Universitat de Girona
Escola Politècnica Superior

Treball Final de Carrera

Estudi: Enginyeria Informàtica

Títol: RAONAMENT BASAT EN CASOS APLICAT AL COMERÇ ELECTRÒNIC

Document: MEMÒRIA

Alumne: Antoni Martínez Arteaga

Tutor: Josep Lluís de la Rosa
Departament: Electrònica, Informàtica i Automàtica

Convocatòria: Gener de 2001

Taula de Continguts

1 REQUERIMENTS I OBJECTIUS DEL PROJECTE 4

1.1 REQUERIMENTS DEL PROJECTE 4

1.2 OBJECTIUS DEL PROJECTE 5

2 INTRODUCCIÓ 6

2.1 DESCRIPCIÓ GENERAL 6

2.2 PER QUE APLICAR EL CBR AL COMERÇ ELECTRÒNIC 7

2.3 HIPÒTESI 8

2.4 ESTRUCTURA DEL PROJECTE 8

3 EL RAONAMENT BASAT EN CASOS 11

3.1 INTRODUCCIÓ AL RAONAMENT BASAT EN CASOS 11

3.1.1 Concepte de cas 13

3.2 ESTRUCTURA GENERAL I CICLE DE VIDA DEL CBR 14

3.2.1 Procés de Retrieve 17

3.2.2 Procés de Reuse 20

3.2.3 Procés de Revise 21

3.2.4 Procés de Retain 22

3.3 ANALOGIA ENTRE ELS SISTEMES EXPERTS I EL CBR 23

4 CONSTRUCCIÓ D'UNA BOTIGA VIRTUAL AMB E-MERCHANDISING 24

4.1 DESCRIPCIÓ DEL SISTEMA 24

4.1.1 La botiga virtual 24

4.1.2 El simulador de la botiga virtual 25

4.2 DESENVOLUPAMENT DEL PROJECTE 26

4.3 TRADUCCIÓ DE LA LLISTA DE PRODUCTES A UN CAS 27

4.3.1 Codi de barres dels productes 28

4.4 DEFINICIÓ DEL DOMINI DEL MODEL DE DADES DEL CBR 29

4.4.1 Definició de les classes del Model de Dades 29

4.4.2 Definició dels atributs del cas 30

4.5 DEFINICIÓ DE LA FUNCIÓ SIMILITUD ENTRE CASOS 31

4.6 DEFINICIÓ DELS PESOS DELS ATRIBUTS DEL CAS 32

4.7 PRODUCTES I SERVEIS OFERTATS PER LA BOTIGA 33

4.8 NOVA ETAPA D'EXCITACIÓ DEL CICLE DE VIDA DEL CBR 34

4.9 FUNCIÓ D'AVALUACIÓ D'USUARI 36

4.10 QUANTIFICACIÓ DE LA SATISFACCIÓ DE L'USUARI 37

4.10.1 Mitja del les similituds del sistema 38

4.10.2 Bondat del Sistema *38*
4.10.3 Rendiment del Sistema *38*
4.10.4 Rendiment de la Botiga *39*

5 ESTUDI I RESULTATS DEL SISTEMA 40

5.1 MOTOR DE SIMULACIÓ 40
5.2 CONSTRUCCIÓ DELS PERFILS D'USUARIS 45
5.3 EVOLUCIÓ DELS VALORS D'UN CAS 46
5.4 REPRESENTACIÓ DE RESULTATS 47
5.5 RENDIMENT DEL SISTEMA 49
5.6 RELAXACIÓ I SOBREEXCITACIÓ DEL SISTEMA 54
5.7 LA BOTIGA VIRTUAL 57

6 CONCLUSIONS 60

ANNEX A – CONTINGUT DEL CD-ROM 63

1 Requeriments i objectius del projecte

1.1 Requeriments del projecte

Fins ara en el comerç tradicional s'han fet grans esforços per donar al client un servei amb qualitat, aconsellant-li la compra d'un producte que s'adapti més amb les seves necessitats, però en canvi avui en dia el comerç electrònic té una mancança molt important que és l'atenció personalitzada al client.

Els requeriments explícits del projecte són:

- Construcció d'un producte nou i inexistent en el mercat per tal de respondre a la demanda dels negocis el comerç electrònic per donar una atenció personalitzada al client.

- Adaptar una eina relativament innovadora dins del camp de la intel·ligència artificial per tal de resoldre aquest problema: El Raonament Basat en Casos (*Case Base Reasoning* ó CBR).

- Utilització del paquet de funcions del JavaKate que ens ajudarà en la tasca d'implementació de la tecnologia del Raonament Basat en Casos aplicada al comerç electrònic.

- Implementació d'un sistema que a partir de la llista de la compra de diversos articles que realitza un client en una botiga sigui capaç de proposar altres articles, productes o serveis a aquest client amb una determinada certesa que el client el compri.

- El sistema ha de permetre que el número d'articles a la venda sigui totalment variable i que és pugui canviar amb gran facilitat el número d'articles ofertats.

- El sistema no pot conèixer cap dada personal de la persona que està comprant i per tant no pot saber les compres que ha realitzat aquest client anteriorment.

- La implantació del sistema ha de ser la més directe possible.

1.2 Objectius del projecte

Els objectius específics del projecte són els següents:

- Estudiar la viabilitat, característiques i potència d'aplicació del Raonament Basat en Casos en el comerç electrònic.

- Construir una botiga virtual que a partir de la llista de la compra que realitza un client la botiga sigui capaç de proposar altres productes o serveis, amb una determinada certesa que aquest producte o servei sigui adquirit pel client.

- Construir un simulador de la botiga virtual que mostri el funcionament intern del Raonament Basat en casos i per tant el funcionament intern de la botiga virtual.

- Construir un espai web de lliure accés on es pugui realitzar les operacions de compra en la botiga virtual i a on es pugui executar remotament el simulador.

2 Introducció

2.1 Descripció General

El Raonament Basat en Casos (*Case Base Reasoning ó CBR*) es un nou paradigma de la intel·ligència artificial. El raonament basat en casos és un mètode per resoldre problemes mitjançant el record de situacions similars prèvies a la actual i reutilitzant el coneixement de la informació de l'experiència que ens aporta l'antiga situació a la nova situació: "Problemes similars tenen solucions similars".

La part més important del projecte per als usuaris és la botiga intel·ligent a on es pot comprar una sèrie de productes i després la botiga fa una proposta d'un producte que creu, amb una certa certesa l'usuari comprarà. A partir de la llista de productes de la compra que realitza el client aquesta botiga virtual és capaç d'extreure el perfil d'aquest client. I a partir d'aquest perfil que identifica el tipus de client, la botiga li pot ofertar productes que tinguin unes grans possibilitats de ser adquirits. S'ha implementat una botiga virtual on l'usuari realitza una compra de productes alimentaris i la botiga és capaç de proposar altres productes alimentaris, a més d'aconsellar entrades de cinemes, teatre, i a més recomanar sopars en restaurants i inclòs et pot presentar ofertes de viatges amb una certesa que aquest productes o serveis seran adquirits.

La part més important del projecte per al personal docent, alumnat, investigador i professional és la possibilitat de veure interactivament el funcionament intern del Raonament Basat en Casos de la botiga intel·ligent. Aquesta part és el simulador de la botiga, on després d'establir una sèrie de paràmetres de simulació, podem seguir pas a pas el raonament de la botiga i consultar el rendiment de la botiga amb links a fitxer de dades. Una gran diferència amb els sistemes actuals, és que en aquest projecte disposem de mesures de rendiment del sistema i del simulador de compra-venda de productes que ens aporta uns resultats amb les característiques i mancances d'aplicar el Raonament Basat en Casos al comerç electrònic. Un cop ja s'hagi treballat amb la botiga i es conegui el funcionament, existeix la possibilitat d'interactuar manualment amb aquesta i fer consulta del tipus del 'veí més proper (*Nearest Neighbour*)', fer la cerca en arbre amb la consulta inductiva (*Induction consultation*), i per últim afegir un cas manualment a la base de casos.

Tenim doncs, en les nostres mans un treball de recerca que serà útil en centres docents, per estudiar de forma interactiva un nova tecnologia del camp de la intel·ligència artificial. I ens serà útil també, com a estudi de viabilitat, per aquelles empreses o persones que vulguin aplicar o adaptar la mateixa filosofia del raonament basat en casos al seu espai de telecomunicacions.

Per a la construcció d'aquest projecte s'ha fet servir la darrera tecnologia de programació orientada a Internet: Servlets, Java Server Pages, a més del clàssic Java i HTML. Per a la implementació del Raonament Basat en Casos s'ha treballat amb un dels millors paquets existents en el mercat d'aquesta tecnologia, el paquet de funcions del JavaKate.

2.2 Per que aplicar el CBR al comerç electrònic

En aquest moments el comerç electrònic pateix un gran dèficit, cosa que no succeeix en situacions de comerç normal, és el suport intel·ligent o bé assistència personalitzada al client. És evident que la selecció de productes i serveis quan es navega per un espai complex d'ofertes pot resultar una tasca que porti a l'usuari a no trobar exactament el que necessita o bé que no quedi prou satisfet amb la utilització del comerç electrònic. En aquest moments no existeix cap portal en Internet que sigui capaç de guiar, aconsellar o donar suport a la compra de productes o serveis a un client.

Si això s'afegeix la dificultat de garantir les transaccions electròniques, que el disseny de l'interface dels catàlegs on-line encara és un 'art' molt jove i inclòs està en la seva fase experimental, resulta que molts usuaris no accepten totalment el mitjà ja sigui per una raó o una altre. Amb tot això podem arribar al punt en que el comerç electrònic pot no evolucionar com actualment s'especula. Es per això que el raonament basat en casos ens pot donar una empenta més al èxit del comerç electrònic.

El Raonament basat en casos està començant a aplicar-se amb èxit en un gran ventall d'aplicacions com poden ser anàlisis financer, manteniment tècnic, control de processos, control de qualitat, diagnosi mèdic. Inclòs hi ha aplicacions a on altres tècniques de intel·ligència artificial han fracassat i el Raonament Basat en Casos ha donat un resultat satisfactori.

Ara és el moment a on s'estan dedicant grans esforços per aplicar aquesta tecnologia en el comerç electrònic. Podem pensar que degut al moviment tant dinàmic que té les telecomunicacions pot ser d'aquí uns anys surti una altre tecnologia o bé un altre sistema que sigui millor per ser aplicat al comerç electrònic. Però actualment el CBR és del millor camí que hi ha per arribar a un bon punt de rendiment.

Cal destacar empreses com *Acknosoft* una de les més importants dedicada a l'aplicació del Raonament Basat en Casos al món industrial, s'ha reestructurat i ha canviat de nom per *Kaidara* i actualment es dedica íntegrament a les vendes automàtiques i al suport de clients en Internet.

2.3 Hipòtesi

Preguntant a una gran base de dades d'observacions de casos d'anteriors compres i donat un nou cas de compra d'un nou usuari podem trobar la similitud entre el perfil d'usuari del cas anterior i el perfil del nou usuari, i com a conseqüència d'aquesta similitud podem oferir i donar un suport o servei personalitzat al nou usuari amb una certesa que el nou usuari ens acceptarà la nostra proposta, i tot això sense tenir les dades de les persones físiques en un suport informàtic.

2.4 Estructura del projecte

Podem enfocar el projecte des de dos punts de mira, el vessant docent i l'empresarial. Segons l'interès del lector pot dedicar més temps a uns temes que a uns altres. El primer punt en comú és la introducció al projecte, a on es presenten els requeriments i s'especifiquen els objectius.

En el tercer tema s'explica la tecnologia del raonament basat en casos. Tot el que hi ha explicat en aquest tema és una introducció teòrica sobre el raonament basat en casos i serveix per entendre el treball realitzat en aquest projecte. És imprescindible que si el lector que no coneix la tecnologia del raonament basat en casos llegeixi aquest tema. Per als lectors que ja tinguin alguns coneixements i desitgin llegir el tema hi podran trobar algunes impressions i comentaris personals sobre el raonament basat en casos. És un tema bàsicament orientat a tots aquells que desitgin treballar a fons i dels que vulguin analitzar i estudiar els resultats del projecte. No és fins al quart i cinquè tema on hi ha exposat tot el treball realitzat en el projecte.

En el quart tema hi ha presentat el tot procés d'anàlisi i d'implementació d'un sistema merchandising. En primer lloc es presenta una descripció global del sistema que s'ha construït. En segon lloc es porta a terme tot l'anàlisi de l'aplicació des del punt de vista del raonament basat en casos. Concretament es parla de tots els paràmetres que cal establir en el CBR, es defineix el domini del model de dades que és requerit per tota aplicació que inclou el raonament basat en casos, s'estableixen la relació entre les classes del model de dades i per últim es configuren els pesos dels eslots. A més de tots aquest passos comuns a tot sistema CBR s'introdueixen nous punts i idees innovadores que son importants tenir-les en consideració per aquelles aplicacions que desitgin aplicar el CBR en el comerç electrònic.

El cinquè capítol està dedicat als experiments i els resultats obtinguts en projecte, que és on es veuen les idees més interessants i és on s'extreuen totes les conclusions del projecte que es presenten en el capítol sisè.

Per la realització del projecte s'ha construït un espai web a on es pot trobar la millor i darrera la informació d'aquest treball. Dintre d'aquest espai podrem trobar aquesta documentació, podrem també anar a la botiga virtual i veure iteractivament el comportament d'aquesta. A més podrem trobar el mateix simulador que s'ha fet servir per fer l'estudi del sistema i veure com es comporta el sistema internament. També tenim la possibilitat de treballar a baix nivell amb la nostra botiga i comprovar el seu funcionament intern.

Aquest espai web és el primer lloc de tot Internet, on es pot veure de manera interactiva tot el procés de raonament d'una botiga virtual intel·ligent amb CBR. Podrem examinar totes les passes que es van produint en el Raonament Basat en Casos per tal d'arribar a una oferta d'un producte o servei. Aprofitant la programació en Java Server Pages podem enviar l'ordre d'execució del simulador al ordinador servidor i que aquest ens retorni la resposta dinàmicament al nostre ordinador. O sigui que al crear la resposta de manera dinàmica, és realment la nostra simulació la que ens és retornada. Aquesta filosofia és la que actualment també s'està treballant en el comerç electrònic, és crear de manera personalitzada planes web per a diferent tipus d'usuaris.

Cal assenyalar que aquest simulador permet al personal docent mostrar el funcionament insitu d'aquesta tecnologia amb els seus avantatges i els seus inconvenients. Al final de cada execució del simulador podrem recollir els resultats de la nostra simulació i de les simulacions anteriors,

els podrem graficar, a més de treballar-los i modificar-los. És per tot això que aquesta documentació només és una de les grans parts que s'afegeixen a tot el conjunt del projecte, que en fan d'aquest un pioner en el seu camp.

Per accedir a l'espai web del projecte es fa referència a la meva plana personal http://eia.udg.es/~tmartinz a on hi ha un link explícit del Raonament Basat en Casos aplicat al comerç electrònic a http://eia.udg.es/~tmartinz/botigaCAT.htm. Aquí es troba la botiga virtual, el simulador de la botiga, amb tot el codi de l'aplicació dels programes en Java en HTML i en Java Server Pages, juntament amb el model de l'aplicació del JavaKate, amb tota la documentació del projecte i per últim s'afegiran tots articles que es vagin publicant sobre el treball i també es publicaran comentaris que aportin les persones i institucions relacionades amb el Raonament Basat en Casos.

