

Tests Adaptativos utilizando Lógica Difusa

Gerardo Rossel

Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana- Av. Montes de Oca 745
Buenos Aires, C1270AAH, Argentina.
Departamento de Computación, FCEN-Universidad de Buenos Aires
Ciudad Universitaria, Pab. I.
Buenos Aires, C1428EGA, Argentina
grossel@computer.org

y

Andrea V. Manna

Departamento de Computación, FCEN-Universidad de Buenos Aires
Ciudad Universitaria, Pab. I. Buenos Aires, C1428EGA, Argentina.
amanna@dc.uba.ar

Abstract

Paper and pencil traditional test not include the flexibility that can be obtained by means of the automation. The computer adaptive testing offers a great flexibility, permitting to adapt the test to examinee level. The most it utilized underlying theory is the item response theory. It is theory offers a lot of complexity for the implementation, in this article we show an implementation alternative based on fuzzy logic. This alternative is integrated in a multiagent system in order to achieving architecture for an adaptive educational system Web-based.

Keywords: Computer Adaptive Testing, Fuzzy Logic, Item Response Theory

Resumen

Los test tradicionales en papel y lápiz no cuentan con la flexibilidad que puede obtenerse mediante la automatización de los mismos. La implementación de sistemas de test adaptable por computadora (CAT Computer Adaptive Testing) brinda una gran flexibilidad, permitiendo adaptar los test al nivel del alumno. La teoría subyacente más utilizada es la teoría de respuestas a ítems. Esta teoría ofrece bastante complejidad para su implementación. En este artículo mostramos una implementación alternativa basada en lógica difusa sobre un motor de prolog. Esta alternativa es integrada en un sistema de agentes con el fin de lograr una arquitectura para un sistema de enseñanza adaptativo basado en Web.

Palabras claves: Testeo adaptativo por computadora, lógica difusa, teoría de respuesta a ítems.

1 INTRODUCTION

Un test adaptativo por computadora (CAT) es sistema de evaluación administrado por software en el cual el ítem presentado al alumno, la secuencia de los mismos y la finalización del test se resuelve en forma dinámica a medida que el alumno responde basándose en el desempeño del alumno a lo largo del test. A partir del desempeño y eventualmente otras variables (en nuestro caso hemos utilizado desempeño y tiempo de respuesta) se decide cuál es el próximo ítem a presentar.

Los test adaptativos por computadora presentan algunas ventajas con respecto a los test tradicionales. Las ventajas son tanto de índole operativa, como ser administración automática de los ítems, soporte informático para la evaluación, etc. como de índole propiamente pedagógica como reducción del tiempo de evaluación (con la consiguiente mejor predisposición de los alumnos) y la estimación más precisa del nivel de los alumnos (lo cual como veremos facilita además la creación de sistemas de enseñanza adaptativos). Como desventaja más importante se encuentra la imposibilidad de revisar respuestas anteriores.

La utilización de test adaptativos por computadora permite superar algunas de las deficiencias detectadas en la teoría clásica de los test (TCT). La TCT [5] se basa en que la puntuación empírica o sea el valor medido por el test puede ser dividido en dos partes: la habilidad verdadera y el error de medida. La habilidad verdadera no puede ser observada directamente. En su lugar se obtiene una habilidad medida (entendiendo por habilidad el atributo a ser medido por el test) que es una medida indirecta de la habilidad real y está asociada a algún grado de error. La principal dificultad de la teoría clásica de test y la que se intenta resolver mediante los test adaptativos, es la interdependencia entre los examinados y el test. Por un lado, las mediciones dependen del test y por otro, las propiedades estadísticas del test dependen de los alumnos examinados. Esto es así dado que para la teoría clásica de test la habilidad se mide según el número de respuestas correctas, lo cual condiciona el resultado al grado de dificultad del test. Pero a su vez, para calcular la dificultad de la pregunta se toma el porcentaje de alumnos que la han contestado correctamente lo cual hace difícil extrapolar y generalizar los resultados. Por ello los test adaptativos (y fundamentalmente la teoría de respuesta a ítems) son manejados por ítems (*item-driven*) en contraste con la teoría clásica que es manejada por los test (*test-driven*).

