

Más allá del Razonamiento Basado en Casos y una Aproximación al Modelado de Sociedades Utilizando Minería de Datos

Alberto Ochoa^{1,4}, José-Alberto Hernández², Francisco J. Alvarez³, Gennadiy Burlak², & Jaime Muñoz³

¹ Universidad Autónoma de Zacatecas

Jardín Juárez 147 Centro Histórico, Zacatecas, México, CP. 98000

⁴ Instituto Tecnológico de León (Maestría en Sistemas Inteligentes)

² Centro de Investigación en Ingeniería y Ciencias Aplicadas, UAEM

Av. Universidad 1001 Col. Chamilpa, Cuernavaca, Morelos México, CP. 62209

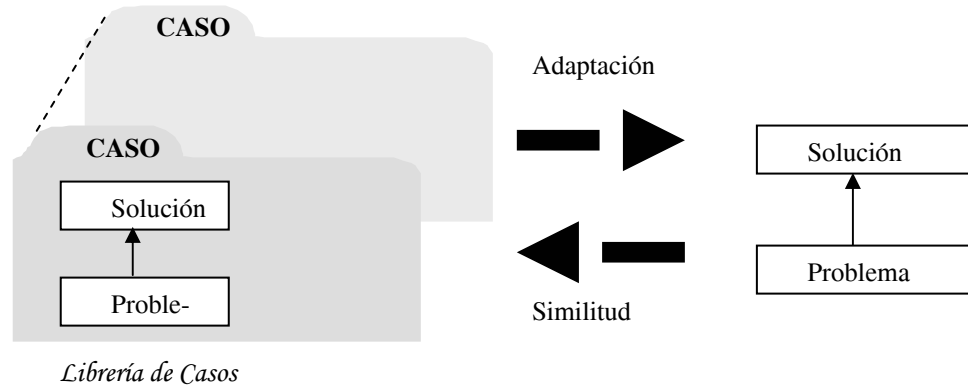
³ Decanato, Universidad Autónoma de Aguascalientes, México, CP. 20280

Resumen. El objetivo de este artículo es ofrecer una visión general de fácil comprensión de la tecnología del RBC y también de instrumentos para que el lector pueda aplicar la tecnología en la práctica. Hoy en día no existe literatura suficiente sobre técnicas y métodos de RBC, a pesar de la popularidad que el tema ha venido obteniendo. La falta de literatura sobre RBC debe, en parte, al hecho de que el RBC, al contrario de la mayoría de los métodos y técnicas de informática tradicionalmente desarrolladas y lanzados al mundo en los Estados Unidos, alcanzó su “madurez” como tecnología de gestión y manipulación de conocimiento en la primera mitad de la década de 1990 en algunos grupos de investigación en Europa, principalmente en Alemania. Debido a esto, muchos libros y publicaciones técnicas avanzadas sobre RBC fueron editados en Inglés o Alemán y no son accesibles al público en general. Con este artículo, se pretende enmendar esa laguna. Aunado a una aproximación al Modelado de Sociedades desde la perspectiva de la Minería de Datos Social.

Keywords: Razonamiento basado en Casos, Minería de Datos, Modelado de Sociedades.

1. Introducción

RBC es un nuevo paradigma para la resolución de problemas que en muchos aspectos difiere en forma fundamental de otros enfoques de la Inteligencia Artificial [AP94]. En vez de basarse únicamente en el conocimiento generalizado acerca de un dominio de problemas o de realizar asociaciones lógicas a lo largo de relaciones abstractas entre los descriptores de problemas y conclusiones, RBC es capaz de utilizar el conocimiento específico de soluciones de problemas concretos, experimentadas anteriormente, denotadas como casos.



Librería de Casos
Figura 1 Modelo básico del enfoque RBC
Elementos básicos de RBC.