Juntament amb la documentació del projecte s'adjunta un CD-ROM que conté exactament el mateix que hi ha en el meu espai web, la documentació creada fins al moment, tot el codi de l'aplicació i el model de l'aplicació del JavaKate.

3 El Raonament Basat en Casos

3.1 Introducció al Raonament Basat en Casos

El Raonament Basat en Casos (*Case Base Reasoning ó CBR*) es un nou paradigma de la intel·ligència artificial. Fins al moment aquesta tecnologia ha sigut aplicada amb èxit en el món industrial i ha arribat a tenir bons resultats a on els sistemes experts no funcionaven.

El raonament basat en casos és un mètode per resoldre problemes, mitjançant el record de situacions similars prèvies a la actual i reutilitzant el coneixement de la informació de l'experiència que ens aporta l'antiga situació a la nova situació.

> *A case-based reasoner solves new problems by adapting solutions that were used to solve old problems.* (Riesbeck and Schank 1989)

Els primers treballs es varen començar als inicis de 1985, però no és fins a finals dels noranta a on ha tingut el seu creixement més espectacular. Actualment és quan més recerca s'està invertint per tant podem preveure que els millors resultats encara estan per venir. Les primeres aplicacions del raonament basat en casos han sigut en el món industrials, i no és fins ara en aquest projecte on acaben de començar a aplicar-se al comerç electrònic.

En aquest capítol es fa una introducció sobre la tecnologia del raonament basat en casos i en cap cas s'intenta donar una formació complerta del raonament basat en casos. Existeixen molts llibres i articles que expliquen amb molta claredat i extensió diversos punts del CBR. Diem punts per que no existeix un sol llibre o una sola publicació a on sigui tot, tot just ara estem treballant per estudiar totes les seves propietats i al cap d'uns anys presentar el CBR com una tecnologia rentable en moltes aplicacions.

Sobre la tecnologia basat en casos existeix molts articles i llibres publicats. És per això que per a tots aquells que vulguin treballar amb el CBR personalment es recomana la següent adreça web, on a més de tota la informació, la trobarà actualitzada, afegint que trobaran calendaris de xerrades i conferències, i amb tot això cal sumar tot un llistat d'exemples d'aplicacions del CBR.

- http://www.cbr-org.com

Recomanen també la lectura dels articles que apareixen en les referències del projecte, [Aamodt – Plaza] com a introducció i [Klaus – Brigitte, 1996] com guia per desarrollar aplicacions CBR.

Basat en la meva experiència en el treball amb el CBR cal advertir a tots els lectors que vulguin entrar a fons en aquesta tecnologia per una possible o imminent aplicació, que de tota la literatura existent triï la que més s'avingui amb els seus objectius, ja que si no pot entrar en temes o punts que no porten cap informació per les seves necessitats.

La idea del CBR és molt bàsica i la majoria d'articles publicats de treballs personal que tracten de les seves experiències o les seves inquietuds que han aparegut al treballar amb el CBR.

Una altre recomanació per al lector que vulgui experimentar amb el CBR i es que al principi no es construeixi ell mateix un sistema de CBR ja que existeixen diversos que son de domini públic o bé son versions de demostració que permeten treballar amb tot el seu potencial. Hi ha diferents programes CBR en el mercat que cadascun té les seves propietats i avantatges respecte a els altres. No tots els sistemes CBR que hi ha en el mercat son iguals, hi ha que tenen un enfoc diferent i permeten fer unes tasques que altres no tenen o que encara no han desenvolupat, és per això que depenent de les necessitats del lector cal fer una tria d'un o altre CBR.

A continuació s'explicarà l'element més important del CBR, el cas. Després s'explicarà el cicle vida del CBR.

3.1.1 Concepte de cas

La noció més bàsica del Raonament Basat en Casos és la noció de cas. Podem dir que un cas consisteix en una tupla amb la descripció del problema més la solució d'aquest. El lloc a on es guarden els casos se'n diu la Base de casos. Una Base de Casos conté tots els problemes que han anat apareixent en el temps més la seva solució. Semblar ser que quan més gran sigui la base de casos més problemes amb la seva solució corresponent tindrem. Dit d'una altre manera el coneixement o intel·ligència que tindrem a l'hora de resoldre problemes depèn del número de casos que tinguem emmagatzemats en la nostra base de casos.

Un cas consta de diferents eslots de diferents tipus. Per una part aquest eslots son atributs que em descriuen el problema o que ens defineixen el problema. Per una altre part tenim els atributs que ens descriuen la solució d'aquest problema. Els casos poden ser representats de diferent manera per diferents persones, depenent de la complexitat del problema. L'enginyeria del coneixement és la que ens dirà quina estructura té un cas i com representa el domini de totes les tuples solució-problema. Una regla molt important a l'hora de definir l'estructura d'un cas és la que diu definir casos amb pocs atributs i senzills.

Kolodner defineix cas com a:

> *A contextualized piece of knowledge representing an experience that teaches a lesson fundamental to achieving the goals of the reasoner*. (Kolodner 1993)

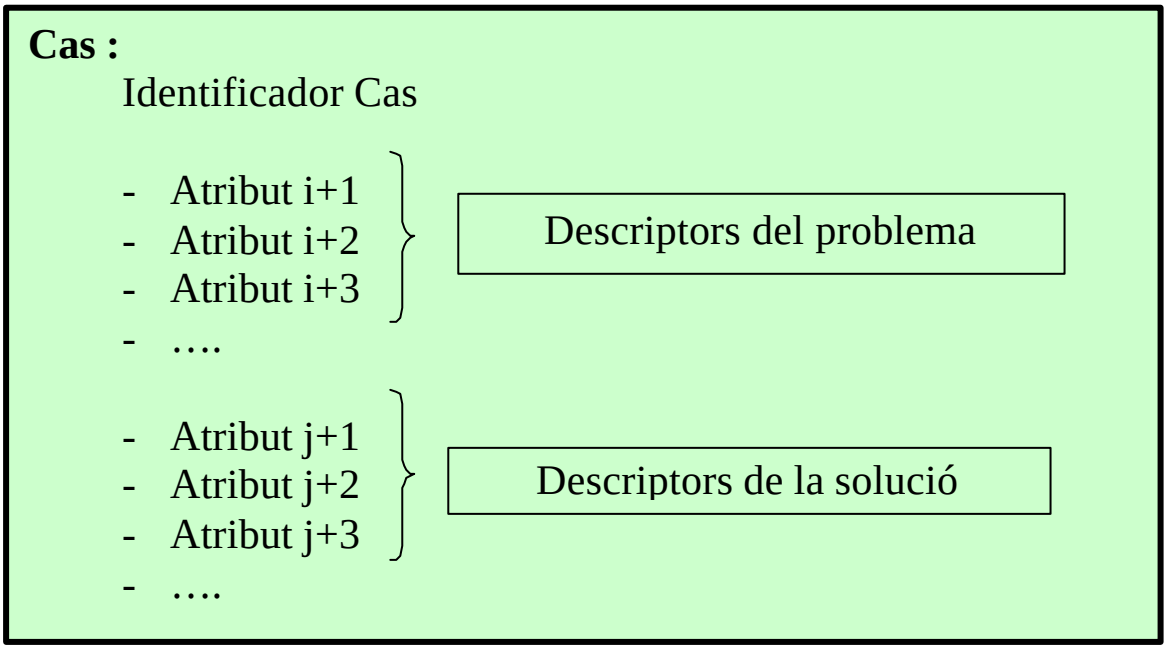

Fig3.1. Estructura d'un cas.

3.2 Estructura general i cicle de vida del CBR

Un sistema CBR enregistra les experiències passades a l'hora de resoldre problemes, això significa que calen algorismes de recuperació dels casos enregistrats amb les seves solucions i adaptar-los a les noves situacions.

Leake (1996) exposa que els cinc problemes principals de la IA que poden ser millorats pel CBR:

- Adquisició del coneixement
- Manteniment del coneixement
- Augmentar l'eficiència de la resolució de problemes
- Augmentar la qualitat de les solucions
- Augmentar l'acceptació de l'usuari.

L'adquisició del coneixement es veu enormement millorat per que en el CBR no cal representar explícitament tot el coneixement, podem establir tota una sèrie de paràmetres o variables com a eslots que sabem que influeixen en el nostre problema. Unes variables poden ser condicionades per unes altres, o be poden ser totalment independents. L'adquisició i representació del coneixement ve simplificat a només fer una tria de les variables que nosaltres creiem que poden ser importants. És possible inclòs que ens deixem alguna variable que identifica el problema, però mitjançant les altres variables podem arribar a determinar la seva influència. La gran riquesa del CBR és que el coneixement el tenim totalment desordenat, però el tenim, i això ens redueix la complexitat problema a trobar algorismes d'ordenació i de recuperació del coneixement.

El manteniment del coneixement del CBR resulta una tasca molt més fàcil que en qualsevol altre tecnologia d'IA, ja que el coneixement és emmagatzemat en una base de dades comuna, per tant el tenim totalment estructurat , el podrem consultar les seves dades, podem aplicar filtres el podem esborrar, modificar i tot el que es pugui fer amb qualsevol base de dades.

Un altre punt que anomena Leake és augmentar l'eficiència de la resolució de problemes, que aquí no és per qüestió de temps de computació dels ordinadors sinó que és per temps de treball dels seus perifèrics.

Els darrer dos punts esmentats per en Leake en el cas del CBR es poden agrupar en un de sol, l'acceptació de l'usuari depèn de la qualitat de les solucions, ja que les solucions proposades per un sistema CBR son directament les enregistrades pel mateix usuari.

Leake també explica que:

> *The world is regular: similar problems have similar solutions. Consequently, solutions or similar prior problems are a useful starting point for new problem-solving.*

Es per això que per qualsevol que vulgui construir una aplicació CBR s'ha de plantejar les següents qüestions: Com és el món de regular ? Què és el que s'ha de recordar ? Qui és el més qualificat per establir la similitud entre els problemes, l'expert o l'operari ?

Hi ha diferents propostes sobre el cicle de vida que segueix el CBR, unes de les més exteses i més acceptades és la que és presenta a continuació donada per Aamodt i Plaza (1994). [Aamodt – Plaza]

L'event que dispara el procés de resolució apareix quan tenim un nou problema. El problema que ens ve donat el traduïm, si es necessari, al que nosaltres hem representat com a cas. Pot ser que un nou problema ens doni directament els valors dels eslots del cas, o bé pot ser necessari fer un fase prèvia d'adaptació de la definició del problema al nostre cas específic.

El primer pas del primer cicle de vida és l'anomenat RETRIEVE, el qual consisteix en recuperar un problema anterior de la Base de Casos que sigui semblant o igual al nou problema. El problema recuperat consta de dues parts, la descripció del problema més la seva la solució. Si recordem la nostra hipòtesis " Problemes similars tenen solucions similars" ens resultarà el segon pas que s'anomena REUSE. Aquest pas és reutilitzar la solució del problema anterior aplicant-la al nou problema. Generalment l'acció d'aplicar una solució al anterior al nou problema no és directa, ens cal un pas que adapti la solució antiga al nou problema, d'aquest pas d'adaptació o de modificació de la solució antiga s'en diu REVISE. En aquest punt tindrem un nou problema amb una solució antiga adaptada o sigui que tenim un nou problema amb una nova solució, aquest serà un nou cas que podria ser afegit a la Base de Casos per poder resoldre altres problemes semblants en el futur. Per saber si aquest nou cas és prou bo com per ser retingut en la Base de Casos s'aplica una funció d'avaluació del cas. Si la funció d'avaluació

accepta el cas es produeix el procés de RETAIN, el qual reté el nou cas en la Base de Casos. Si la funció d'acceptació rebutja el cas, llavors simplement es descarta aquest. Ara ja tenim el cicle de vida complert i el raonament basat en casos es queda a l'espera de l'aparició d'un nou problema.

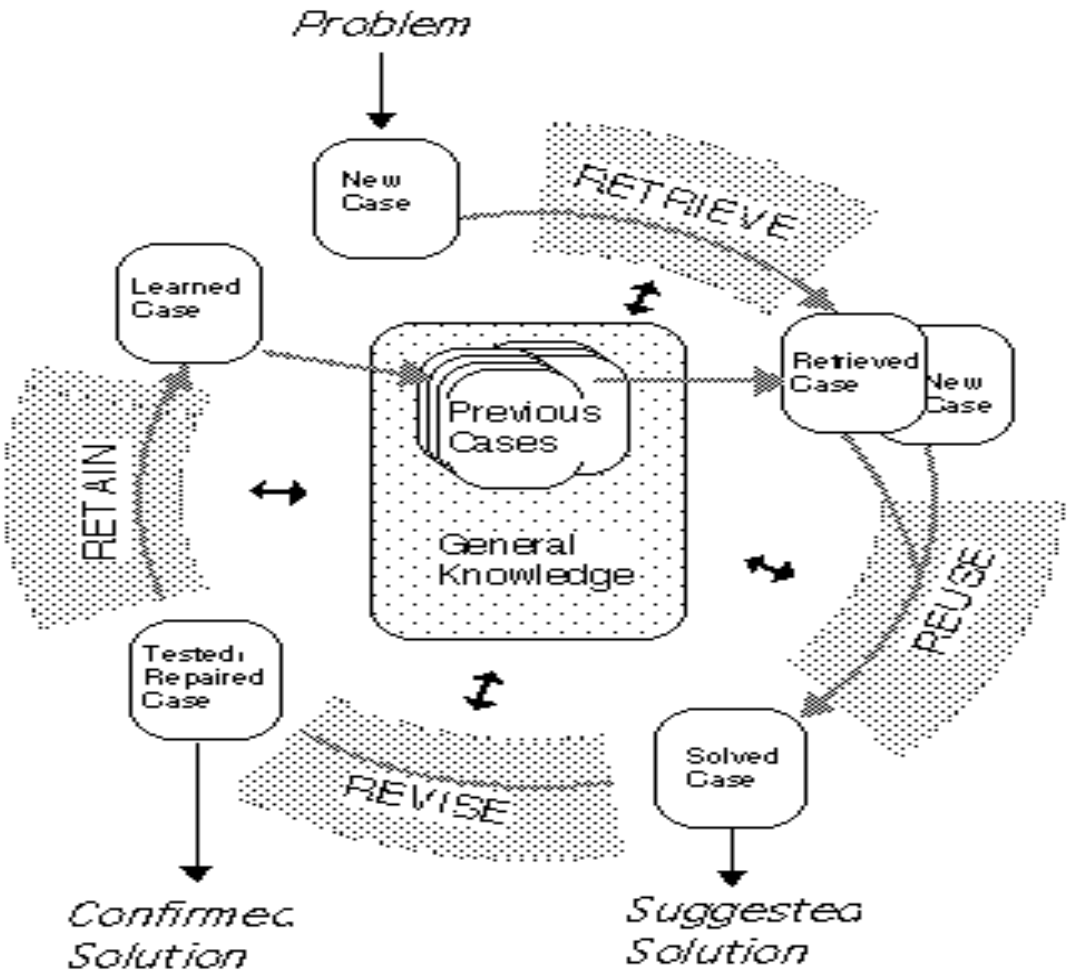

Fig3.2. Cicle de vida del CBR [Aamodt – Plaza]

En el cicle de vida cal notar dos punts a on intervé l'usuari, el primer punt és a on es presenta la solució al usuari *Suggested Solution*. El segon punt és a on l'usuari confirmarà la solució, ja sigui acceptant-la o la rebutjant-la:

Dintre del coneixement general (*General Knowledge*) no només podem recuperar els casos previs que més s'assemblin, sinó que existeixen tota una sèrie de mecanismes de recuperació del coneixement que d'alguna manera ens diu altres mètodes de com recuperar els millors casos

i de com poder adaptar-los en la fase d'emmagatzematge. Aquest mecanismes depenen de cada implementació del CBR.