Los test adaptativos por computadora siguen un algoritmo esquematizado en el siguiente pseudocódigo:

```
mientras not (condicion _ fin) hacer  
    item = seleccionar _ item _ optimo  
    presentar _ al _ alumno(item)  
    obtener _ respuesta  
    estimar _ habilidadd  
fin _ mientras
```

La teoría más utilizada para la construcción de test adaptativos es la teoría de respuesta a ítems que ha sido aplicada exitosamente [12][2][9][18]. En este trabajo, en cambio, presentamos como instrumento para la adaptación del test a la utilización de la lógica difusa.

Este trabajo se organiza de la siguiente manera: en la sección 2 describimos brevemente los

conceptos básicos de lógica difusa utilizados, luego detallamos el diseño de las variables y reglas difusas para el soporte de test adaptativos y luego la integración del programa lógico difuso en el contexto de un sistema de agentes para el soporte de sistemas de enseñanza basado en Web.

2 LÓGICA DIFUSA COMO HERRAMIENTA

La lógica difusa o borrosa [15] [16] [17] (sería más correcto hablar de lógicas difusas en plural) es una herramienta ideal para crear programas que puedan adaptar su respuesta al estado del entorno, fundamentalmente en condiciones donde hay cierto grado de imprecisión en la clasificación de las variables de entrada. La respuesta a cambios en el entorno se basa en la aplicación de reglas difusas sobre variables difusas produciendo un resultado acorde. Las aplicaciones de este tipo de lógicas son enormes: desde sistemas de control no lineales hasta sistemas de diagnóstico. Son básicamente lógicas multivaluadas que extienden a las lógicas clásicas.

A continuación, definiremos algunos de los términos utilizados en la estructura de los programas lógicos difusos y en nuestra implementación. Hay dos componentes principales: las *variables difusas* y las *reglas difusas*.

Las variables difusas requieren del concepto de *conjunto difuso*. Las variables difusas tienen asociado un calificador lingüístico (de allí la fórmula propuesta por Zadeh *Fuzzy logic = computing with words* [18]) y una función de membresía a un conjunto difuso. Muchos conjuntos no cumplen con la condición de que un elemento *es o no es* miembro de él, tal cual se establece en teoría de conjuntos clásica donde un elemento x pertenece a un conjunto o pertenece a su complemento. Los límites de algunos conjuntos son *borrosos*, por ejemplo el conjunto de las personas *altas*. ¿Cuál es la condición para ser miembro de ese conjunto? Una persona de estatura mayor a 1.80 m puede ser considerada alta ¿pero una de 1.79? ¿Dónde se pone el límite? La propuesta de los conjuntos difusos es, en lugar de definir membresía como una cualidad *verdadera o falsa* de un elemento, definir un *grado* de membresía de tal forma que se logra una transición gradual entre membresía y no membresía. Para cada valor definido (*crisp*) de una variable difusa, se pueden obtener los grados de membresía de ésta a los conjuntos difusos. A su vez, los grados de membresía pueden ser calificados mediante lo que se denomina un calificador o “*hedge*” lingüístico.

Una regla difusa (*fuzzy rule*) es una regla que cuenta con una o más variables difusas en su antecedente y una o más variables difusas en su consecuente. El proceso de aplicar las reglas difusas es conocido como propagación. Las reglas difusas pueden también estar reunidas en una memoria asociativa difusa o matriz difusa.

Un programa de lógica difusa es entonces un proceso de tres estados: *fuzzification*, propagación y *de-fuzzification*.

