-b)}}$$

### Modelo de tres parámetros

El modelo de tres parámetros es una modificación del modelo de dos parámetros para incorporarle un tercer parámetro asociado a la posibilidad de adivinar la respuesta correcta. Esta modificación fue propuesta por Birnbaum en 1968 [14]. La función correspondiente es:

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-a(\theta-b)}}$$

En la cual

$a$  es el parámetro de discriminación.

$b$  es el parámetro de dificultad.

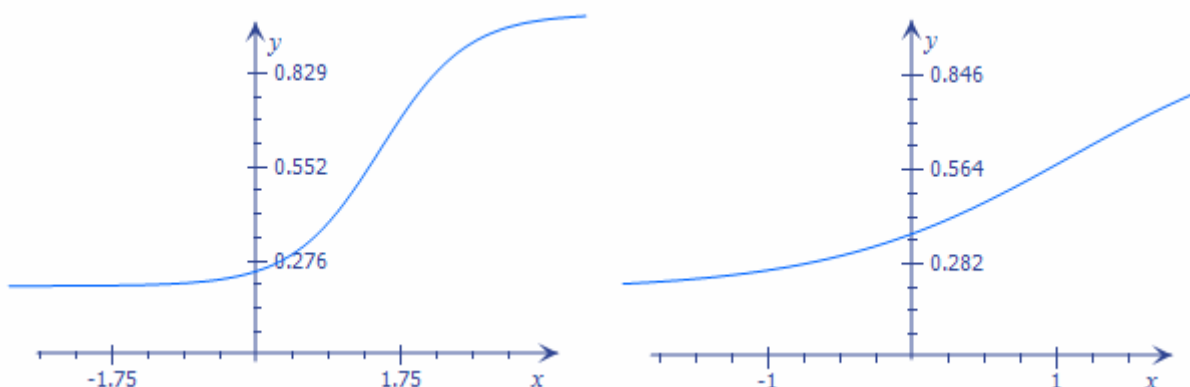
$c$  es el parámetro de adivinación (posibilidad de adivinar).

$\theta$  es el nivel de habilidad.

El parámetro  $c$  es la novedad. Este parámetro indica la probabilidad de que un alumno adivine la respuesta correcta. Es un parámetro independiente del nivel de habilidad, dado que todos los estudiantes tienen la misma probabilidad de adivinar. En general los valores aceptables de  $c$  son valores en el rango de 0 y 0.35, aunque teóricamente el rango válido es de 0 a 1.

La inclusión del parámetro de probabilidad de adivinación afecta la definición de  $b$  como parámetro de dificultad. Ahora el límite inferior de la curva no es 0 sino  $c$ . La definición del parámetro de dificultad como el punto donde  $P(\theta) = 0.5$ , debe cambiarse a un punto donde  $P(\theta)$  esté entre  $c$  y 1. Entonces el nivel de dificultad del ítem es el punto en donde  $P(\theta) = (1+c)/2$ . Al mismo tiempo, la pendiente de la curva en  $b$  es  $a(1-c)/4$ .

Las siguientes son dos curvas características para valores de  $a=1.9$   $b=1.5$  y  $c=0.2$  y para valores  $a=1.2$   $b=1.1$  y  $c=0.2$ .



La primera tiene un mayor nivel de dificultad y es más discriminante como se puede observar de las figuras. Es importante destacar que el parámetro de dificultad como valor de posicionamiento del ítem es simplemente un punto en la escala de habilidad, para este último modelo, aquel donde  $P(\theta_i) = (1+c)/2$ . Esto quiere decir que un alumno con habilidad  $\theta_i$  tendrá esa probabilidad de responder correctamente el ítem.