3.2.1 Procés de *Retrieve*

Ja s'ha introduït quan és cridat el procés de R*etrieve*. L'objectiu d'aquest pas és recuperar un cas que sigui igual o semblant a un anterior per tant és necessari saber la similitud entre el nou problema i el problema recuperat. És defineix funció de similitud a la funció que ens retorna un valor de com son d'iguals un cas amb l'altre. És important saber quina funció de similitud s'aplica a l'hora de recuperar un cas de la base de casos, i inclòs saber si ens retornen més d'un cas. Si ens retorna més d'un cas podem analitzar cas per cas i seleccionar-ne un a partir d'un anàlisi posterior. Alguns atributs d'un cas poden ser més importants que uns altres, es per això que a la funció de similitud s'afegeix una ponderació dels atributs, per tal de que la funció de similitud ens retorni uns valors més alts segons les característiques del cas.

$$F_{similitud}(X, BC) = G_{distància}(X_0, BC_{0j}) \times W_o + ... + G_{distància}(X_i, BC_{ij}) \times W_i$$

on:

F: És la funció de similitud entre dos casos.

BC: Conjunt de tots els casos de la base de casos.

j: És número de cas de la base de casos.

X: És el nou cas.

i: Indica cada un dels atributs del cas **X** i de tots els casos enregistrats en **BC**

G: És la funció de distància entre els atribut **i**.

w: És el pes de l'atribut **i** dintre del cas.

Tenir un gran coneixement a l'hora de resoldre els problemes de qualsevol aplicació implica tenir Bases de Casos grans, és per això que el procés de retrieve té un punt crític que és el temps de recuperar el cas de la Base de Casos. Existeixen dos tècniques diferents a l'hora de recuperar els casos, la del veí més proper i la de l'inducció. Normalment aquestes dues tècniques s'apliquen per separat, o bé apliquem una o bé apliquem una altre. Existeixen però alguns entorns de treball es poden aplicar les dues a la vegada.

3.2.1.1 Mètode del veí més proper (Nearest Neighbour)

En el procés de *Retrieve* tenim la necessitat que donat un cas hem de recuperar el cas que més s'assembli dintre d'un gran conjunt de casos. Generalment la funció de similitud només s'aplica quan es treballa amb aquesta tècnica per que aquesta ens retorna un valor numèric entre 0 i 1, que ens diu com son d'iguals els casos. El procés càlcul consisteix en anar calculant la distància entre els mateixos atribut dels casos, i anar-ho sumant en una similitud total. Molt sovint ens interessarà donar més pes a uns certs atributs per que son més importants que uns altres, llavors introduirem un valor de ponderació per cada atribut.

Pel càlcul de la distància entre dos atributs es basa en el tipus d'atribut. Si son numèric es sol fer la diferència entre els dos valors de l'atribut dividit pel rang de valors d'aquest atribut. Quan son atributs simbòlics cal construir una taula de *nxm* elements on introduirem manualment la semblança entre el valor del símbol *n* amb el símbol *m*.

En la següent figura es presenten casos amb dos atributs, per tant podem representar els valors en un domini de dos dimensions.

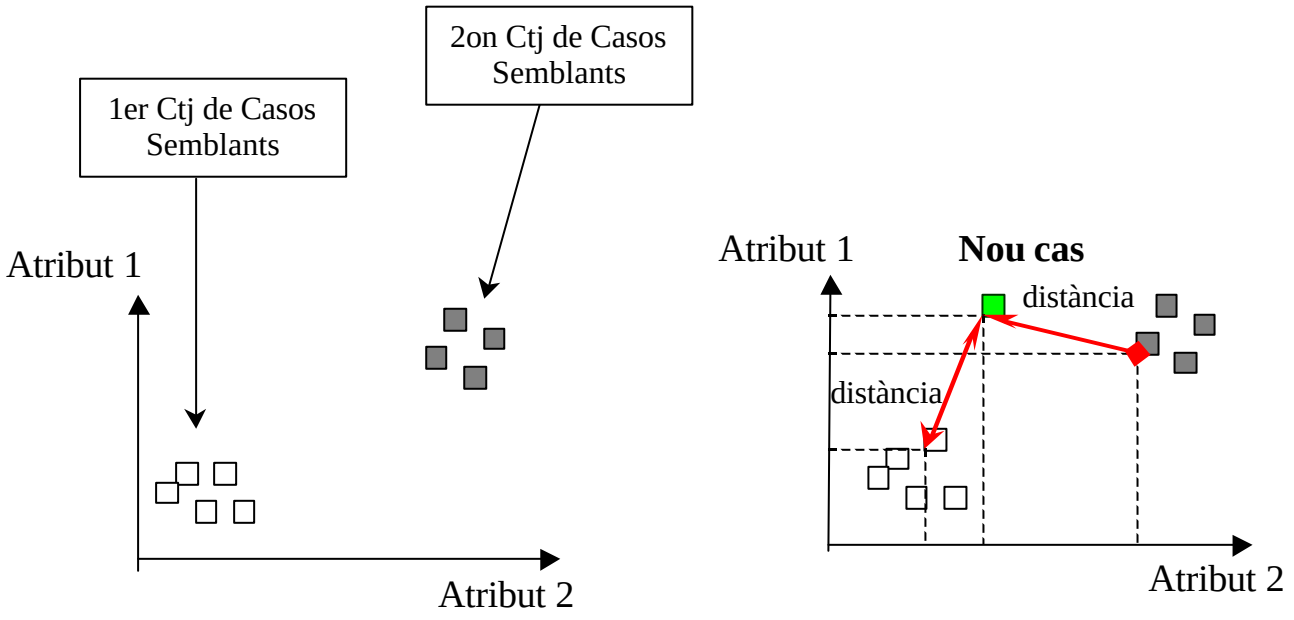

Fig3.3. Representació del domini dels casos per al 'veí més proper'

3.2.1.2 Mètode d'inducció (Induction Retrieval)

El mètode d'inducció consisteix en construir un arbre de cerca dels casos o arbre de decisió. Realment el que s'està fent és una ordenació dels casos segons els atributs dels cas. Per la inducció cal que els atribut siguin linealment separables per les branques de les cerques, ja que si no ens trobarem amb el gran problema d'aquest mètode, ens quedaran casos amagats en les branques. Des de la meva opinió personal aquest mètode només és vàlid per quan tinguem bases de dades immensament tan grans que el temps del procés de *Retrieve*, amb la tècnica del veí més proper, sigui inviable.

Si es dona el cas anterior de grans aplicacions i que es doni que tenim atributs linealment separables i d'altres que no ho son o bé no sabem es pot aplicar les dues tècniques a la vegada. Primer s'aplica el mètode d'inducció només als atributs que siguin linealment separables, un cop fet això ens quedaran tot un conjunt d'atributs que hi aplicarem la tècnica del veí més proper. D'aquesta manera solucionem que sempre trobarem el cas antic que més s'assembli a nou cas.

Cada node de l'arbre discrimina un rang de valors d'un atribut, si anem seguint un camí de l'arbre anirem descartant els casos que no compleixi el rang de valors del seu atribut i ens anirem quedant amb els casos que si compleixen que contenen un rang de valors que s'acosti al valor del nou cas.

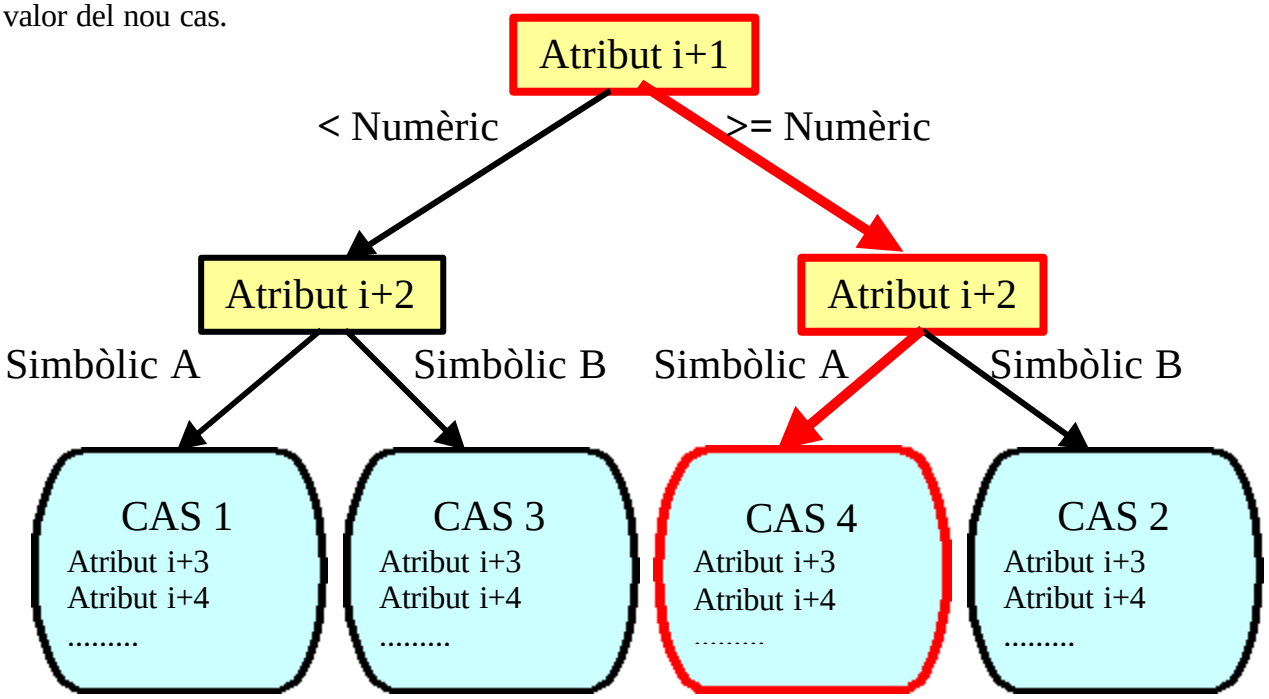

Fig1.4. Camí seguit en un arbre de cerca per al mètode d'inducció

Si ens fixem en el dibuix podem deduir que és important l'ordre dels atributs en el procés de cerca, ja que si podem un atribut molt important i que ens discrimina amb claredat uns grans conjunts de casos, el més lògic posar-lo al principi de l'arbre de cerca i així escurçarà molt el camí cap a la reca de casos desitjats.

Hi ha aplicacions en que el cas no conté directament la solució, sinó que conté com calcular la solució o bé com arribar a solució. Existeixen també aplicacions en que els casos tenen jerarquies i categories, a on un cas pot tenir diferents casos associats o bé un cas conté a tot un conjunt de casos. Hi ha moltes variants de configuracions que depenen principalment de l'entorn amb el qual es treballa.

Concretament en l'entorn en el qual es treballa en aquest projectes és amb Base de Casos homogènea, el que vol dir que tots els casos tenen exactament el mateix número i tipus d'atributs.

Base de Casos				
	Tipus A	Tipus B	.....	Tipus N
Cas 0	Atribut 0	Atribut 1	.....	Atribut i
Cas 1	Atribut 0	Atribut 1	.....	Atribut i
.....	.....	.....	.....	.....
Cas j	Atribut 0	Atribut 1	.....	Atribut i

Fig3.5. Estructura d'una base de casos homogènia

3.2.2 Procés de *Reuse*

La diferència entre el nou cas i el cas recuperat es té en compte en aquest pas, en el qual es mira fins a quin punt és possible reutilitzar la solució del cas anterior per a la solució del nou problema. Si la funció de similitud retorna un valor prou elevat es tindrà en consideració la solució del problema anterior. Si les diferències son molt grans llavors es descartarà la solució del problema antic i és dirà que per al nou problema no tenim una solució vàlida.

En aquest pas existeixen dos maneres de reutilitzar la solució recuperada:

1. **Reutitilització transformacional**: Es quan es diu que el cas recuperat conté exactament la solució al nou cas i no cal una adaptació de la solució anterior. Aquesta situació només tindrà lloc si la funció de similitud ens ha donat un valor molt a prop del seu màxim possible.

2. **Reutilització derivacional**: Es quan es diu que el cas recuperat no conté directament la solució al nou problema sinó que conté la metodologia de com s'ha de calcular o construir la nova solució. Aquest segon mètode de reutilització generalment només s'aplica a problemes de planificació, a on la solució ens diu el camí a seguir per arribar a la solució. Cal dir però el procés de derivació cap a la solució hauria de ser un procés fiable i exacte, ja que sinó ens portaria a solucions errònies i sovint aquest inconvenient ens faria caure el rendiment de la tecnologia del Raonament Basat en Casos.

3.2.3 Procés de *Revise*

En aquest pas es quan tenim un nou problema amb una nova solució recuperada anteriorment, la pregunta que ens apareix ara és si realment aquest cas és correcte. Per decidir això definim funció d'avaluació com a una funció que ens acceptarà el cas si el cas és correcte i ens rebutjarà si no ho és. Si un cop avaluada la solució l'acceptem llavors podem passar a la següent fase d'aprenentatge.

És evident que en el punt anterior ens apareix una qüestió important que cal aclarir. Si tenim una funció que ens diu que per un problema si la solució es correcte o no, ve a dir que podem saber per tots els problemes totes les solucions. El raonament anterior és incorrecte, ja que es dona que el domini de totes les possibles situacions dels problemes és tan gran o tan complexe que és intractable. Donem per exemple un problema de la intel·ligència artificial clàssica, el joc dels escacs. L'espai dels millors moviments de totes les figures dels escacs, que cal fer pe guanyar la partida per qualsevol estat del joc, resulta un problema exponencial que és intractable pels computadors actuals.

Tenim doncs que l'única manera de saber si un cas és correcte o no es provar-lo i veure la resposta. Sempre tindrem situacions en que el cost de provar una solució i de que no sigui la correcte sigui molt elevada. En aquest cas no podrem aplicar el Raonament Basat en Casos, recordem la definició que ens basem en experiències prèvies, per tant hem d'experimentar les noves solucions i observar el seu resultat.

3.2.4 Procés de *Retain*

En aquest pas és on apareix l'aprenentatge del Raonament Basat en Casos. El procés pot consistir en diverses subtasques, però el procés sempre acabaran enregistrant el nou cas en una base de casos., L'algoritme d'aprenentatge d'alguna manera ja es va començar en el pas anterior que és quan s'avalua la solució proposada, però no és fins aquest moment quan és configura el cas a emmagatzemar. Podem enumerar diferents tasques dintre d'aquest procés.

1. Seleccionar quina informació ha de ser guardada.
2. Establir com ha de ser aquest informació guardada.
3. Establir com s'indexarà el nou cas per a futures recuperacions.
4. Determinar com s'integrarà el nou cas dintre de la base de casos.

No sempre es realitzarà el procés de *retain*, podem trobar aplicacions a on es desitja tenir una base de casos que sigui capaç de resoldre un determinat número de problemes i per tant no existeix una dinàmica que requereixi un canvi o una evolució constant del coneixement de la base de casos. En aquesta situació podem dir que el nostre sistema està entrenat.