3 LÓGICA DIFUSA PARA TEST ADAPTATIVOS

Nuestro modelo de test adaptativo utiliza la calidad de las respuestas (si fueron correctas o no y la cantidad de información que brindan) y el tiempo promedio como variables difusas de entrada. Los ítems a ser contestados por los alumnos tienen un tiempo máximo por lo cual es posible utilizar el tiempo que se demora en contestar como medida de la complejidad que representa para el alumno el ítem.

La salida del programa lógico difuso es un ajuste acorde para la selección del próximo ítem a presentar al alumno evaluado. De esta forma, se está seleccionando el ítem que provee mayor información acorde al desempeño del alumno hasta el momento [13] [12]. La definición de las variables y todo el programa lógico difuso está realizada sobre prolog utilizando la herramienta de

razonamiento en condiciones de incertidumbre de LPA llamada Flint [8].

3.1 Definición de variables difusas

Hemos utilizado variables difusas cuyo calificador de curvatura es lineal. De todas formas esto puede fácilmente ajustarse si en algún tipo de test es necesario. La variable que mide el tiempo de demora de la respuesta y la variable que mide el desempeño del alumno en el test tienen la siguiente definición.

fuzzy_variable(tiempo_respuesta) :-
[0,360];
muy_rapido, \ , linear, [0,60];
rapido, ^ , linear [40, 75,110];
normal /-\ , linear [100,120,160,180];
lento /-\ , linear [170,200,230,260];
lentisimo / , linear, [240,300].

fuzzy_variable(desempeño) :-
[0,100];
muy_bajo, \ , linear, [0,25];
bajo, /-\ , linear, [20,30,40,50];
aceptable, /-\ , linear [40, 50,60,70];
bueno /-\ , linear [55,60,70,75];
muybueno / , linear [70,90].

El desempeño es medido en un rango de 0 a 100 y es el promedio de respuestas correctas pesado por la información que brindan esas respuestas. Todos los alumnos comienzan con un desempeño inicial. El mismo puede ser el resultado de las evaluaciones anteriores en caso de haberlas o un desempeño medio si no hay evaluaciones anteriores.

Los ítems son ordenados de mayor a menor en un rango de 0 a 10 siendo 0 los ítems de menor dificultad y 10 los de mayor dificultad. La variable *nivel_actual*, que indica el nivel del ítem que está siendo presentado al alumno, será ajustada por el resultado del programa difuso a partir de la *de-fuzzification* de la variable difusa *ajuste*. La variable *ajuste* tiene un rango de -3 a 3 que indica cuanto debe descender o ascender el nivel del ítem en la próxima respuesta. A continuación, la definición de la variable difusa *ajuste*:

fuzzy_variable(ajuste) :-
[-3,3];
bajar_nivel, \ , linear, [-2,0];
mantener, ^ , linear, [-1,0,1];
subir_nivel, / , linear [0,2,];
centroid(largest_membership, mirror_rule, shrinking).

En este caso, la variable *ajuste* debe ser *de-fuzzificada* por lo cual se debe especificar la función de utilizar para ello, usamos la técnica del *centroide*. El método del centroide trabaja buscando el centro de gravedad de los grados de membresía para todas las membresías (aunque esto puede

variar en función de los parámetros usados) para todos los calificadores lingüísticos para la variable difusa.

3.2 Reglas difusas

El programa utiliza reglas difusas en prolog basadas en las variables definidas en el punto anterior. Por razones de espacio y simplicidad sólo mostraremos algunas de las reglas definidas. Todo programa prolog basado en lógica difusa sobre Flint debe definir los operadores para escribir las reglas:

```
:- op( 1150, fty, ( if ) ),  
op( 1150, xfy, ( then ) ), op( 1150, xfx, ( else ) ),  
op( 1100, xfy, ( or ) ),op( 1000, xfy, ( and ) ),  
op( 700, xfx, ( is ) ), op( 600, fty, ( not ) ).
```

De esta manera podemos escribir las reglas en una sintaxis clara. A continuación una de las reglas utilizadas para ajustar a cuando el tiempo de respuesta es rápido y el desempeño es aceptable:

```
uncertainty_rule( ajustar1 ) :-  
    if tiempo_respuesta is rapido  
    and desempeño is aceptable  
    then ajustar is subir_nivel.
```