### **ITR en tests adaptativos por computadora**

Los test adaptativos utilizan la teoría de ITR para ir adecuando los ítems presentados al alumno a partir de la estimación de la habilidad del mismo. Para crear un test adaptativo por computadora es necesario contar con un conjunto de ítems estadísticamente evaluados. Estos ítems serán presentados al usuario de acuerdo a un algoritmo de secuenciación preestablecido. A medida que se van presentando ítems, se realiza un procedimiento para evaluar la habilidad del alumno según su respuesta. Para estimar la habilidad del alumno existen varias técnicas [15], [16], [17]. Nosotros hemos usado una técnica propuesta por Runder [12]. Se comienza con una estimación inicial (basada en las habilidades promedios de grupos anteriores)  $\theta_s$  y a medida que se van respondiendo ítems, se realiza un ajuste. La forma de ajustar la estimación se ve en la siguiente fórmula:

$$\theta_{s+1} = \theta_s + \frac{\sum S_i(\theta_s)}{\sum I_i(\theta_s)}$$

En donde  $I(\theta)$  es una función que mide la cantidad de información de un ítem para el nivel de habilidad  $\theta$ . Esta información es también utilizada por el agente de test para la elección del próximo ítem a presentar. Normalmente, el ítem seleccionado es aquel que provee más información,

entendiendo que un ítem demasiado fácil para el nivel del estudiante, así como uno demasiado dificultoso, no provee demasiada información. La selección de ítems también está sujeta a otras restricciones de contenido (por ejemplo el docente puede querer que los alumnos contesten al menos una cantidad determinada de ítems correspondientes a cada tema). La fórmula que indica la información para cada ítem según el nivel de habilidad es:

$$I_i(\theta) = \frac{P'_i(\theta)^2}{P_i(\theta)(1 - P_i(\theta))}$$

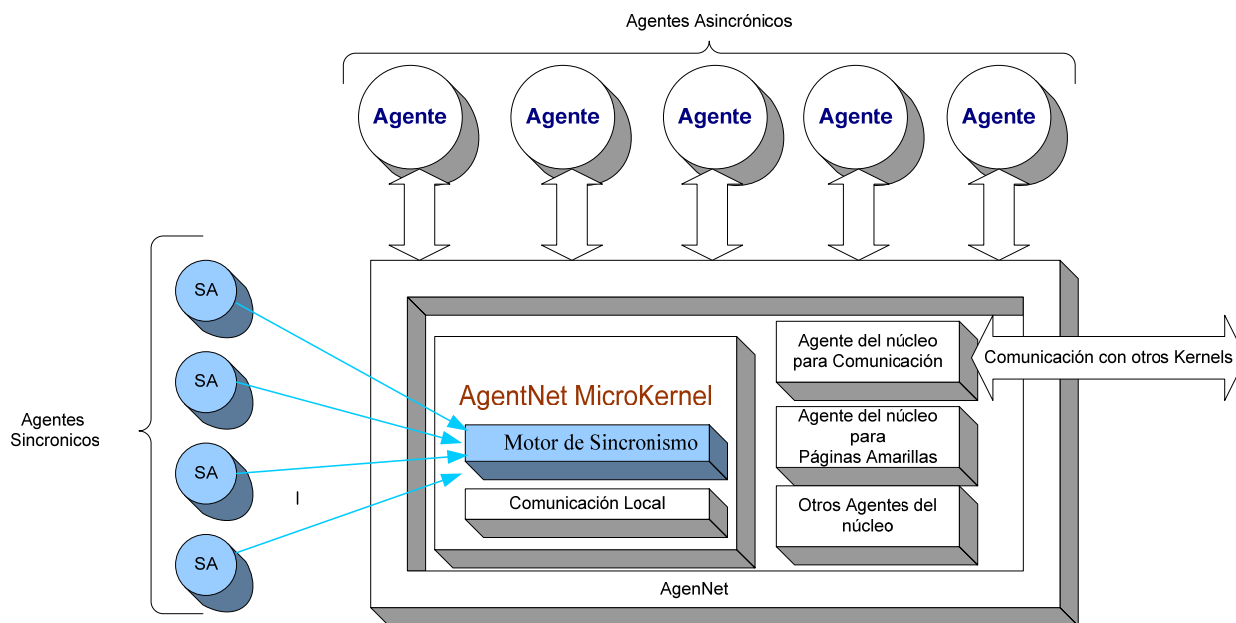
La función  $S_i$  es similar a  $I_i$  pero resta o suma según la respuesta al ítem  $i$  sea incorrecta o correcta. De esta forma, se van eligiendo los mejores ítems y ajustando la habilidad estimada del alumno hasta que se cumple un criterio de finalización determinado.