En sistemes dinàmics que sigui necessari esborrar o actualitzar un cas caldrà fer:

1. Si el cas és reutilitzat i adaptat per a generar una nova solució, el nou cas ha de ser generalitzat com a part d'un nou cas que ha d'absorbir al antic cas, o bé crear-ne un cas completament nou.
2. Si la solució necessita l'intervenció de l'usuari, cal que un el nou cas sigui totalment construït con a un nou cas.

Un cop hem afegit el cas a la nostre base de casos es possible que calgui reindexar tota la base de casos, ja que s'ha afegit un cas nou. La necessitat de reindexar depèn de les característiques de la nostra aplicació, ens podem trobar aplicacions que mai calgui reindexar, d'altres que es faci cada un cert número de casos nous o inclòs cada nou cas.

3.3 Analogia entre els sistemes experts i el CBR

En aquest punt és convenient introduir una analogia entre els sistemes experts i el raonament basat en casos, que per aquelles persones que tinguin experiència amb els sistemes expert resultarà més fàcil entendre el concepte i els elements del CBR fent una analogia entre els dos.

Podem dir que un cas és equivalent al que seria un fet, el qual el podem deduir-lo i esborrar-lo a una base de coneixement, que seria la Base de Casos. L'experiència que es necessita per construir un fet amb una estructura que sigui capaç de contenir coneixement, és idènticament la mateixa que es necessita a l'hora de definir un cas. És per això que en el Raonament Basat en Casos també surten preguntes a l'hora de definir l'estructura dels casos, com poden ser: quin tipus d'eslots tindrà, si son del tipus numèric, simbòlic, o si bé cal definir pocs eslots i molt complicats o molts eslots i molt senzills.

4 Construcció d'una botiga virtual amb e-Merchandising

4.1 Descripció del sistema

El sistema contruit es basa en dues grans parts. La primera consta d'una botiga virtual en la qual hi tenim tota una sèrie de productes els quals podem comprar. Un cop els hem seleccionat, la botiga de comerç electrònic ha de ser capaç d'extreure el nostre perfil d'usuari i donar-nos un servei d'atenció al client per oferir-nos més productes o serveis que s'avinguin amb les nostres necessitats. Aquesta tècnica de facilitar-nos o bé presentar productes en llocs directes a nosaltres se'n diu *Merchandising.*

La segona gran part del sistema CBR construït és el simulador, a on podem establir els paràmetres de simulació i observar l'evolució i del raonament del CBR. Aquest simulador ens retornarà mesures de rendiment i també tindrem informació del número de casos dintre de la base de casos, a més dels valors de la funció de similitud.

Podem dir que la nostra botiga serà com un agent de vendes. Segons [Wolfgang – Ralph – Stefan] la tasca de l'agent de vendes durant el procés es l'adaptació iterativa de la demanda de l'usuari, afegint casos o canviant-los i l'adaptació iterativa de la configuració dels producte amb l'objectiu de trobat un punt d'acord en l'espai multidimensional demanda/producte, entre la demanda i el producte.

En el nostre simulador podrem veure com s'executa pas a pas el procés del cicle de vida del CBR, a més d'això tenim també tres apartats on podem treballar directament amb el cas i per tant podrem repetir manualment tots el passos del cicle de vida.

4.1.1 La botiga virtual

En la botiga virtual hi haurà diversos tipus de productes que podran ser adquirits, generalment seran alimentaris ja que són els que es compren més sovint. La botiga virtual estarà implementada en una plataforma real i de lliure accés en la xarxa internet, d'aquesta manera qualsevol persona podrà comprar i examinar els resultats de la botiga.

La implantació la botiga virtual en una empresa ha de ser la més directe possible. El sistema ha de permetre que el número de productes a la venda sigui totalment variable i que és pugui canviar amb gran facilitat el número de productes ofertats.

Un altre requisit imposat es que la botiga virtual no pot conèixer cap dada personal de la persona que està comprant i per tant de cap manera pot saber les compres que ha realitzat aquest client anteriorment. Com a opinió personal estic en desacord amb els actuals sistemes que emmagatzemen informació dels usuaris, sense que aquest sàpiguen realment que s'està fent amb les seves dades. Actualment existeix la llei LORTAD (*Lei orgánica que regula el tratamiento automático de los datos*) que ens protegeix de totes les empreses que es vulguin aprofitar de dades personals, però tot i així encara no és respectada per la majoria d'empreses. Només les empreses més responsables i generalment les més importants, que és quan es hi ha en joc més capital, respecten la desinformació dels drets dels seus clients.

4.1.2 El simulador de la botiga virtual

Un dels problemes més gran del raonament basat en casos és l'obtenció del seu element fonamental, el cas. Posem un exemple: La realització d'aquest projecte. No és pot construir un sistema del CBR a la primera, es necessita fer un primer prototipus, després calen fer proves de rendiment, proves de consistència canviant el domini del model de dades de l'aplicació, configurar els pesos dels atributs del cas, escollir un determinat mètode pel procés de *Retrieve*, i en general la depuració de l'aplicació. És per això que no es pot fer una primera versió de la botiga virtual, fer una promoció per que la gent compri, observar els canvis que calen fer, implantar els nous canvis en una nova versió i començar un altre cop. Aquest procés d'implementació no ens portaria mesos sinó anys.

Una de les primeres conclusions que s'avancen d'aquest projecte és la construcció obligatòria d'un simulador per tot sistema CBR aplicat al comerç electrònic, i sempre es farà abans de posar la botiga virtual en funcionament cara al públic. Amb aquest simulador podrem canviar els paràmetres del CBR, veure el seu funcionament, descobrir aspectes que estaven ocults i com ha succeït en aquest projecte deduir nous conceptes del CBR per aquest camp.
Aquest simulador s'ha construït amb una de les darreres tècniques que s'apliquen al món d'internet, els Servlets amb els Java Server Pages, juntament amb el clàssic Java i HTML.

4.2 Desenvolupament del projecte

Per desenvolupar aquest projecte és necessari adquirir coneixements de programació de llenguatges orientats a les telecomunicacions Java, HTML, Java Server Pages. També és necessari conèixer tota la tecnologia del CBR per entendre el paquet de funcions del JavaKate. JavaKate és una implementació del raonament basat en casos creada per Acknosoft en Java. Dintre d'aquest paquet existeixen desenes d'objectes, els quals cal conèixer per poder afegir casos, esborrar-los, poder fer consultes dintre de la base de coneixement.

Existeix també l'aplicació JavaKate que ens facilitarà la configuració d'alguns paràmetres del CBR. Cal dir però que en tot sistema que es desitgi una integració del CBR amb el JavaKate caldrà anar fins al punt d'implementació de baix nivell amb les crides al paquet de funcions del JavaKate. Actualment totes les aplicacions, i més per al comerç electrònic s'han de realitzar a mida del client que desitgi implantar aquesta tecnologia.

Primer de tot és necessari tenir o ser capaç d'adquirir suficients casos per permetre tenir una representació de tot el domini de l'aplicació.

Un punt molt important és que hem de tenir en compte que amb aquest disseny, no podrem saber directament quin tipus de producte prefereix un usuari amb un perfil determinat. Això ens pot sobtar, perquè el CBR sí sap quin producte pot ofertar i és perquè el coneixement està implícit en la base de casos (al no estar explícit no el podem veure).

Amb aquest disseny només podrem fer preguntes en una sola direcció. Per poder respondre algunes de les preguntes en l'altre sentit es podria provar de fer consultes SQL directament a la base de casos com a base de dades.

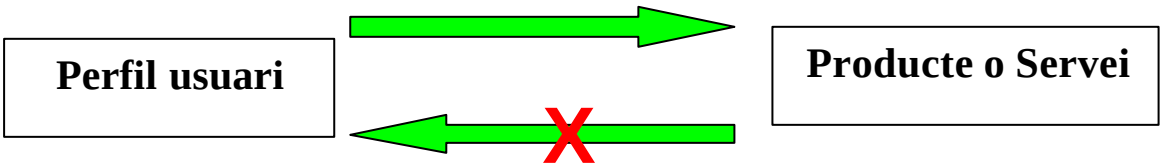

Fig4.1. Infererència de dades a decisions.

4.3 Traducció de la llista de productes a un cas

Abans de la construcció de tot sistema CBR cal fer l'anàlisi i el disseny de l'estructura d'un cas. Aquest cas degut a la configuració del JavaKate el cas ha de tenir una estructura homogènia, tots els casos han de ser iguals i a més l'estructura d'un cas no pot ser modificada durant el temps de funcionament del CBR.

En la nostre aplicació l'origen de les dades és una llista de productes que l'usuari ha comprat, per tant aquest llista és variable en quantitat de productes comprats. Tenim doncs un problema de com representar una llista de productes variables en un cas amb estructura fixa, i per tant amb un número d'atributs del cas fix.

Per dissenyar un sistema de fàcil traducció de la llista de productes comprats a un cas en ajudarem en la codificació dels productes tenen en el seu codi de barres. Del codi de barres dels productes extraurem les característiques d'aquest. Aquestes característiques les passarem directament als atributs del cas.

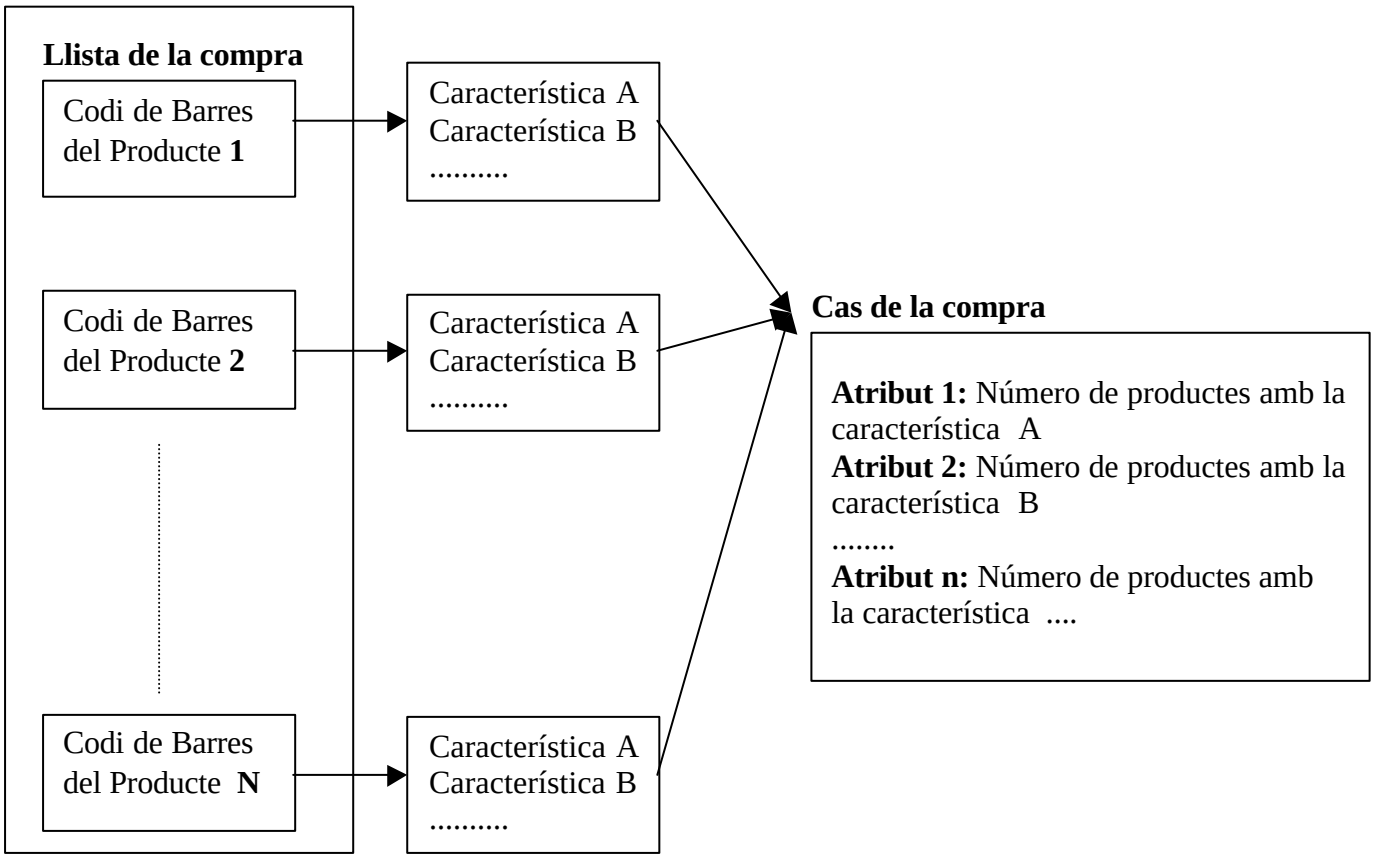

Fig4.2. Transformació del la llista de productes a un cas.

4.3.1 Codi de barres dels productes

Aquí és on hi una penalització de cost a l'hora d'implementar el CBR. Aquesta no serà feina del sistema CBR si no de l'empresa que comercialitza els seus productes o serveis. Inclourà en el seu codi de barres la informació addicional necessària pel nostre sistema.

Actualment els codis de barres estan establerts seguint una normativa, però cada empresa que posa a la venda els seus productes els codifica. Actualment ja existeixen normes per classificar els productes segons les seves característiques de pes, emmagatzematge, tipus de producte. Només caldrà afegir a aquesta norma uns atributs estàndards de classificació de productes que és de ben segur que es faran servir en tota plataforma de comerç electrònic.

Fig4.3. Dibuix típic d'un codi de barres.

Dins del codi de barres en el camp ISSN és el que es pot fer servir com a base per a la construcció del codi de barres. El centre Nacional de l'ISSN (http://www.bne.es/issn04.htm) no proporciona codi de barres, però sí proporciona l'ISSN a partir del qual es pot obtenir el codi de barres i després demanar-ho als fabricants de masters o mitjançant algun programa informàtic especialitzat. El codi de barres amb ISSN segueix la norma EAN-13, amb la simbologia més estesa en la codificació comercial.

Exemple:

977	1144875	00	7	03
Prefix	**ISSN**	**Codi de preu**	**Dígit de control**	**Adendum**

4.4 Definició del Domini del Model de Dades del CBR

4.4.1 Definició de les classes del Model de Dades

Per a la definició de les classes del model de dades hem d'establir els conjunts de les característiques del productes. Aquesta agrupació de les característiques dels productes ens determinaran les classes del model de dades. Una classe del model de dades ens agrupa tots aquells atributs que s'identifiquen amb la classe.

La definició del model de dades ve determinada en gran part per les possibilitats que ens permet el paquet de funcions del JavaKate. Podem tenir sistemes CBR que ens permetin fer altres definicions del model de dades.

Dividirem els atributs dels productes de la nostra botiga en cinc classes:

- Producte o servei ofertat: **Merchandising**
- Sistema d'embalatge del producte : **Empaquetat**
- Tipus del producte : **Tipus**
- Lloc a on es guarda el producte: **Magatzem**
- Els aspectes nutricionals dels productes: **Aspectes Nutricionals**
- Cost d'adquisició del producte: **Preu**

Amb aquesta definició de productes podem tenir en la base de coneixement del CBR els perfils d'usuari. Cal observar que dins del cas no tindrem especificat el perfil de l'usuari, ja que aquest està implícit en el model. El que contindrem seran característiques que compleixen el perfil.

En aquesta definició de les classes del cas podem afegir altres atributs que s'associen a les característiques dels diferents productes com poden ser:

- País d'origen del producte: França, EEUU,...
- Propietats organolèptiques del producte com: Sabor o gust, Olor,...
- A la persona que va orientada el producte: Homes, Dones, Nens,..