Cuando el tiempo de respuesta es normal y desempeño es bueno entonces se debería mantener el nivel actual. Eso se expresa en la siguiente regla:

```
uncertainty_rule( ajustar2 ) :-  
    if tiempo_respuesta is normal  
    and desempeño is bueno  
    then ajustar is mantener
```

Cada regla tiene un átomo que es el nombre de la misma y se puede usar para referirse a ella. Otras reglas usadas son:

```
uncertainty_rule( ajustar3 ) :-  
    if tiempo_respuesta is lento  
    and desempeño is bueno  
    then ajustar is bajar_nivel
```

```
uncertainty_rule( ajustar2 ) :-  
    if tiempo_respuesta is muy_rapido  
    and desempeño is bueno  
    then ajustar is subir_nivel
```

3.3 Cálculo de variables difusas

El cálculo del nivel actual se realiza en base a las respuestas dadas por el alumno evaluado.

Representa el porcentaje de respuestas correctas según su nivel de información aportado, en forma similar a la que lo hace la teoría de respuesta a ítems. Una respuesta correcta a una pregunta de nivel muy bajo (de poca dificultad) a un alumno de alto nivel no aporta mucha información. Lo mismo que una respuesta incorrecta a un ítem de alto nivel para un alumnos de nivel bajo. Para calcular el promedio, se utiliza la distancia del ítem al nivel al momento de contestar. La fórmula utilizada es la siguiente

$$des = \frac{\sum \left(\frac{1}{d_i} * u_i \right)}{\sum \frac{1}{d_i}} * 100$$

Donde d_i es:

$$d_i = (nant - nivel_i) + 1$$

Siendo $nant$ el nivel actual (es decir el nivel del ítem que está respondiendo el alumno), des el nuevo desempeño calculado y $nivel_i$ el nivel del ítem i .

3.4 Esquema general de funcionamiento

El diagrama siguiente muestra el esquema general de utilización de las reglas y variables definidas anteriormente para adaptar el test según las repuestas del alumno evaluado