Los elementos básicos de un sistema de RBC son (ver figura 1):

- **Representación del Conocimiento:** En un sistema de RBC, el conocimiento es representado principalmente en forma de casos que describan experiencias concretas. En tanto, si fuera necesario, también otros tipos de conocimiento sobre el dominio de aplicación pueden ser almacenados en un sistema de RBC (por ejemplo, casos abstractos y generalizados, tipos de datos, modelos de objetos usados como información).
- **Medida de Similitud:** Tenemos que ser capaces de encontrar un caso relevante para el problema actual. Sistemas de RBC avanzados tienen mecanismos y conocimiento para adaptar los casos recuperados completamente, para verificar si satisfacen las características de la situación presente.
- **Adaptación:** Situaciones pasadas representadas como casos difícilmente serán idénticas a las del problema actual. Sistemas de RBC avanzados tienen mecanismos y conocimiento para adaptar los casos recuperados completamente, para verificar si satisfacen las características de la situación presente.
- **Aprendizaje:** Para que un sistema se mantenga actualizado y se desarrolle continuamente, siempre que el resuelva un problema con éxito, deberá ser capaz de recordar esa situación en un futuro como una característica de un nuevo caso.

Representación del Conocimiento.

La representación del conocimiento es un aspecto tan importante del RBC que a esta le será dedicada una sesión especial dentro de este artículo.

La forma principal de representación de conocimiento en un sistema de RBC son los casos. Un caso es una pieza de conocimiento contextualizado que registra un episodio en que un problema o una situación problemática fué totalmente o parcialmente resuelta.

Un caso puede también contener otros ítems, como los efectos de la aplicación de la solución o la justificativa para aquella situación y su respectiva explicación [Kol93]. Puede también ser enriquecido por datos administrativos, como el número del caso, y la fecha de su creación o el nombre del ingeniero de conocimiento que lo incorpore a la librería.

Los casos contienen primordialmente experiencias concretas, vividas en una situación específica. No en tanto, podemos también crear casos abstractos, que realizan la hipótesis de asumir las experiencias adquiridas en un conjunto de situaciones [Ber96].

2. ¿Cómo funciona RBC?

Cuando un nuevo problema es encontrado, RBC recuerda casos similares y adapta las soluciones que funcionaron en el pasado al problema actual (ver figura 2). Subyacente a este enfoque está la suposición de que problemas cuya descripción posea forma similar presentan soluciones similares. Consecuentemente soluciones de problemas previos similares al actual son un punto de partida útil para las soluciones de un nuevo problema.

El Ciclo de RBC

El modelo más aceptado para el proceso de RBC, es el **Ciclo de RBC** propuesto por Aamondt y Plaza [AP94], que engloba un ciclo de razonamiento continuo compuesto por cuatro tareas principales.