4.4.2 Definició dels atributs del cas

Un cop tenim definida les classes que conté l'estructura del cas cal definir els atributs que conté cada classe. Els atributs de cada classe han de ser els més representatius i a més els atributs han de capaços de definir un perfil d'usuari. Recordem que el coneixement que volem tenir dins de la base de coneixement son els diferents tipus d'usuaris amb els productes que s'adaptin al seu perfil. Els atributs del cas seran del tipus numèric i contindran un valor que ens digui el número de productes de la llista de la compra que compleixen aquell atribut.

Exemple:

Producte	***Atributs que compleix***
Oli d'oliva	Embotellat, cru, sec, alt calories, normal
Pèsols congelats d'oferta	Bossa, verdura, fred, baix calories, oferta
Fuet extra	Natural, curat, sec, calories normals, luxe

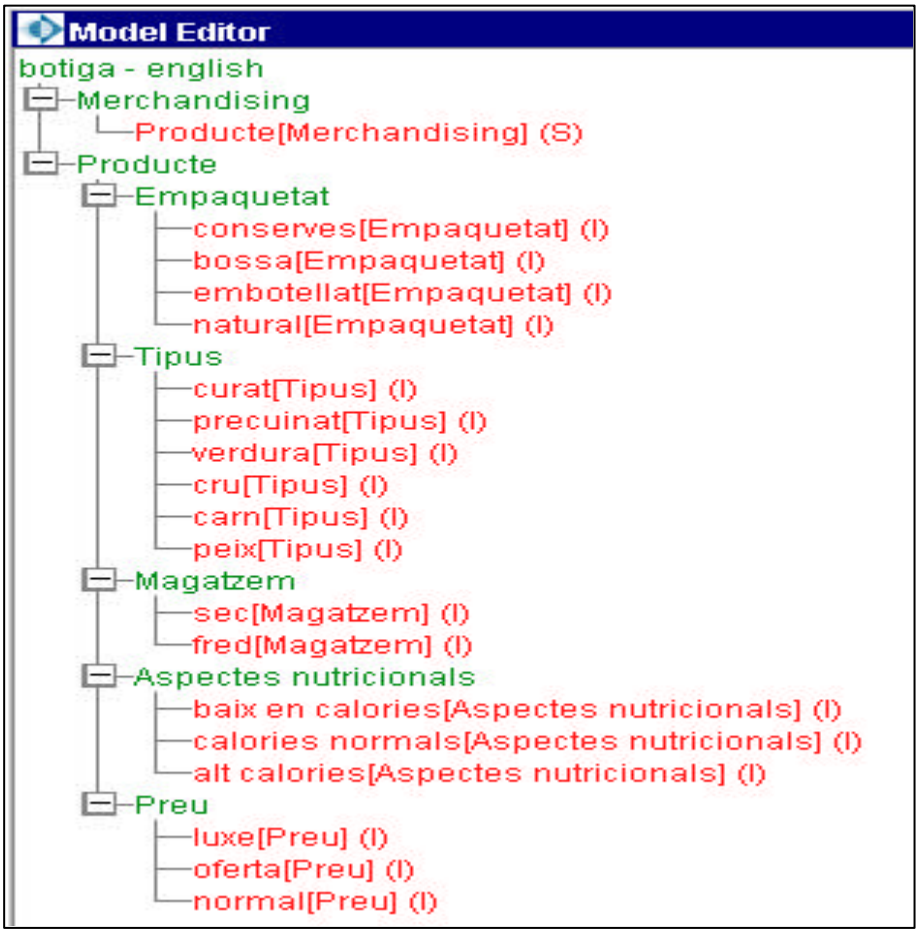

Fig4.4. Configuració dels atributs del cas.

4.5 Definició de la Funció Similitud entre casos

La funció de similitud que ens calcula la semblança entre els casos ve determinat internament pel paquet de funcions JavaKate i és de la forma:

$$f_{similitud}(X,Y) = \sum_{i=1}^{N} g_i(Xi,Yi) \times Wi$$

on:

X: És el nou cas

Y: El cas a comparar dintre de la base de casos

N: Número d'atributs del cas

i: Atribut del cas X i Y

g: Funció de distància de entre els dos atributs dels dos casos.

W: Pes de l'atribut i

Com ja s'ha definit anteriorment els tipus atributs seran del tipus numèric i contindran el número de productes que compleixen la característica de cada atribut.

La funció **g** distància entre dos atributs la podem implementar nosaltres mateixos o bé podem fer servir algunes de les que ens facilita el JavaKate. En el nostre cas agafem una funció facilitada pel JavaKate que calcula la distància numèrica entre dos valors numèrics dividit pel rang del valor numèric.

$$g_i(Xi,Yi) = \frac{|Xi - Yi|}{Rang(i)}$$

Per últim comentem que en el paquet de funcions del JavaKate s'ha d'especificar el rang que té cada atribut i és calculat de la següent manera:

$$Rang(i) = max(i) - min(i)$$

4.6 Definició dels pesos dels atributs del cas

En la funció de similitud entre dos casos podem donar més importància a uns atributs més que a uns altres, de manera que la funció de similitud es donarà valors alts per a aquells atributs que em discriminen més els casos. La manera que té el JavaKate per donar més importància a uns atributs o a uns altres es multiplicant la funció distància per un valor de ponderació. Cada valor de ponderació se'n diu el pes de l'atribut que té en el conjunt de tots els atributs. Els valors de ponderació van des de el 0.0 fins al 10, on el 0 indica que l'eslot no és rellevant en el càlcul de similitud i el 10 indica que és el màxim de rellevant.

Per aquest model s'han establert que els productes de luxe, els d'oferta i els productes curats com els embotits, precuinats, els alt en calories i baix en calories són els que discriminen més els perfils d'usuaris que tindrem a la botiga. Els pesos dels altres atributs només tenen una importància menor.

1.0	Producte[Merchandising]
4.0	conserves[Empaquetat]
4.0	bossa[Empaquetat]
4.0	embotellat[Empaquetat]
4.0	natural[Empaquetat]
10.0	curat[Tipus]
10.0	precuinat[Tipus]
10.0	verdura[Tipus]
6.0	cru[Tipus]
6.0	carn[Tipus]
6.0	peix[Tipus]
5.0	sec[Magatzem]
5.0	fred[Magatzem]
10.0	baix en calories[Aspectes nutricionals]
5.0	calories normals[Aspectes nutricionals]
10.0	alt calories[Aspectes nutricionals]
10.0	luxe[Preu]
10.0	oferta[Preu]
5.0	normal[Preu]

Fig4.5. Configuració dels pesos dels atributs del cas.

Cal dir que els valors d'aquest pesos no són estrictes, podem modificar els valors dels pesos i obtenir els mateixos resultats en el sistema. O bé si no tenim cap coneixement sobre el tipus d'usuari que ens comprarà en la nostra botiga posarem a tots els atributs el mateix pes.

Per a la implementació del projecte s'ha establert que els pesos representats en la taula anterior son una de les millors configuracions que diferencien els casos, per tant que la funció de similitud en el procés de *Retrieve* té els millors resultats.

4.7 Productes i serveis ofertats per la botiga

Els productes i serveis que poden ser ofertats per la botiga poden ser de qualsevol tipus, ja que no es fa cap distinció entre un i altre. Poden ser oferits tant productes d'alimentació com entrades de cinema, electrodomèstics, viatges, i inclòs et pot recomanar anar a sopar a un restaurant determinat.

Productes ofertats:

- Delícies de mar
- Ordinador
- Televisor
- Torrons xocolata
- Camera fotos
- Entrada cinema
- Croissant
- Entrada teatre Liceu
- Formatge semicurat
- Viatge Tenerife
- Vi negre
- Sopar restaurant TotMil
- Mòbil
- Llonganissa
- Pernil salat
- Equip musica
- Magdalenes
- Natilles
- Viatge Balears

El número total de productes o serveis que poden ser oferts als clients són 19. S'ha construït en el model que aquest número no sigui fix i així podrem augmentar el número de productes ofertats sempre que vulguem. Poder canviar les ofertes és de vital importància, ja que permet oferir productes segons les necessitats immediates de la botiga. Aquest necessitats ens faran entrar en un dinàmica molt gran a l'hora de canviar els productes ofertats. Per exemple en un moment determinat en la nostra botiga ens podem trobar amb una acumulació de productes en estoc, per tant ens interessaria fer sortir els productes quan més aviat millor.

Possibles necessitats de canvi de la llista de productes ofertats:

- Publicitat personalitzada als diferents tipus de clients
- Acumulació de productes en estoc.
- Promoció de nous productes.
- Productes en oferta o rebaixa..

- Servei d'ajuda al client per seleccionar els productes.

Per augmentar o disminuir el número de productes ho haurem de fer modificant la definició del domini del model de dades des del JavaKate. L'atribut del producte ofertat és de tipus simbòlic.

En el nostre model al no fer-se servir el mètode d'inducció en el procés de *Retrieve* aprofitem aquest tipus d'eslot simbòlic per tenir la particularitat de tenir un llista de productes ofertats variable.

4.8 Nova etapa d'excitació del cicle de vida del CBR

En el tema sobre el Raonament Basat en casos em vist com en el cicle de vida quan es trobaven casos acceptats per l'usuari aquest eren guardats dins de la base de casos. Aquesta base de casos el que realment està guardant és coneixement, el coneixement de com resoldre els problemes. A més en el CBR només es guarden els casos que realment serveixen per resoldre els problemes, ja que aquest cas ha sigut presentat i ha sigut acceptat per l'usuari. En definitiva la base de casos té una evolució dels casos més ben adaptats per resoldre els problemes. Només ens faltaria aplicar un mecanisme al CBR per a que tingui una evolució natural: creuament, aleatorietat i mutació de casos. Fent un símil amb la Unibiodiversitat podríem implementar altres mecanismes d'evolució basant-se amb la *Unicasdiversitat* que faria que el CBR ens plantegés nou problemes.

Podem observar que el cicle de vida del CBR és un cercle tancat, per tant si no existeix cap mecanisme que aporti més solucions a la base de coneixement aquesta es queda estancada com una basa d'oli. Es per això que necessitem excitar el sistema, proposar ja sigui amb l'aleatorietat, mutació o creuament de casos, noves solucions. D'aquesta manera en el temps s'aniran creant nous casos, dels quals només una part seran els acceptats i afegits a la base de casos.

Per al nostre objectiu en el projecte s'ha implementat un nou mecanisme per tal de fer evolucionar la base de casos. Entre el procés de *Reuse* i *Revise* s'ha introduït un nou pas, el procés de *Excitació*. Pot haver-hi gent que pensi que aquest pas s'ha d'incloure dintre del procés de *Revise*, però els conceptes introduïts son prou importants com per crear un nou procés.

En aquest projecte el tipus d'excitació implementada és l'aleatorietat. Quan un cas és recuperat en el mètode de retrieve és mira la seva funció de similitud. Aquest nou procés es comporta mirant si el valor de la funció és molt baix llavors és millor proposar un altre solució encara que sigui aleatòria abans de donar una solució dolenta. El punt on decidim si hem d'excitar o no en direm *llindar de la funció d'excitació*. Existeixen moltes possibilitats a l'hora de construir una funció d'excitació com per exemple tenir en compte la mitja d'encerts del sistema. L'estudi dels valors del llindar d'excitació es veuran en el següent capítol en l'estudi de resultats del sistema.

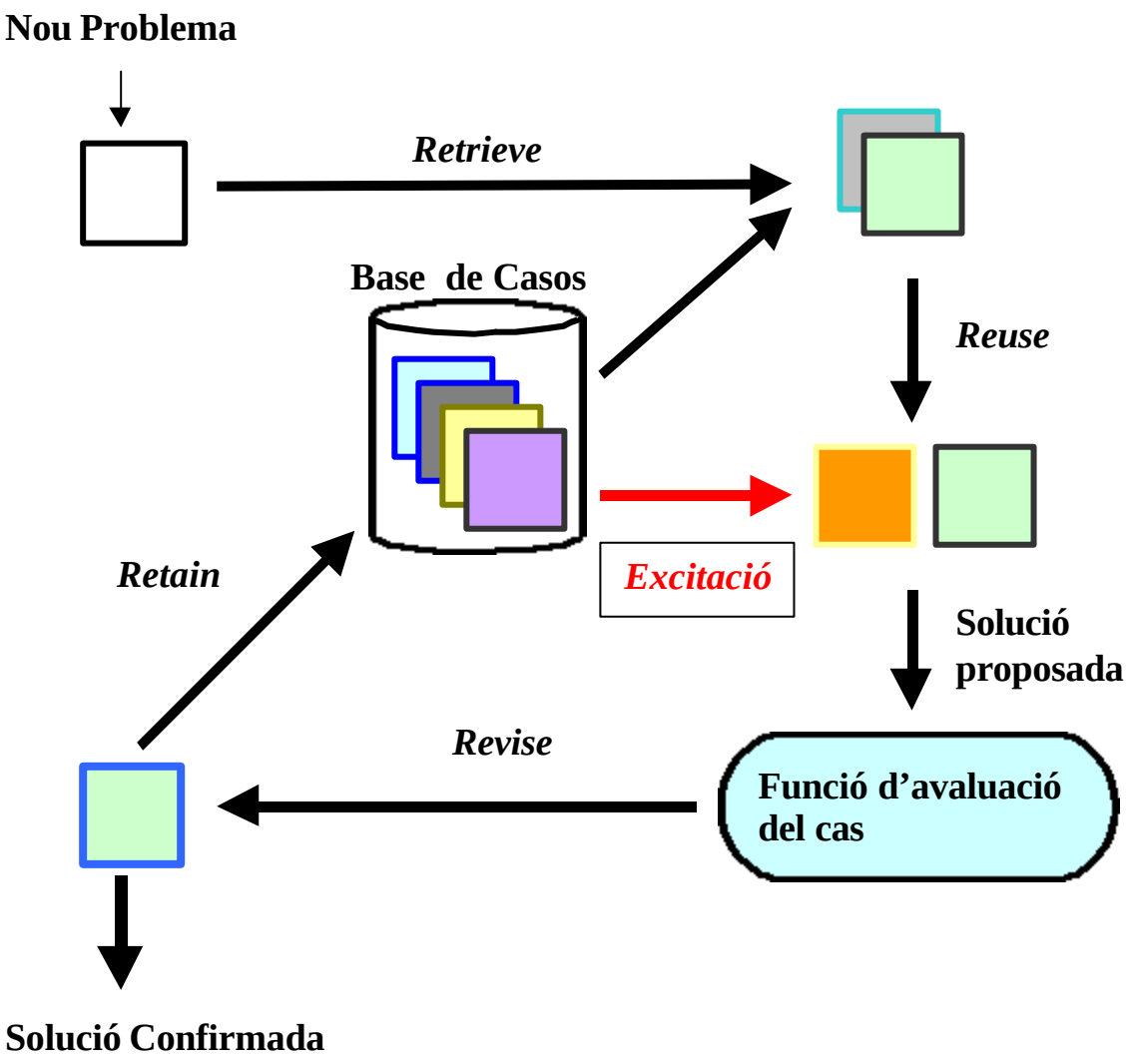

Fig4.6. Cicle de vida proposat per al CBR aplicat al comerç electrònic.