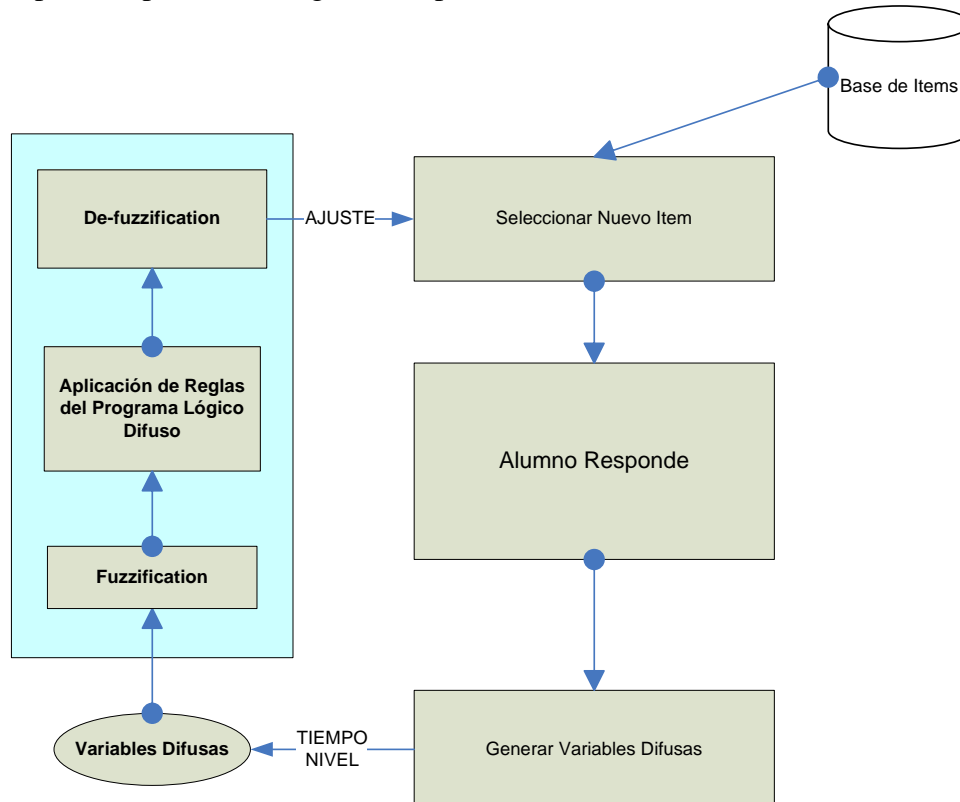


Fig. 1 Esquema general

Los supuestos presentes en la teoría de respuesta a ítems, son asumidos en nuestro modelo basado en lógica difusa. Particularmente asumimos la independencia estadística de las respuestas a los

ítems y que la probabilidad de contestar correctamente un ítem es una función cuya curva se encuentra dentro de la familia de las curvas acumulativas de la función logística [2]. La existencia de esa curva para un nivel de habilidad dado presupone la consistencia de las respuestas, es decir que un alumno con un nivel determinado responderá correctamente los ítems de nivel más bajo y responderá incorrectamente los ítems de nivel más alto (según la probabilidad dada por la curva). Si en cambio responde la mayoría de los ítems más altos correctamente pero no los ítems de nivel más bajo, entonces nos encontramos ante un caso de inconsistencia de las respuestas [1] que puede deberse a diversos factores (entre ellos a una mala calificación de los ítems). El análisis de la consistencia de las respuestas permitirá entonces ajustar el nivel de cada ítem. La figura 2 es un ejemplo de dicha curva, llamada en el contexto de IRT curva característica del ítem:

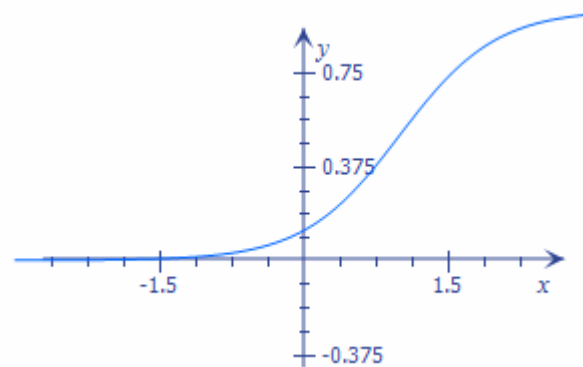


Fig.2 Ejemplo Curva Característica de un ítem

4 AGENTES DE CONTROL DEL PROGRAMA DIFUSO

En esta sección describiremos la arquitectura general basada en agentes para la utilización del programa lógico difuso en el marco de un sistema de enseñanza adaptativo basado en Web (SEAW) (SEAW) [3] [4]. Para la implementación utilizamos el marco de trabajo *AgentNet* [7], dado su flexibilidad y facilidad para embeber agentes inteligentes en aplicaciones Web. Esta arquitectura utiliza como base la que hemos implementado utilizando IRT [6].

Dada las diversas definiciones de agentes y sistemas multiagentes, utilizamos este trabajo 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” [14]. 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, incluyen una serie de actitudes mentales, implica además: movilidad, veracidad, benevolencia y racionalidad [14].