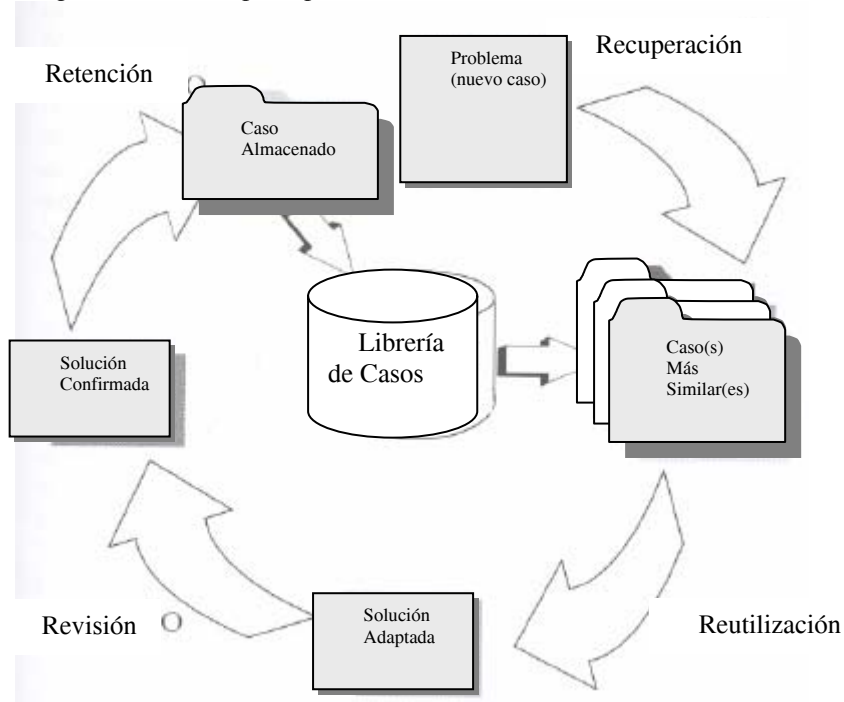


Figura 2 Ciclo del Razonamiento Basado en Casos [AP94].

Recuperar lo(s) caso(s) más similar(es) de la librería de casos;

Reutilizar este(s) caso(s) para resolver el problema;

Revisar la solución propuesta; y

Retener la experiencia representando el caso actual (o partes de esta experiencia) para la reutilización futura.

2.1 Componentes básicos de un caso.

Un caso es compuesto típicamente de los siguientes componentes principales:

- El *problema* que describe el estado del mundo cuando el caso ocurre;
- La *solución* que postula la solución derivada para aquel problema. Una solución puede también ser una acción, un plan o una información útil al usuario.

La *descripción del problema* describe el problema que fue resuelto o la situación que fue comprendida. Esto puede incluir, por ejemplo, el objetivo del razonador, la respectiva tarea que debería ser apoyada, bien como cualquier característica de ambiente que tiene alguna participación en la solución escogida o en el resultado final alcanzado. La descripción del nuevo problema a ser resuelto es situada en el *espacio del problema*. La *recuperación* es el proceso que identifica el caso con la descripción del problema más similar (la flecha rotulada R en la figura 4.1.) y su solución almacenada es encontrada. La solución cuenta al razonador como un problema fue resuelto en una situación particular. Esto puede registrar la solución en sí o el protocolo de cómo la solución fue derivada [Vel96,Car86] y/o las razones asociadas con cada parte de la solución. Si es necesario, la *adaptación* puede ocurrir (la flecha rotulada A) y una nueva solución es creada. Este modelo conceptual de RBC presupone que existe un mapeado directo uno-para-uno entre los espacios del problema y de la solución.

3. La metodología detrás del Razonamiento basado en Casos

El objetivo del Razonamiento Basado en Casos es la utilización de soluciones conocidas en el contexto de un problema nuevo, de la solución también desconocida. En función de esto, la determinación de ejemplos de caso adecuados, que no precisan necesariamente ser idénticos a la situación actual, es uno de los problemas centrales de esta técnica de la Inteligencia Artificial.

El concepto de utilidad de casos es, desde un punto de vista abstracto, central para el Razonamiento Basado en Casos. Durante la recuperación de casos se busca por un caso en la librería de casos, que en el contexto de descripción del problema actual, es útil para determinar su solución.

Es difícil definir exactamente lo que significa que sea útil un caso para solucionar un determinado problema, como la utilidad debería ser la medida o lo que retorna un determinado caso, el cuál es el más apropiado para solucionar un problema del que puede solucionar otro. En especial, la adecuación de un caso para solucionar un determinado problema puede ser determinada con certeza solamente después que se intento elaborar una solución para este problema con base en el caso seleccionado. Esta forma de determinación de utilidad, sin embargo, en que un sistema construye soluciones hipotéticas a partir de todos los casos almacenados en la librería de casos y después determina cuál es el mejor para aplicarla, es impráctico. Por tanto, se necesita

de un método heurístico para analizar un caso y juzgar su utilidad sin realizar todo el trabajo de intentar construir una solución a partir de este e intentar aplicar esta solución, hasta encontrarla porque en muchos casos, como en el diagnóstico, esto no es posible de ser realizado.