4.9 Funció d'avaluació d'usuari

Recordem que en el procés de *Reuse* un cop quan ja hem recuperat un cas, aquest és presentat a l'usuari. Si repassem el cicle de vida del CBR quan el cas es presentat com a solució cal algun mecanisme que ens avaluï aquest cas abans de passar al procés de *Retrieve.* Necessitem saber si el cas recuperat és realment bo o no, hem d'avaluar si la solució que ens dona el cas és correcte. La funció que ens dirà si el cas recuperat és correcte és la funció *d'avaluació d'usuari.*

Per a la botiga virtual la funció d'avaluació és directament l'usuari, que ens avaluarà el cas recuperat i ens dirà que accepta el cas proposat o bé el rebutja. No passa però, el mateix en el simulador de la botiga i una qüestió se'ns fa crucial: En un simulador de CBR que captura característiques de persones com podem construir una funció d'avaluació que ens digui si un cas és correcte o no? Bé la resposta és que no podrem construir mai una funció que raoni igual que una persona humana, i que aquesta funció tingui els mateixos gustos i preferències que una persona. El que sí que podrem fer en el procés de simulació crear un perfil d'usuari i executar una funció d'avaluació sabent el perfil d'usuari.

Defineixo funció d'avaluació d'usuari com a aquella funció que a partir d'un perfil d'usuari i un producte ofertat ens retornarà SÍ, si el cas proposat a l'usuari és acceptat i retornarà NO si l'usuari rebutja el producte ofertat.

Aquesta funció serà un funció hash, la qual sempre ens retornarà el mateix valor per una determinada tupla perfil i producte ofertat. És totalment indiferent per quins parells de perfil-producte aquesta funció ens accepti o ens rebutgi, el que si volem es que sigui consistent i sempre ens doni els mateixos valors de sortida pels mateixos valors d'entrada.

$$f_{avaluacio}(\text{Perfil}, \text{Producte}) = \begin{cases} \textbf{SI} & \text{si } (f_{Hash}(\text{Producte}) + f_{Hash}(\text{Producte}))\ \%2 == 0 \\ \textbf{NO} & \text{altrament} \end{cases}$$

Fig4.7. Funció d'avaluació d'usuari proposada.

Podem fer però que la funció d'avaluació d'usuari sigui òptima. Per construir-la caldria fer un estudi de mercat dels productes que compren els diferents tipus d'usuaris i aplicar-ho a al simulador de la nostra botiga virtual.

Per construir aquesta funció primer caldrà dotar el nostre simulador la capacitat de crear diferents perfils d'usuaris. D'aquesta manera segons el perfil d'usuari amb el que es faci la simulació tindrem que l'espai de casos simulats s'ha costa a una classificació de casos determinada.

4.10 Quantificació de la Satisfacció de l'usuari

Tant en la botiga com en el simulador necessitem una mesura que ens digui com de bé està funcionant el sistema. No s'ha implantat una sola mesura de funcionament del sistema, ja que amb diferents mesures ens aporten més informació de tot el conjunt.

S'han implementat quatre mesures de funcionament. La Mitja de les Similituds del Sistema, Bondat del Sistema i Rendiment del Sistema per al simulador i per a la botiga tindrem la mesura de Rendiment de la Botiga.

Totes les mesures que es presenten contemplen tot l'històric de la base de casos. Altres variacions de mesures de la satisfacció, que només les nomenarem, seran aplicar la tècnica de la finestra a totes les mesures satisfacció. D'aquesta manera podrem observar l'estat de la satisfacció de l'usuari en diversos punt en el temps, com per exemple: Quan la base de casos només conté un perfil d'usuari, quan la base de casos només té un número mínim de casos, quan el CBR ja està entrenat i té un perfil, dos perfils, etc.

El simulador de la botiga ens dona la possibilitat d'accedir a tots els valors de totes les mesures en fitxer de dades. Aquest fitxers són creats dinàmicament per cada simulació que fem, per tant contenen directament els resultats de la nostre execució. Podem descarregar-los de la xarxa i graficar i estudiar els resultats de simulació.

4.10.1 Mitja del les similituds del sistema

És un valor entre 0 -1 que conte la mesura de confiança mitja del sistema de tots el casos que s'han recuperat i l'usuari ha acceptat. Aquest valor es fa servir pel llindar d'excitació del sistema partir del qual es fa servir la solució recuperada o bé si està per sota es proposa una altra.

El sistema amb aquesta mesura determina si és millor crear una solució aleatòria abans que donar una solució dolenta. Això també em serveix per enriquir la base de casos, ja que si no es queda 'estancada' al no crear-se casos nous dins de la base de casos.

MitjaSimilitudAccepats = SumaSimilitudAccepats / CasosRecuperatsAcceptats

4.10.2 Bondat del Sistema

És un valor entre 0-1 indica que indica el tant per cent de casos que l'usuari ha acceptat sense tenir en compte que hagin sigut recuperats pel sistema de la base de casos. Aquesta és una mesura que només es calcula en el simulador de la botiga. Aquesta mesura no té en compte si el producte ofertat prové del procés d'excitació o bé prové del procés de *Reuse*. Amb la bondat del sistema podem veure la capacitat de resposta del sistema incloent quan no es troba un cas igual anterior. Ja hem vist que quan no és trobat un cas anterior el que fa el sistema és generar una oferta d'un producte aleatòriament.

BondatSistema = SumaCasosAcceptats / NumeroSimulacionsTotals

4.10.3 Rendiment del Sistema

El rendiment del sistema és un valor de 0-1 que indica el tant per cent de casos que l'usuari ha acceptat, però només tenint en compte que els productes ofertats hagin sigut recuperats pel sistema de la base de casos. Aquesta és una mesura molt més estricte que la bondat del sistema, per que ens penalitza molt si el cas no ha sigut recuperat de la base de casos.

RendimentSistema = CasosRecuperatsAcceptats / NumeroSimulacionsTotals

És convenient inicialitzar el rendiment del sistema un cop tenim una base de casos amb un mínim de coneixement. A l'inici de la vida de la base de casos tindrem pocs casos per tant la majoria d'operacions que es realitzaran seran excitacions del sistema i l'encert de la base de casos serà baix. Si no inicialitzem el rendiment del sistema ens trobarem que el rendiment del sistema anirà pujant molt a poc a poc ja que el número de simulacions serà alt. El número de simulacions inicials a descartar o dit d'una altre manera el moment de inicialització del rendiment de la botiga no té un punt determinat pot variar quan tinguem 100 o 200 casos. De vegades ens interessarà inicialitzar el rendiment de la botiga cada un cert període de temps, per veure la seva evolució en el temps, sense tenir en compte la seva història.

4.10.4 Rendiment de la Botiga

La mesura per a la botiga virtual que es proposa i que és la més entesa per tothom, és el tant per cent de compres que s'han acceptat, o sigui les vegades que el client ha dit que si a la oferta del producte. D'aquesta mesura en direm Rendiment de la Botiga.

Rendiment de la Botiga = Núm. Propostes Acceptades / Número de Compres

Aquesta mesura és exactament la mateixa que la bondat del sistema en el simulador, però té un nom diferent per que des del punt de vista del client és el rendiment que realment observa en la botiga.

Aquesta mesura ens apareixerà en la botiga virtual com a nota informativa abans de fer una compra, i ens servirà per donar un vot de confiança per a que el client compri el producte.

5 Estudi i Resultats del Sistema

5.1 Motor de Simulació

En el motor de simulació podem establir alguns paràmetres de simulació. Dins de la simulació tenim dos grups de paràmetres, els paràmetre d'execució i el de la configuració del perfil d'usuari. En els paràmetres d'execució introduirem el número de simulacions que volem fer, amb un màxim de 100 simulacions. Cal dir que la velocitat de resposta del simulador depèn de moltes la càrrega que hi hagi en la xarxa. Encara que el simulador s'executi en el servidor es necessita un temps en enviar les dades cap al client. El número màxim de productes de cada tipus és per dir indicar longitud té la llista de la compra. En el simulador no sabrem mai el número de productes que hi ha a la llista, el que sabrem serà el número de productes que compleixen una característica establerta pel model de dades.

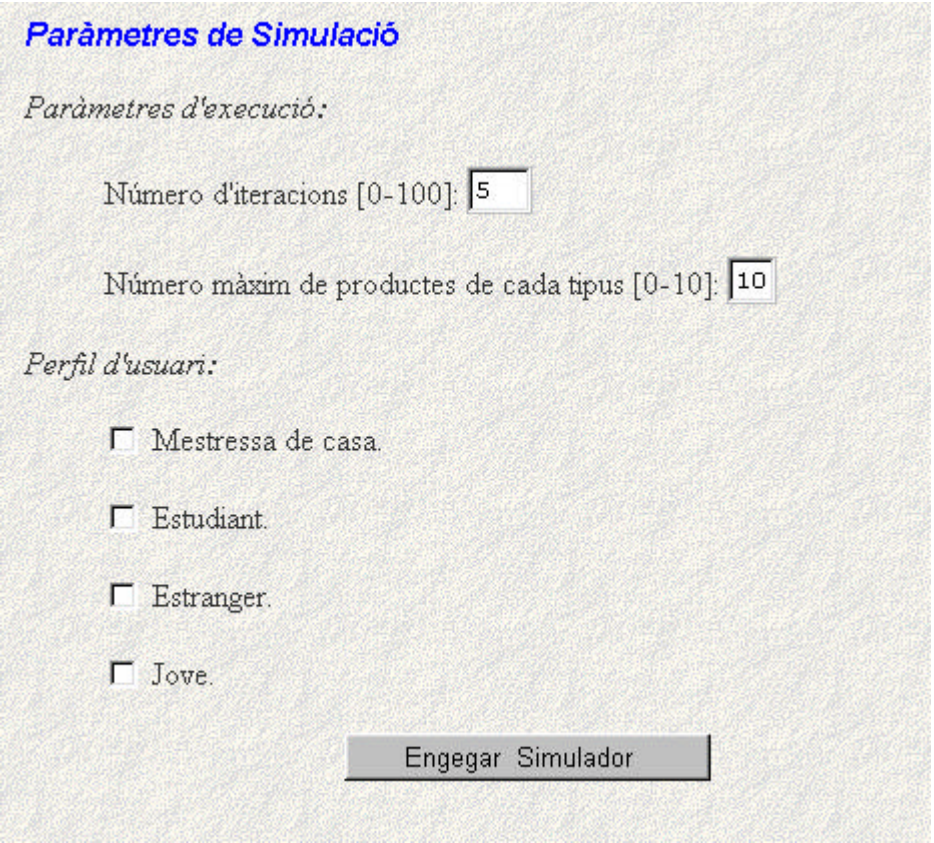

Fig5.1. Pantalla principal del simulador.

Un cop establert els paràmetres d'execució podem seleccionar el perfil d'usuari que volem que simuli. Cal dir que si no es selecciona cap perfil el simulador assumeix que crearà els casos amb totes les varietats possibles, es com si diguéssim tots els perfils possibles. Cal notar també que es poden fer combinacions de perfils i marcar més d'un perfil a l'hora, el simulador el que farà és fusionar els dos perfils a l'hora, resultant un nou perfil d'usuari.

El següent pas és polsar el botó de *Engegar Simulador*, on es començarà a executar el simulador en el servidor. Un cop s'hagi acabat l'execució ens retornarà el resultat a la nostra plana web.

El simulador per cada iteració ens mostrarà el passos que ha seguit el cicle de vida del CBR, i quan la funció d'excitació del sistema entri en funcionament mostrarà el seu valor.

Exemple d'execució del Simulador:

1. Creació d'un nou cas segons el perfil seleccionat i mostreig dels valors del cas.

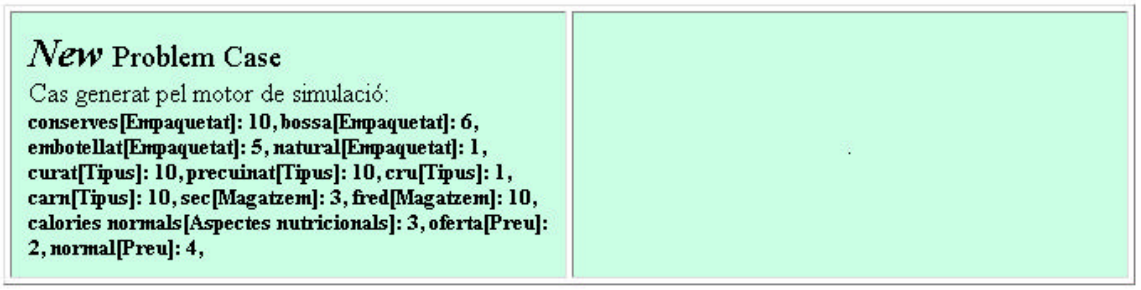

Fig5.2. Creació d'un Nou cas en el simulador.

2. Recuperació del cas de la base de casos amb la similitud més gran i mostreig del seu valors

Similarity	Retrieved Case Values	
76.0%	Producte[Merchandising]: Entrada cinema, conserves[Empaquetat]: 10, embotellat[Empaquetat]: 5, natural[Empaquetat]: 2, curat[Tipus]: 10, precuinat[Tipus]: 10, verdura[Tipus]: 10, cru[Tipus]: 8, carn[Tipus]: 7, peix[Tipus]: 9, sec[Magatzem]: 2, fred[Magatzem]: 5, baix en calories[Aspectes nutricionals]: 6, calories normals[Aspectes nutricionals]: 2, oferta[Preu]: 2, normal[Preu]: 3,	☑

Retrieve **Previous Case** La semblança màxima del cas generat amb els casos de la Base de Casos és de **76.06837749481201 %.**	

Fig5.4. Procés de Retrieve en el simulador

3. Execució de la funció d'excitació per decidir si es crea una nova solució o bé s'utilitza la solució recuperada.

	Threshold to excite the system Es supera de la mitja de les similituds de la funció d'excitació**67.4445228535554 %**. Ofertem el producte del cas recuperat.

Fig5.5. Superació del llindar de la funció d'excitació en el simulador

4. Si no es dispara la funció d'excitació llavors es mostra el procés de reutilització amb la solució del cas recuperat

Reuse **Previous Case** Producte ofertat pel sistema: **Entrada cinema**	

Fig5.6. Procés de Reuse en el simulador.

5. Si es dispara la funció d'excitació llavors es mostra la nova solució creada pel sistema per construir un nou cas.

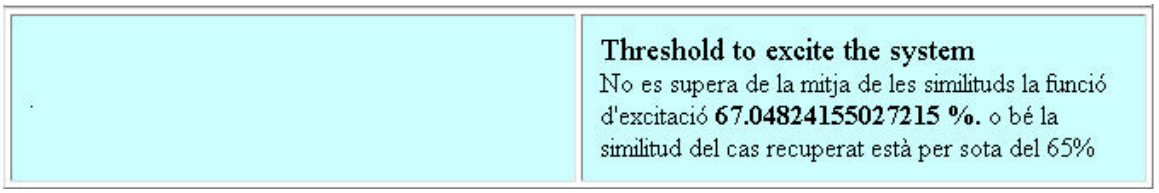

Fig5.7. No superació del llindar d'excitació en el simulador.

6. Si s'ha disparat la funció d'excitació llavors es mostra el procés d'excitació i la solució del cas generat, per tal de construir un nou cas.

Fig5.8. Procés d'excitació en el simulador.

7. Un cop hem creat el cas amb la nova solució, executem la funció d'acceptació de l'usuari per saber si el producte ofertat és acceptat.

Fig5.9. Procés de Revise en el simulador.

8. Si l'usuari accepta el nou producte llavors guardem el cas en la base de casos nom a un nou cas.

Retain **Previous + New** El sistema guardara el cas creat amb el nou producte ofertat com a un nou cas, ja quel'usuari ha revisat el cas i l'ha acceptat.	
Cas gravat amb èxit amb Id: **K34832501**	

Fig5.10. Procés de Revise i Retain en el simulador.