Los sistemas de enseñanza adaptativos para Web basados en los sistemas hipermediales adaptativos y en los sistemas tutores inteligentes permiten mejorar los sistemas de enseñanza a distancia sobre Web personalizando el acceso al contenido. Es importante la adaptación personalizada del contenido porque la adaptabilidad de los sistemas hipermediales clásicos basada en la libre navegación no es suficiente, ya que el alumno carece de guía para seguir el contenido. La adaptación de los SEAW se da en dos facetas, la adaptación de la navegación y la adaptación del contenido. A las ventajas que nos proveen los test adaptativos en general, nuestra implementación en particular permite agregar su integración en SEAW. Para ello, desarrollamos una arquitectura que incorpore el sistema de test adaptativo basado en lógica difusa a un SEAW utilizando tecnología de agentes. La figura 3 muestra un diagrama de la arquitectura general del sistema

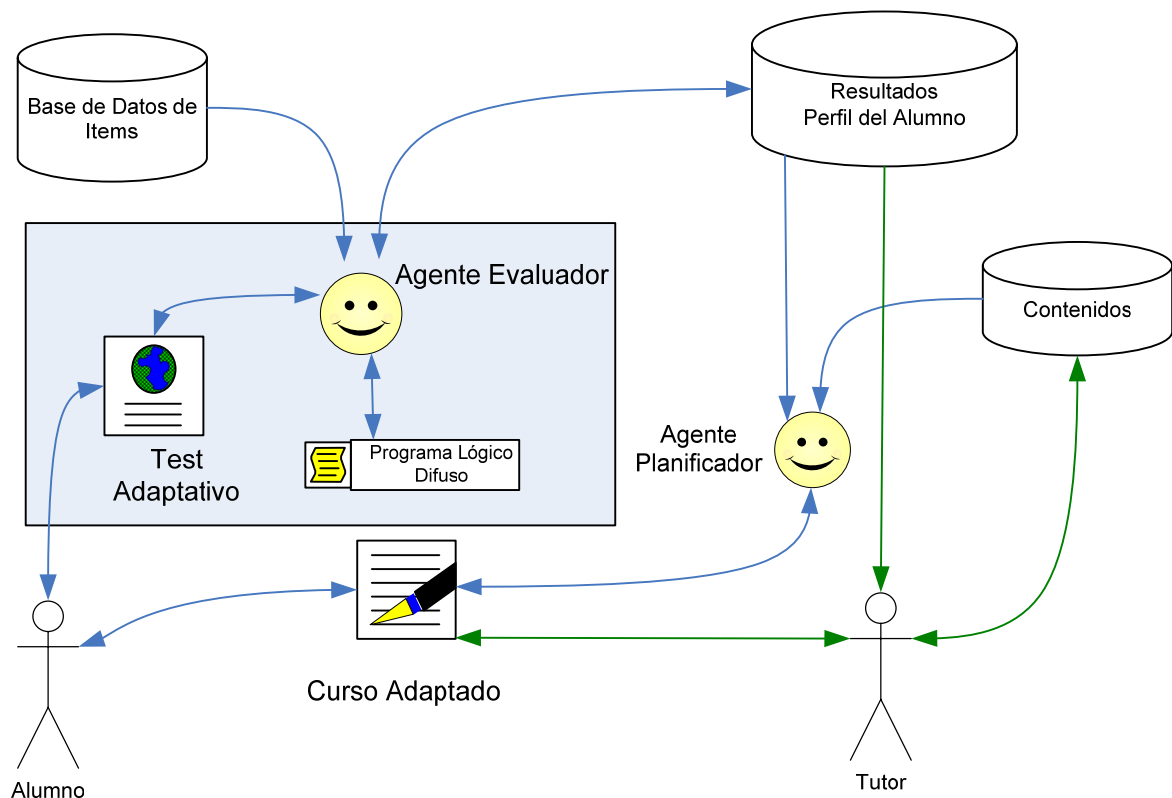


Fig.3 Arquitectura del SEAW

El resultado del test adaptativo es incorporado al perfil del alumno para ser usado por el agente planificador con el objetivo reorientar la práctica docente según las necesidades evidenciadas por el resultado de la misma, acorde a lo planteado en [10].