## Agentes Inteligentes y Sistemas Multiagentes

Si bien no hay una definición universalmente aceptada del término *agente*, sí existen criterios más o menos generales acerca de algunas de las características que deben poseer los sistemas basados en agentes. En ese sentido, surgen las nociones débiles y fuertes de agencia [4]. Se considera entonces a un agente como a “*un sistema de computación que está ubicado en algún ambiente y que es capaz de accionar autónomamente en dicho ambiente en orden a cumplir sus objetivos*” [6]. La definición dada se basa en la noción débil de agencia que implica que los agentes tienen autonomía, habilidad social, son reactivos y pro-activos. La noción fuerte de agencia, en donde se ubican también los agentes inteligentes e incluyen una serie de actitudes mentales, implica además: movilidad, veracidad, benevolencia y racionalidad [4].

En [8] se clasifican los agentes en función de tres propiedades fundamentales: autonomía, cooperación y aprendizaje. La autonomía implica pro-actividad e iniciativa, es decir los agentes tienen control sobre su estado interno y su comportamiento. La cooperación es la base de la habilidad social y es la propiedad que permite a los agentes comunicarse e interactuar con otros. Los lenguajes de comunicación de agentes como FIPA-ACL son esenciales en dicho sentido, ya que estos lenguajes basados en la teoría de los actos del habla estandarizan la comunicación. Además de la cooperación ínter agentes clásica de los sistemas multiagentes, también se considera la cooperación con los humanos. El aprendizaje es la habilidad del agente de mejorar su comportamiento a medida que transcurre el tiempo, es decir de aprender de su interacción con otros agentes y/o humanos. En [8] se establece a partir de esos atributos una tipificación de agentes en: agentes colaborativos, agentes colaborativos que aprenden, agentes de interfaz y agentes inteligentes.

Para implementar el sistema adaptativo, hemos utilizado el marco de trabajo *AgentNet* [1] que permite la construcción de sistemas multiagentes en diversos lenguajes sobre la plataforma .NET. Para ello, aprovechamos las facilidades brindadas por *AgentNet* para construir agentes y embeberlos en aplicaciones Web. La siguiente figura es un esquema de la arquitectura de AgentNet.



## Arquitectura del Sistema de Enseñanza Adoptivo basado en Web.

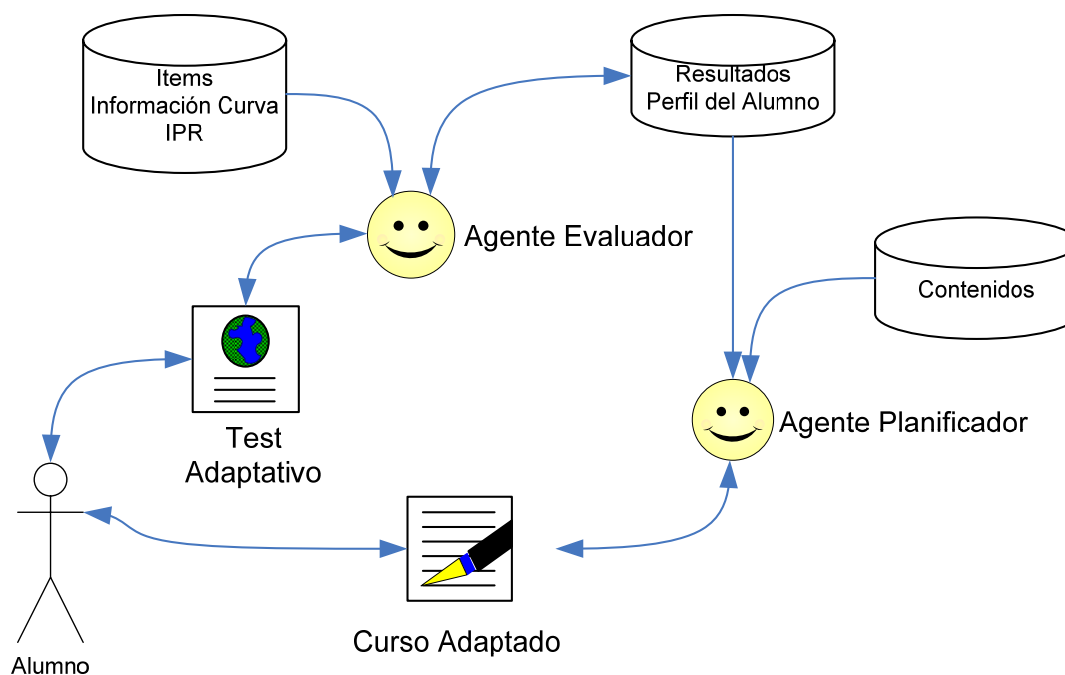
Nuestro sistema se basa en un entorno Web sobre el cual se le van presentando al alumno los materiales correspondientes a cada módulo del curso. Al finalizar cada módulo, el alumno debe realizar un test de evaluación. Dicho test permite una mejor selección del material del próximo módulo, indicando por ejemplo material para reforzar falencias o materiales de distintos niveles de dificultad, así como ejercicios acordes, no solamente para los alumnos más atrasados sino dando materiales avanzados a aquellos que estén en condiciones de asimilarlos. Al principio, el alumno realiza una evaluación de diagnóstico que sirve como punto de partida para el curso. Esta evaluación tiene mucha importancia porque define los arreglos que deben ser realizados a fin de lograr el mejor desempeño del alumno. Cumple el mismo rol que la evaluación diagnóstica de la enseñanza tradicional permitiendo al docente realizar los ajustes que considere convenientes. En este caso, además, la mayor parte de los ajustes son realizados por un agente inteligente: el agente planificador.