9. Si l'usuari no accepta el nou producte llavors simplement no guardem el cas.

	Revise **Previus + New** Resultat de la funció d'acceptació d'usuari és: NO
Not Retain **Previus + New** El sistema no guardara el cas creat ja que el producte ofertat no ha sigut acceptat	

Fig5.11.Procés de Revise en el simulador.

5.2 Construcció dels perfils d'usuaris

A l'hora de crear els casos en el simulador de la botiga no els crearem aleatòriament del tot. Si són totalment aleatoris no podríem estudiar si el sistema CBR proposa els productes correctament per als diferents tipus de clients, ja que no es veuria la classificació de clients.

Amb el CBR implícitament classifiquem els tipus de clients amb els productes que més s'avenen a ells. D'aquesta manera a l'hora de recuperar un cas d'una compra anterior, oferim el mateixos productes als mateixos tipus d'usuaris. Com hem dit, construir un simulador del comportament humà que ens simuli els productes que agraden més per tots els tipus d'usuaris és molt difícil. De fet si ho poguéssim calcular-ho aquest projecte no tindria sentit ja que sempre sabríem quin producte ofertar a un determinat perfil d'usuari. Així doncs crearem els casos amb uns valors dels atributs aleatoris i després aplicarem una factor de ponderació per acostar el cas cap a una classificació determinada. És totalment indiferent el valor de factor de ponderació, només ens interessa que en l'evolució del temps hi hagin conjunts de casos diferents.

Núm. de productes = número aleatori [0-1] * Factor de ponderació (Perfil, Atribut)

S'han establert quatre tipus de clients i per cada tipus s'ha establert un valor de ponderació que s'ha costi a la realitat el màxim possible. En la següent taula podem observar tots els valors establerts per la creació dels perfils d'usuaris.

		Atribut del Cas																	
	Factor De Ponderació	Conserves	Bossa	Embotellat	Natural	Curat	Precuinat	Verdura	Cru	Carn	Peix	Sec	Fred	Baix calories	Calories Normals	Alt calories	Luxe	Oferta	Normal
Perfil de clients	Mestressa Casa	1	1	1	3	2	0.2	3	3	2	2	1	2	0.2	1	3	0.2	3	1
	Estudiant	3	1	3	0.2	2	3	1	1	2	0.2	1	2	3	1	0.2	0.2	3	1
	Jove	3	1	1	1	3	2	2	1	1	1	1	1	2	2	0.2	0.2	1	3
	Viatger	0.2	1	1	3	3	0.2	0.5	1	1	2	1	0.2	0.2	3	3	3	0.2	1

Fig5.11. Taula de configuració dels perfils d'usuaris.

Per la lectura de la taula primer ens imaginem una persona que compraria diferents productes sense tenir cap preferència, això seria la creació d'un cas aleatori, després escollim un perfil de client (pex Mestressa Casa) i un atribut del cas (pex Oferta). Per últim llegirem la taula dient: Generalment les mestresses de casa compren 3 vegades més productes d'oferta que una persona sense cap preferència.

5.3 Evolució dels valors d'un cas

A continuació es mostra un exemple dels valors que poden tenir els casos. Els casos mostrats en la taula son seleccionats a l'atzar i poden ser originats per diferents perfils. Si ens fixem en la taula podem observar que hi ha valors que tenen un interrogant. Aquesta és una particularitat que te el JavaKate per indicar que el seu valor és desconegut, que és diferent que el seu valor sigui zero. Si el valor de l'atribut és zero, podem fer que aquest valor no es tingui en compte en el càlcul de la funció de similitud o bé podem establir un valor fix de la distància per un atribut desconegut.

El sistema creat en aquest projecte es permet construït manualment els nostres propis casos i executar la funció de similitud entre dos casos. Per posat un exemple podríem seleccionar un cas qualsevol de la taula següent posar els seu valors en el formulari i afegir el cas, o bé realitzar la consulta inductiva, o bé la consulta del veí més proper.

CAS	Producte	Conserves	Bossa	Embotellat	Natural	Curat	Precuinat	Verdura	Cru	Carn	Peix	Sec	Fred	Baix calories	Calories Normals	Alt calories	Luxe	Oferta	Normal
Cas t+1	Llonganissa	2	9	?	10	8	1	7	10	7	6	7	4	?	?	7	1	10	4
Cas t+2	Delícies de mar	10	7	4	9	10	6	8	3	5	3	5	7	7	1	1	1	1	9
Cas t+3	Torrons xocolata	?	2	3	5	10	?	3	2	6	10	8	1	1	10	10	7	?	5
Cas t+4	Camera fotos	5	1	9	10	5	?	5	7	10	10	9	3	?	7	10	?	10	2
Cas t+5	Magdalenes	?	4	3	1	10	10	1	8	3	8	5	8	10	10	1	1	4	4
Cas t+6	Magdalenes	?	4	2	3	1	3	?	5	1	1	8	2	?	7	2	?	?	?

Fig5.12. Evolució dels valors dels casos.

5.4 Representació de resultats

Després de cada execució del simulador aquest ens mostra una taula amb un resum dels resultats. En la taula apareix els valors que tenien les mesures abans de la simulació i els que tenen després de la simulació. Dels set resultats representats només tres aporten més informació de l'estat del sistema abans de l'execució de la simulació.

.	Abans de la simulació	Després de la simulació
Número de Casos Totals	77	114
Rendiment del Sistema[0-100]%	52.0 %	50.66666666666667 %
Bondat Total del Sistema[0-100]%	76.0 %	75.33333333333333 %
Similitud Màxima de la Simulació[0-100]%	-------	79.16666269302368 %
Similitud Mínima de la Simulació[0-100]%	-------	36.66665852069855 %
Temps de Retrieve Màxim	-------	60 ms.
Temps de Retrieve Mínim	-------	0 ms.

Fig5.13. Resum dels resultats del procés de simulació

Altres paràmetres com el temps de recuperació de cas o la similitud màxima i mínima, només tenen un caràcter merament informatiu.

Juntament amb la taula que resumeix el procés de simulació apareix els links als fitxer de dades amb els resultats de la simulació del perfil seleccionat i a més els links als fitxer de dades que contenen tots els resultats de totes les simulacions que s'han anat executant. Aquest fitxers de dades estan en format .csv i contenen dos columnes separades per comes de manera que el podem descarregar i tractar el fitxer i tractar les dades amb qualsevol editor.

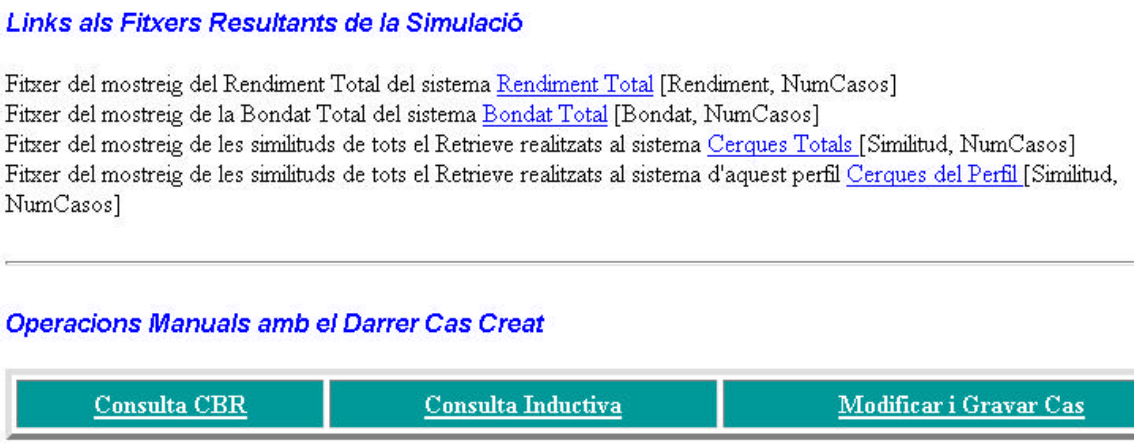

Fig.5.14. Links als fitxers de dades.

Ja per acabar el sistema construït dona la possibilitat de treballar amb el darrer cas creat pel simulador. Es poden fer tres operacions amb el cas.

1.- Modificar i gravar el cas dintre de la base de casos.
2.- Fer una consulta mitjançant l'algorisme del veí més proper
3.- Fer una consulta mitjançant l'algorisme d'inducció

El conjunt de tot el simulador resulta la primera eina que existeix en tot el món que és capaç de mostrar interacctivament el funcionament intern del CBR, i a més té la capacitat de proporcionar fitxers estadístics on hi consten les mesures de rendiment del sistema.

Quan es realitzen proves amb el simulador sovint desitjarem esborrar la base de casos, per començar des de zero i anar afegint casos per veure el seu funcionament. Per raons de seguretat

del servidor l'operació d'esborrar té accessos restringits. Si el lector desitja en algun moment esborrar la base de casos, preguem que es posi en contacte amb l'autor del projecte.

5.5 Rendiment del sistema

En aquest punt es representen el comportament del rendiment del sistema per diferents valors en els paràmetres d'execució. Per fer un estudi del comportament del sistema generalment el que farem serà anar fent simulacions de cent o de cinquanta iteracions i anar canviant el perfil de l'usuari. D'aquesta manera anirem fent execucions del sistema i observarem el seu resultat. Pel següent estudi partim de una base de casos buida.

Primer experiment:

1.- Execució:
Simulacions: 250
Número màxim de productes: 10
Perfil: Mestressa Casa

2.- Execució:
Simulacions: 250
Número màxim de productes: 10
Perfil: Jove

A continuació es mostra una taula amb el resum dels resultats de les dues execucions:

Simulacions Totals	550.0
Casos recuperats acceptats	130.0
Suma casos acceptats	224.0
Suma similitud acceptat	90.721929728984836

Fig5.15. Resum dels resultats del primer experiment.

El més interessant per estudiar el comportament del sistema és aprofitar la possibilitat que tenim de descarregar el fitxer de dades que contenen les mesures de rendiment i estudiar la seva evolució.

En la gràfica següent podem observar com el rendiment del sistema va oscil·lant fins que arriba a un punt d'estabilitat. A partir d'aquest punt el rendiment gairebé no varia.

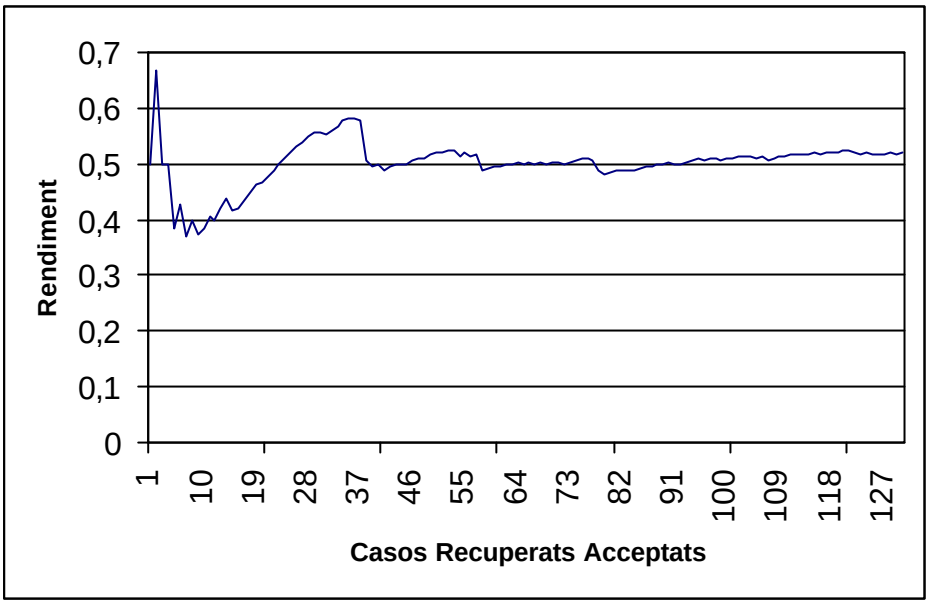

Fig5.16. Evolució del rendiment del primer experiment.

Anem a veure ara a estudiar un altre processos d'execució. Al igual que el punt anterior partim d'una base de casos buida per estudiar l'evolució de la base de casos.

<u>Segon experiment</u>:

1.- Execució:

Simulacions: 999

Número màxim de productes: 10

Perfil: Mestressa Casa

Simulacions Totals	999.0
Casos recuperats acceptats	544.0
Suma casos acceptats	788.0
Suma similitud acceptat	379.3711621761322

Fig5.17. Resum dels resultats del segon experiment.

Entre l'evolució del rendiment de la gràfica anterior i l'evolució d'aquesta podem veure clarament que existeix un comportament comú. Els valors de rendiment a partir d'una mida de la base de casos el sistema s'estabilitza. S'ha d'observat el mateix comportament del rendiment per moltes altres simulacions i podem concloure que amb aquest model una base de casos està entrenada quan la mesura de rendiment s'estabilitza.

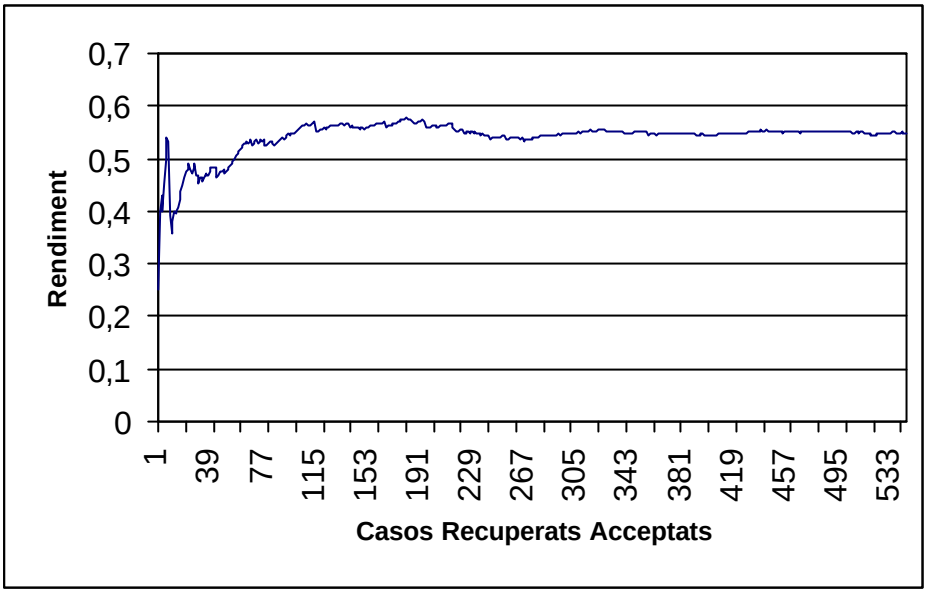

Fig5.18. Evolució del Rendiment del Sistema en el segon experiment.

Observem també que es poden donar situacions en el sistema oscil·la inicialment per sobre del punt d'estabilització per després estabilitzar-se en ell, o bé pot començar a oscil·lar per sota del punt d'estabilització. Cal notar que quan l'oscil·lació està per sota del punt d'estabilització el sistema necessita més casos per arribar al punt d'estabilització, per tant podem dir que el sistema té una resposta lenta. En el primer experiment el sistema s'estabilitza més ràpidament.