3.1 Recuperación de Casos.

El objetivo de la recuperación de casos es encontrar un caso o un pequeño conjunto de casos en la librería de caso que contengan una solución útil para el problema o la situación actual. Por ejemplo, dada la descripción de un problema ocurrido en una impresora, un sistema de RBC debería ser capaz de recuperar un caso describiendo una solución apropiada al problema (por ejemplo, *cambiar el cartucho de tinta*).

Para realizar esa recuperación, es necesario identificar la descripción del problema actual para analizarlo con los problemas almacenados en la librería de casos, aplicando una medida de similitud. Varios enfoques diferentes son aplicados para la realización del proceso de correspondencia consulta-casos de la librería para recuperar un conjunto de casos útiles. En este artículo discutiremos primeramente los principios generales que sustentan la recuperación de casos y después presentaremos en detalle los principales abordajes de la recuperación.

El proceso de recuperación de casos puede ser formalmente descrito por medio de un conjunto de sub tareas que deben ser realizadas por el sistema de RBC:

- *asesoramiento de la situación* enfocándose a la formulación de una consulta representada por un conjunto de descriptores relevantes de la situación o del problema actuales;
- *matching*, enfocándose a la identificación de un conjunto de casos suficientemente similares a la consulta;
- *selección*, que selecciona el mejor matching o matchings con base en el conjunto de casos seleccionado.

3.2. Reutilización de Casos

Una vez que un caso adecuado es recuperado de la librería de casos, la solución sugerida por este caso es objeto de una tentativa de reutilización para la solución del problema actual. Durante este paso, se da una reutilización del conocimiento (la descripción de la solución) del caso anterior, conocido, para el caso actual, sin embargo este todavía no ha sido solucionado.

Un sistema de RBC debe ser capaz de aplicar el problema actual a la solución descrita para un problema similar recuperado de la Librería de Casos. Llamamos a este paso de **reutilización**. La reutilización consiste principalmente de **adaptación** de la solución del caso anterior al caso actual, y las técnicas tratadas en la reutilización de casos intentan resolver los problemas envueltos en la adaptación de casos, que son: cuáles aspectos de la situación deben ser adaptados, cuáles modificaciones deben ser realizadas para esta adaptación y como controlar este proceso. En el transcurso de este artículo, analizaremos estos problemas y veremos técnicas para su solución.

Una vez que el caso recuperado de la Librería puede no satisfacer completamente los requisitos dados por la nueva situación, se puede tornar necesario adaptar la solución descrita en el caso recuperado antes de la aplicación al caso actual. En la mayoría de las circunstancias y dominios de aplicación de RBC, no entonces, generalmente será suficiente que se copie la solución del caso encontrado para el caso actual y se aplique esta solución, o en tanto que se adapte el caso manualmente.

En función de esto, una estrategia frecuentemente utilizada para construir sistemas de RBC es realizar un contorno del problema de adaptación enteramente [Lea96], o que puede ser perfectamente adecuado para tareas analíticas como clasificación, diagnóstico o soporte de decisión. En vez de intentar adaptar casos recuperados a la nueva situación, la estrategia más simple es la de inflar la librería de caso al máximo, para garantizar que todo problema posible pueda poseer en la Librería un caso cuya solución sea lo suficientemente similar, de manera que la solución no es necesaria que sea adaptada. Aplicaciones de este tipo muchas veces son utilizadas desde una filosofía de interacción con el usuario en el que el sistema sugiere una solución o conjunto de soluciones potenciales y un conjunto de herramientas para navegación para esta solución, de manera que permita al usuario explorar las respuestas sobre todos los aspectos y reutilizar la más adecuada a sus necesidades, modificando manualmente una solución, se considera necesario. Esta estrategia ofrece, por ejemplo, una forma simple de implementarse, para realizar un contorno del problema de adaptación que pudiera surgir en nuestro ejemplo de sistema *on-line* de la agencia de viajes: el usuario recibe un conjunto de paquetes de viaje como respuesta y puede navegar por cada paquete y manualmente escoger otro hotel u otra compañía aérea a partir de una lista de opciones válidas para cada paquete.