La evaluación adaptativa sirve al agente planificador para reorientar la práctica docente según las necesidades evidenciadas por el resultado de la misma, acorde a lo planteado en [25].

Los tests al final de cada curso (y el test de diagnóstico) son administrados por el agente evaluador según la metodología presentada, utilizando la técnica de ITR y test adaptativos. El agente evaluador actualiza el perfil del alumno según los resultados del mismo.

Al finalizar el test, el agente de planificación se encarga de organizar y presentar los materiales al alumno y un agente de control informa al docente de la actividad y resultados parciales. El tutor puede además agregar o tomar control del material que el agente de planificación selecciona.

El siguiente diagrama muestra un esquema de la arquitectura que estamos describiendo (se obvió por razones de simplicidad, al agente de control):



Este sistema fue evaluado en cursos básicos de conceptos de orientación a objetos con un resultado satisfactorio, permitiendo a los alumnos avanzar según su ritmo y facilitó al tutor la tarea de seguimiento y control de los mismos.

El funcionamiento final es el de un sistema hipermedia adaptativo ya que la navegación sugerida es en función del nivel de habilidad del alumno adaptando navegación y contenido. Para implementar el curso de prueba de esta arquitectura se realizó una aplicación Web a la cual se le embebió el *kernel* de AgentNet. La aplicación instancia los agentes cuando éstos son necesarios. Estos agentes son agentes asincrónicos con suficiente autonomía para generar los test y para presentar el módulo en cuestión.

## Conclusiones

Hemos presentado la arquitectura para un sistema de enseñanza adaptativo basado en Web [23] [24] apto para la educación a distancia. La utilización de la teoría de respuesta a ítems permite tratar con el problema del modelo del estudiante (tal cual utilizan los sistemas tutores inteligentes), ya que el nivel de conocimiento del estudiante es medido efectivamente a través de tests adaptativos.

Actualmente estamos trabajando en una mejora de los agentes y el motor que evalúa la respuesta a los ítems. Además, planificamos la construcción de un agente de ajuste de los parámetros de los ítems. Es fundamental esto último para calibrar las variables y eliminar ítems defectuosos. Por ejemplo, un ítem cuya curva característica se represente mediante una función de discriminación negativa es un ítem a analizar.

Actualmente, estamos analizando la teoría de respuesta a ítems multidimensional (MIRT). La IRT tiene una visión unidimensional, en el sentido de que todos los ítems en un test miden solamente una habilidad. La teoría de respuesta a ítems multidimensional elimina la restricción de que los ítems midan sólo un tipo de habilidad permitiendo la inclusión de ítems que midan diversas habilidades [26] [27]. Si bien esto agrega complejidad al motor de evaluación puede ser muy útil en contextos donde se necesita medir diversos tipos de desempeños.