El SEAW esquematizado la figura 3 cuenta con dos agentes autónomos:

- 1) El agente evaluador que es el que encapsula el llamado al programa de lógica difusa, da soporte al diálogo con el estudiante y administra el test en general
- 2) El agente planificador. Este agente examina los resultados de los test y presenta el próximo módulo a enseñar. El contenido de este nuevo módulo es adaptado a cada estudiante según su desempeño en los tests.

El tutor interactúa con el sistema pudiendo cambiar u orientar los contenidos presentados a cada alumno por el agente planificador. Este tipo de sistemas se ubica entre la cuarta y quinta generación de sistemas de enseñanza a distancia según la clasificación de James Taylor [11].

5 CONCLUSIONES

Hemos descrito la utilización de lógica difusa para la creación de test adaptativos por computadora. La misma permite el ajuste de la selección del ítem a presentar según el desempeño del alumno hasta el momento. Además presentamos una arquitectura para la incorporación de nuestro modelo de test adaptativo dentro un sistema de enseñanza adaptativo basado en Web. Para ello utilizamos un sistema Multiagente basado en el marco de trabajo *AgentNet*. Si bien las primeras pruebas que realizamos con grupos reducidos de alumnos en cursos introducción a la orientación a objetos han sido satisfactorias, debemos generar más casos de prueba y comparar lo implementado con un modelo más tradicional de teoría de respuesta a ítems que nos permite avanzar en los aspectos

teóricos de la utilización de la lógica difusa en los test adaptativos que permita mejorar el modelo psicométrico asumido. Actualmente estamos comparando esta implementación con una basada únicamente en IRT que presentamos en otro trabajo [6]

Otro de los aspectos que estamos considerando es la mejora del agente planificador para mejorar el desempeño y que además de seleccionar contenidos nuevos y de repaso pueda orientar en la navegación por los mismos.

BILIOGRAFIA

- [1] Betz, N. E. (1975). New types of information and psychological implications. *D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and Prospects* (Research Report 75-5), pp. 32-43. Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- [2] Baker, Frank B. The Basics of Item Response Theory. Second Edition *ERIC Clearinghouse on Assessment and Evaluation*. 2001.
- [3] Brusilovsky, Peter; Eklund, John; Schwarz, Elmar. 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.
- [4] Brusilovsky, Peter. 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
- [5] Novick, M. R. The Axioms And Principal Results Of Classical Test Theory. *Journal Of Mathematical Psychology* 3. (1966).
- [6] Rossel, Gerardo. Usando IRT y Agentes para Educación Distancia Adaptativa. *I Congreso de Tecnología en Educación y Educación en Tecnología* 2006. La Plata. Buenos Aires. Argentina
- [7] Rossel, Gerardo; Manna, Andrea. Desarrollando Sistemas Multiagentes sobre AgentNet. *X Congreso Argentino de Ciencias de la Computación. CACIC* 2004. Argentina. Octubre 2004.
- [8] Shalfield , Rebeca. Flint reference. <http://www.lpa.co.uk>. 2004
- [9]. 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.
- [10] Stenhouse, Lawrence. La investigación como base de la enseñanza. Barcelona: Morata. 1994.
- [11] Taylor, James C. *Distance Education. The Fifth Generation*. IXX Conferencia Mundial del ICDE. Viena. 1999
- [12] Van der Linden, WJ; Hambleton, RK (Editors). Handbook of Modern Item Response Theory. London: Springer Verlag 1997.
- [13] Weiss, D., Kingsbury, G. Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*. 12. 1984

- [14] Wooldridge Michel. *Intelligent Agents in Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Chapter 1. The MIT Press. Gerhard Weiss editor.(1999)
- [15] Zadeh, L.A.. Fuzzy Sets. *Information and Control* Vol8. 1965
- [16] Zadeh, L.A.. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. on Systems, Man and Cybernetics*. January 1973.
- [17] Zadeh, L.A.. Fuzzy Logic *IEEE Computer Mag.* April 1988.
- [18] Zadeh, L.A.. Fuzzy logic = computing with words *IEEE Transactions on Fuzzy Systems* May 1996 Page(s):103 – 111