3.3. Revisión de Casos.

Cuando una solución para un caso generada en la fase de reusabilidad no es correcta, surge una oportunidad para el aprendizaje a partir de esta falla. Esto ocurre en la fase del RBC llamada de **revisión**.

La revisión consiste de dos tareas:

1 Avalar con un criterio cuidadoso la solución generada por la reutilización. Si fuera considerada como correcta, aprenda con el éxito y continúe con la **retención** del nuevo caso en la Librería de Casos.

2 En caso contrario, reparar la solución para el caso, utilizando conocimiento específico sobre el dominio de aplicación o información suministrada por el usuario.

3.4. Retención de nuevos casos.

Retención de los casos es el proceso de incorporación, al conocimiento ya existente, de aquello que es útil de un nuevo episodio de solución de un problema. El objetivo de que se retenga continuamente el conocimiento toda vez que un nuevo problema es resuelto es el de constantemente actualizar y extender la librería de casos. Esto permite a un sistema de RBC continuamente incrementar su conocimiento y convertirse en un solucionador de problemas más poderoso, con el pasar del tiempo es su

utilización. La retención automática de nuevo conocimiento en RBC, cuando este sea además de simple almacenamiento de un caso cuya solución fue confirmada como correcta, utilizando técnicas de **Aprendizaje de Máquina**. Por lo tanto se debe de analizar las variadas filosofías de retención y ver en detalle las técnicas de Aprendizaje de Máquina comúnmente utilizadas en RBC, principalmente el Aprendizaje Basado en Instancias, analizando a fondo, en algunos casos, los algoritmos utilizados.

3.5. Aplicaciones del RBC

El RBC también ha sido usado para la creación de numerosas aplicaciones en un espectro bastante extenso de dominios de aplicación, incluyendo análisis financiero, asesoramiento de riesgos, mantenimiento técnico, control de procesos, control de calidad, diagnóstico médico, sistemas de soporte al software, previsión, planeación, diseño de proyectos, clasificación, interpretación de imágenes, avalar la compra inmobiliaria, comercio electrónico, soporte al usuario, gestión del conocimiento, entre otros [AAB⁺95, ABS96, BBG⁺99].

Para proveer una visión general de la aplicación de sistemas de RBC, haremos una distinción, aquí, entre el dominio de aplicación y el tipo de tarea a ser ejecutada por el sistema.

Una tarea de un sistema de RBC describe el tipo de acción para la cuál el sistema será utilizado, como, por ejemplo, diagnóstico, configuración, planeación entre otras. Esto determinará el tipo de problemas y de soluciones, bien con la naturaleza de las actividades a ser desarrolladas por el solucionador de problemas basado en casos. Como discutimos anteriormente, las tareas pueden ser clasificadas en **tareas sintéticas** y **tareas analíticas**.

Las Tareas Analíticas cubren una amplia gama de aplicaciones que comparten determinadas características. Generalmente un nuevo caso es comparado de aquellos en la Librería de Casos para determinar en que tipo de categoría o clase pertenecen. La solución asociada al caso más similar dentro de la clase correspondiente es entonces presentado [Wat97]. Tareas analíticas de sistemas de RBC típicas son [Ric98]:

- ❖ Clasificación, Diagnóstico, Soporte de Decisiones & Tutorías.

Las tareas de síntesis, por otro lado. Intentan crear una nueva solución por medio de la combinación de partes de soluciones previas [Wat97]. Ejemplos de Tareas sintéticas son [Ric98]:

- ❖ Configuración, Planeación & Diseño de proyectos.

En la práctica, la mayoría de los sistemas comerciales de RBC disponibles soporta solamente tareas analíticas y es dedicada primordialmente a la recuperación de casos.