## Bibliografía

- [1] Rossel, Gerardo; Andrea Manna. *Desarrollando Sistemas Multiagentes sobre AgentNet*. X Congreso Argentino de Ciencias de la Computación. CACIC. Argentina. Octubre 2004.
- [2] Carbonell, J. R. *AI in CAI: an artificial intelligence approach to computer assisted instruction*. IEEE transaction on Man Machine System. V 11 n.4., 1970.
- [3] Wenger, E. *Artificial intelligence and tutoring systems*. Computational and Cognitive Approaches to the Communication of Knowledge. Los Altos C. A. Morgan and Kaufman. 1987
- [18] Murray T. *Authoring Intelligent Tutoring Systems: An Analysis of the state of the art* International Journal of Artificial of Artificial Intelligence in Education.(12) 98-129. (1999)
- [19] Parr, A.; E, A. Reigenaum (Eds). *The Handbook of Artificial Intelligence*, Application Oriented AI Research: Education Kaufmann, Los Altos, CA, 1982.
- [20] D. Pilar da Silva, R. V. Durm, E. Duval, y H. Olivie. *Concepts and documents for adaptive educational hypermedia: a model and a prototype*. Computing Science Reports, Eindhoven University of technology, Eindhoven. 1998.
- [21] P. Brusilovsky, J. Eklund, y E. Schwarz. *Web-based education for all: A tool for developing adaptive courseware*. Computer Networks and ISDN Systems, 30 (1-7) 1998.
- [22] P. De Bra. *Teaching Hypertext and Hypermedia through the Web*. Journal of universal computer science, 1996.
- [23] Peter Brusilovsky, John Eklund, Elmar Schwarz. *Web-based Education for All: A Tool for Development Adaptive Courseware* Computer Networks and ISDN Systems (Proceedings of Seventh International World Wide Web Conference, 14-18 April 1998 30 (1-7), 291-300.
- [24] Peter Brusilovsky. *Adaptive Educational Systems on the World Wide Web: A review of Available Technologies*. Proceedings of Wrokshop "WWW-Based Tutoring" 4ta. Internacional Conference on Intelligent Tutoring Systemas (ITS'98). San Antonio, TX. 1998
- [4] Michel Wooldridge, Nicholas Jennings, *Intelligent Agents: Theory and Practice en Knowledge Engineering Review*. (1995).
- [6] Michel Wooldridge. *Intelligent Agents in Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Chapter 1. The MIT Press. Gerhard Weiss editor.(1999)
- [7] Fuhua Lin, Mohamed Ally. *Designing Distributed Learning Environments with Intelligent Software Agents*., Information Science Publishing. 2005
- [8] Nwana, H. S. *Software agents: An overview*. The Knowledge Engineering Review, 11(3),1996.
- [9] FIPA Standards. *FIPA ACL Message Structure Specification*. Document SC00061G, (2002).
- [10] Van der Linden, WJ y Hambleton, RK (Editors). *Handbook of Modern Item Response Theory*. London: Springer Verlag 1997.
- [11] Frank B. Baker. *The Basics of Item Response Theory*. Second Edition ERIC Clearinghouse on Assessment and Evaluation. 2001.
- [12] Rist, S., Rudner, L. & Wise, L. *Computer Adaptive Tests*. ERIC Digest Series.1989
- [13] Rudner, L. *Computer Testing: Research Needs Based on Practice*. Educational Measurement: Issues and Practice, 2, 1990



- [14] Birnbaum, A. *Some latent trait models and their use in inferring an examinee's ability*. Part 5 in Lord F.M. and M.R. Novick. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.
- [15] Lord, F.M. *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum Associates. 1980
- [16] Kingsbury, G., Zara, A. *Procedures for Selecting Items for Computerized Adaptive Tests*. *Applied Measurement in Education*, 2(4), 1989
- [17] Kingsbury, G., Zara, A. A Comparison of Procedures for Content-Sensitive Item Selection in Computerized Adaptive Tests. *Applied Measurement in Education*, 4 (3) 1991
- [25] Stenhouse, L. *La investigación como base de la enseñanza*. Barcelona: Morata. 1994.
- [26] Ackerman, T.A. Using multidimensional item response theory to understand what items and tests are measuring, *Applied Measurement in Education*. 1994
- [27]. Spencer, S.G. *The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated*. Thesis of Doctorate. Brigham Young University 2004.
- [28] Taylor, James C. *Distance Education. The Fifth Generation*. IXX Conferencia Mundial del ICDE. Viena. 1999