El dominio de aplicación de un sistema de RBC es en el área en la cuál el sistema es aplicado, por ejemplo, medicina, arquitectura, administración, finanzas o ingeniería mecánica. Cada dominio posee características propias, que influyen fuertemente al momento de escoger la forma de representación del conocimiento a ser utilizada.

CBR-Works es uno de dos shells de RBC más poderosos disponibles actualmente. Es un paquete de desarrollo de software adaptado para la creación de soluciones inteligentes en una gran variedad de dominios de aplicación y ambientes operacionales. Los editores gráficos de CBR-Works soportan el desarrollo confortable de los mode-

los de conocimiento más complejos. Conceptos, tipos, medidas de similitud y filtros, todo puede ser creado y mantenido de una forma fácil e intuitiva.

Sus características básicas son:

- Tipos predefinidos y una biblioteca de módulos de aplicación.
- Asistentes de modelaje para tipos y conceptos.
- Pesos asociables a los atributos expresan la importancia de características de conceptos.
- Medidas de similitud a los patrones en forma predefinida.
- Editores gráficos de similitud fáciles de usar.
- Acceso a las bases de datos vía ODBC.
- Interfaces para Internet genéricas sin la necesidad de programarse o proyectar interfaces online.

CBR-Works existe para todos los principales sistemas operativos: Windows, Linux, UNIX (Solaris, AIX, IRIX, HP-UX, entre otros) y MacOS.

4. Aplicación del RBC en otras Técnicas de Inteligencia Artificial

Una de las áreas en donde el RBC se ha complementado adecuadamente es con la Minería de Datos (Data Mining) y en otro idéntico sentido con aplicaciones enfocadas al Procesamiento de Lenguaje Natural específicamente adecuación y complementación de idiomas. Aunado a esto el RBC puede aplicar modelos de predicción para la Migración utilizada en los Algoritmos Genéticos, y puede ayudar a optimizar procesos de Holografía y Aplicaciones Inteligentes para Láser.

El Procesamiento del Lenguaje Natural (PLN) es una de las piedras angulares tempranas de la Inteligencia Artificial (IA). La Traducción Automática, por ejemplo, nació a finales de la década de los cuarenta, antes de que se acuñara la propia expresión «Inteligencia Artificial». No obstante, el PLN ha desempeñado múltiples papeles en el contexto de la IA, y su importancia dentro de este campo ha crecido y decrecido a consecuencia de cambios tecnológicos y científicos. Los primeros intentos de traducir textos por ordenador a finales de los cuarenta y durante los cincuenta fracasaron debido a la escasa potencia de los ordenadores y a la escasa sofisticación lingüística.

Los testores, y en particular los testores típicos son una herramienta utilizada en reconocimiento de patrones (RP), particularmente en problemas de selección de variables, así como también en modelos de clasificación supervisada basados en precedencias parciales. Estos, se dividen con base a la forma en que determinan la similitud entre dos patrones, clasificándose en algoritmos de escala interior y algoritmos de escala exterior, entre los cuales se encuentra BT. El algoritmo BT entre otras aplicaciones, es el que utiliza el sistema PROGNOSIS; no obstante, aunque durante mucho tiempo ha sido considerado como el paradigma de los algoritmos de escala exterior, su desempeño computacional no ha sido totalmente conveniente dados los altos tiempos de ejecución que ha tenido [SSF94]. El cálculo del conjunto de testores típicos, es un problema de complejidad exponencial, por lo que se justifican los esfuerzos por diseñar e implementar nuevos algoritmos (o modificaciones de los existentes) que reduzcan el espacio de búsqueda de la solución, y lo hagan además en un tiempo menor al que actualmente se requiere.

5. Una aproximación al Modelado de Sociedades utilizando Minería de Datos

Las aproximaciones en Minería de Datos para aspectos sociales, buscan situaciones análogas en registros de comportamiento [PGO05]. Los investigadores buscan situaciones donde los grupos de gente están produciendo registros de cómputo (Indicadores de Población, ejemplo: INEGI) como parte de su actividad normal. La información potencialmente útil implícita en estos expedientes se identifica, las técnicas de cómputo para presentar los resultados son diseñadas. Así el cómputo descubre y hace explícito las "tendencias sociales a través del tiempo" creadas por un particular tipo de comunidad [OIM06].

Los sistemas para analizar aspectos sociales con Minería de Datos no requieren a usuarios expertos en ninguna nueva actividad, debido a esto, los investigadores en el tema intentan explotar la información de la preferencia de los usuarios implícita en registros de la actividad existente [OTI⁺05].

Los sistemas para Minería de Datos Social permiten analizar el comportamiento de sociedades [TTO⁺05]. Realizan esto minando y redistribuyendo la información en expedientes de cómputo de la actividad social tales como mensajes de Usenet, bitácoras de sistemas, registros de compra, ligas de interés, listas de canciones en Internet [OMM⁺06]. Por lo tanto se generan dos preguntas generales para evaluar el funcionamiento de tales sistemas: (1) ¿La información extraída es de valor?, y (2) ¿Es posible determinar si un grupo de personas separadas físicamente muestran formas de pensar similares en gustos y capacidades? (ver figura 3).

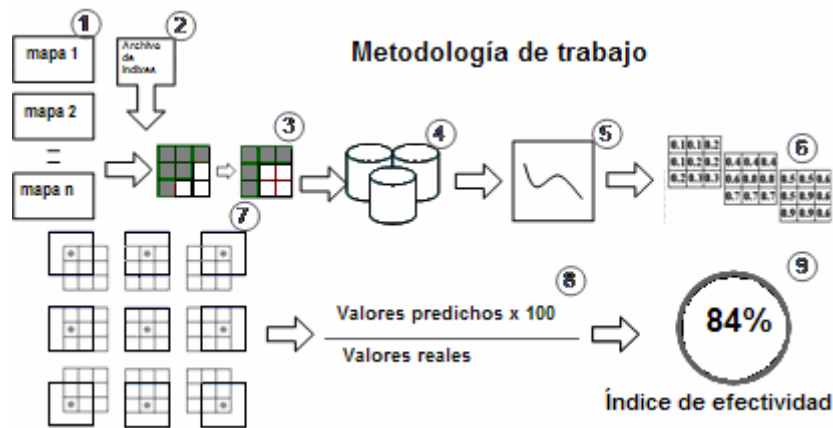


Figura 3.- Metodología propuesta para un modelo de tipo social.

6. Reconocimientos

El primer autor desea agradecer a todos aquellos que participaron, directa o indirectamente, en la elaboración del libro en el cuál nos basamos para realizar este artículo, en especial al Professor Dr. Wolfram Pietsch, del grupo de Sistemas Inteligentes de la Universidad de Ciencias Aplicadas de Aachen, a los investigadores y emprendedores del RBC y el Modelado de Sociedades utilizando Minería de Datos Social: Halina Ivanovic, Daniel Alibabić, Hakkel Žvičy, Harak Murakhaev, Đragoljub Purlija, Fuad Backovic y Lena Phillippson y por fin a los amigos científicos José Alberto Hernández, Miguel Basurto, Qrel Dox, Jiani Cardoso, Miguel Martín, Alice Jetkins-Makalawe, Francisco Álvarez, Javier Martínez, Gennadiy Burlak, Amol Waghlikar, Tasmia Mallor, Lar Gand, Salú Digby, Taardemi Suaromi y Xao Jin, por discusiones fructíferas y opiniones constructivas. A los Drs. Jaime Muñoz y José Alberto Hernández por colaborar en el desarrollo del Capítulo 12

Referencias

- Ochoa A., Hernández A. & Muñoz J. (2006) “Más allá del Razonamiento Basado en Casos y una Aproximación al Modelado de Sociedades utilizando Minería de Datos”, Universidad Autónoma de Aguascalientes Editorial, (Edición Aceptada) Noviembre 2006.