Dels experiments realitzats podem concloure dos punts molt importants que ens aporten una metodologia sistemàtica a l'hora de construir sistemes CBR:

- Podem establir una analogia entre els sistemes de control automàtic i els sistemes CBR.
- El número de casos requerits en un CBR per poder dir que aquest està entrenat és el punt d'estabilització de la mesura presentada com a Rendiment del Sistema.

Passem ara representar l'evolució de la mesura de bondat del sistema per al segon experiment. La bondat del sistema per definició sempre estarà per sobre de la mesura del rendiment del sistema. El que si podem observar es que la mateixa tendència que el rendiment, ja que la mesura de bondat conté els casos que han sigut recuperats i acceptats a més del casos que han sigut proposats pel sistema.

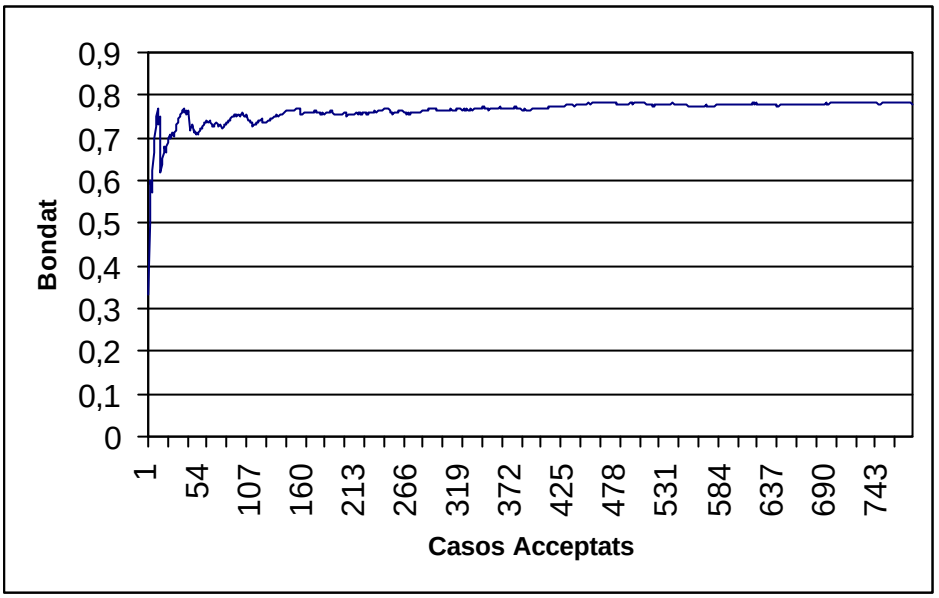

Fig5.19. Evolució de la Bondat del Sistema en el segon experiment.

El que podem concloure de les observacions d'aquesta mesura és el comportament del nostre mecanisme d'excitació. Donem doncs com a conclusió que:

- Per saber si un mecanisme d'excitació és correcte els valors de bondat sempre han d'estar per sobre dels valors del rendiment, de tal manera que si els valors de bondat són els mateixos que els valors de rendiment el mecanisme d'excitació no és vàlid.

Amb aquest estudi podem establir una cota inferior per a la bondat del sistema, però el que no podem establir és quin seria és la cota superior de la bondat per deduir el millor mecanisme d'excitació. Per l'experiència que tinc amb les desenes els processos de simulacions que fet recomano desconfiar d'aquells valors de bondat que s'acostin a 1. Això es justificarà en el següent punt quan es parli de la sobreexcitació del sistema.

A continuació es mostra l'evolució de la mida de la base de casos que segueix un comportament lineal en la majoria de les simulacions observades.

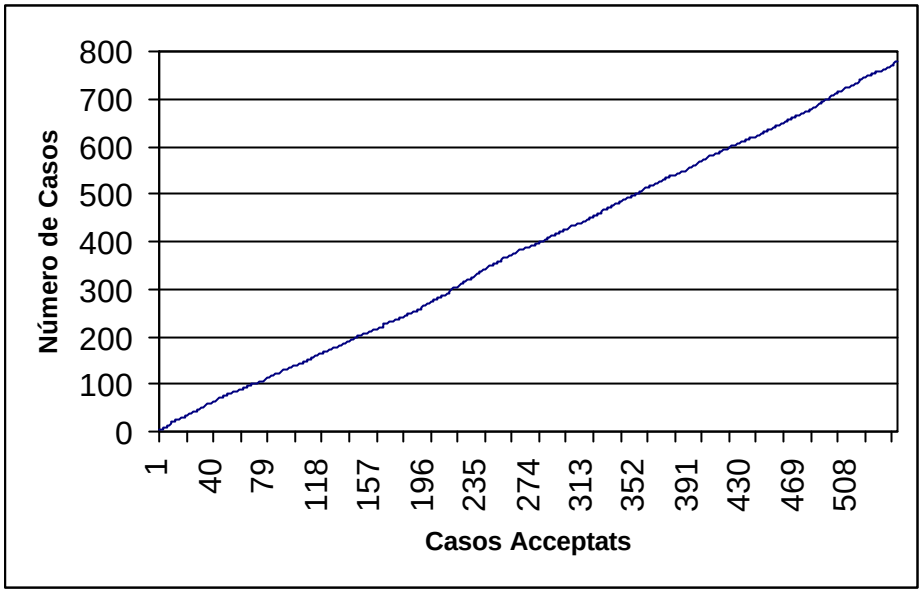

Fig5.20. Evolució de la mida de la base de casos.

Per últim mostrem un altre experiment on es representa el valor de la funció de similitud que s'aplica en el procés de Retrieve.

<u>Tercer experiment</u>:

1.- Execució:

Simulacions: 200

Número màxim de productes: 10

Perfil: Jove

2.- Execució:

Simulacions: 130

Número màxim de productes: 10

Perfil: Viatger

3.- Execució:

Simulacions: 300

Número màxim de productes: 10

Perfil: Cap

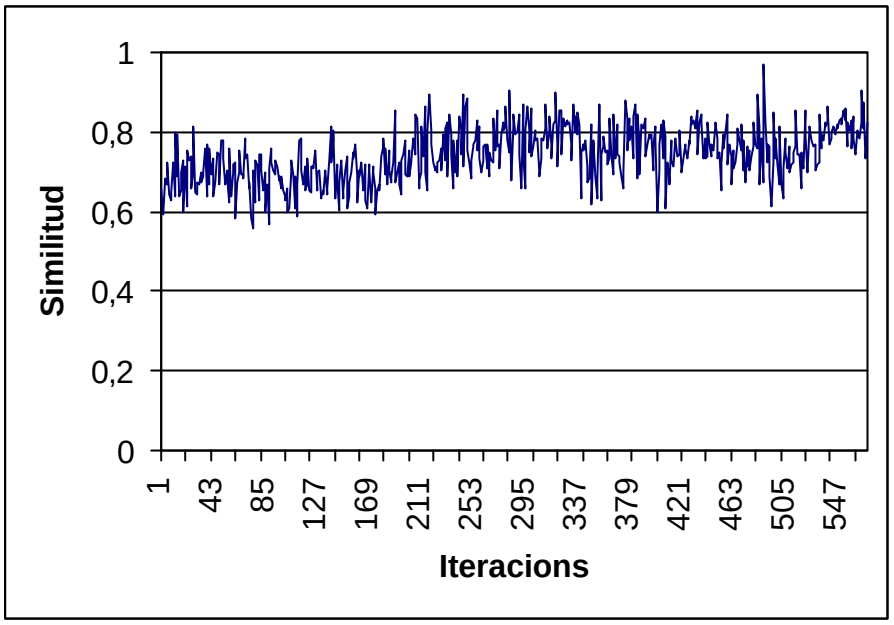

Fig5.21.Evolució dels valors de la similitud

Segons les proves realitzades podem dir que al afegir més varietat de perfils d'usuaris la variabilitat de la funció de similitud és més petita, o dit d'una altre manera tenim prou casos dintre de la base de casos com per trobar-ne algun que sempre s'assembli. El que no es pot afirmar amb aquest sistema és que al afegir més varietat de casos el rendiment del sistema augmenta.

5.6 Relaxació i Sobreexcitació del sistema

Per acabar l'estudi del projecte destacarem la importància que té escollir una bona funció d'excitació. Si no tenim una funció d'excitació del sistema ens trobem que el sistema es no aporta nous productes a l'usuari i com que l'usuari no ens dirà mai quins productes li agraden més, implica que el coneixement de la base de casos no evolucionarà. Del concepte en diré que el sistema està relaxat ja que no s'aplica cap funció d'excitació.

Per determinar si el nostre sistema està relaxat ho farem observant les mesures de Rendiment i Bondat del Sistema. Quan un sistema està relaxat sempre es proposaran els mateixos productes i com que els productes que tenim a la base de coneixement sempre seran acceptats per al

mateix perfil d'usuari, ens resulta que els valors de Rendiment i Bondat del Sistema sempre s'acosten ràpidament a 1.

Detectar la relaxació del sistema per a simulacions que generessin casos de diferents perfils d'usuari consistiria, en observar l'acostament dels valors de Rendiment i Bondat del Sistema a 1, per a un nombre de casos proporcional al número de perfils generats.

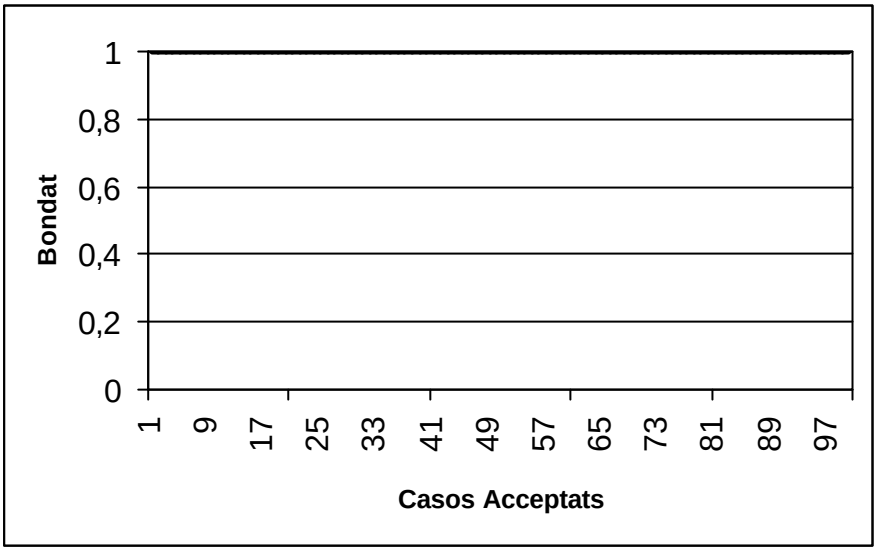

Fig5.22.Evolució de la bondat per un sistema relaxat.

En el nostre sistema ens pot passar també que el sistema es dispari molt sovint la funció d'excitació, el resultat d'això es que el sistema està sobreexcitat i el Rendiment del Sistema baixi en picat.

És important recordar destacar que la Bondat del Sistema no cau a la mateixa velocitat que el rendiment, ja que és una mesura que depèn si l'usuari ha acceptat el ca o no, independentment si el cas ha sigut un cas provenint del procés de Reuse o bé provenint del procés d'Excitació. Per tant no farem servir la mesura de Bondat per determinar si el sistema està sobre excitat.

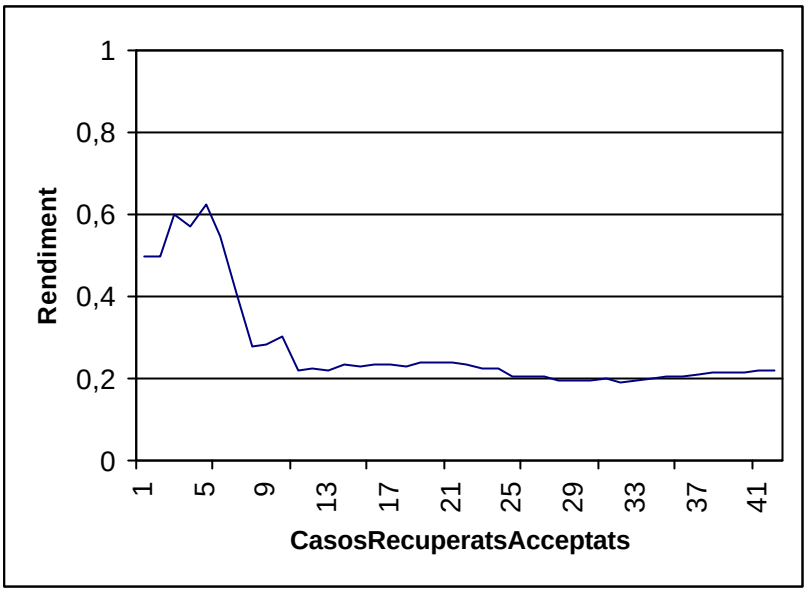

Fig5.23.Evolució del rendiment per a un sistema sobreexcitat.

En la següent taula es resumeix quines mesures s'han de estudiar per saber si el sistema està relaxat o sobreexcitat:

Sistema Relaxat	La Bondat del Sistema ràpidament s'acosta a 1.
Sistema Sobreexcitat	El Rendiment del Sistema ràpidament s'acosta a 0.

Fig5.24.Criteris per la detecció de relaxació i sobreexcitació del sistema.

5.7 La botiga virtual

S'ha intentat que la botiga virtual construïda s'acosti el màxim possible a la realitat. La botiga consta de tres seccions de productes. Una primera secció de productes generals on hi consten els productes de consum més comú. En aquest projecte es volia donar una importància major als productes curats, per tant s'ha creat una segona secció per a embotits. Per últim en la tercera secció podem trobar productes varis, com poden ser begudes alcohòliques, bombons, etc. El client que desitgi comprar productes només ha de marcar el *checkbox* del producte i després fer un click al botó Fer compra. Tot seguit ens apareixerà una nova pantalla amb una sèrie de productes recomanats. Per comprar algun dels productes recomanats només cal prémer el botó que conté el nom del producte

La mesura de quantificació de satisfacció dels usuaris que compren en la botiga es mostra en el valor del Rendiment de la Botiga. Aquest valor es mostrat en temps real per la botiga virtual cada vegada que realitzem una compra. El valor de Rendiment de la Botiga és va calculant per totes cada persona que compra en la botiga per tal tenim un procés que ens detecta si la persona ens ha comprat un producte proposat per la botiga.

La botiga virtual recupera tres productes que s'adaptin més amb el perfil de l'usuari, i a més mostra el valor de recomanació del producte. El valor de recomanació és el valor de la funció de similitud entre el cas de la compra recuperat i el cas actual de la compra de l'usuari. Es pot donar que dels tres productes recuperats per ser ofertat n'hi hagi dos d'iguals entre ells, en aquesta situació la botiga només mostrarà el producte una sola vegada amb la certesa més alta que ha trobat.

Un cop presentat els productes a l'usuari si aquest en compra un, el sistema guardarà la nova compra realitzada amb el producte ofertat com a un nou cas. Observem que totes les operacions que s'estan fent en el sistema, són les indicades en el nou cicle de vida proposat en aquest projecte.

Abans d'afegir el nou cas de la compra mirem si la similitud del cas és més gran del 90%, per què si es més gran no afegirem el cas per que sinó tindríem dos casos iguals en la base de casos, i això causaria redundància en la base de casos.

A més la nostra botiga al igual que el simulador d'aquesta, té la capacitat d'excitar a l'usuari segons el llindar de la funció d'excitació. Quan l'oferta és d'un producte provenint del procés d'excitació apareix al costat del producte (*Excitació*).

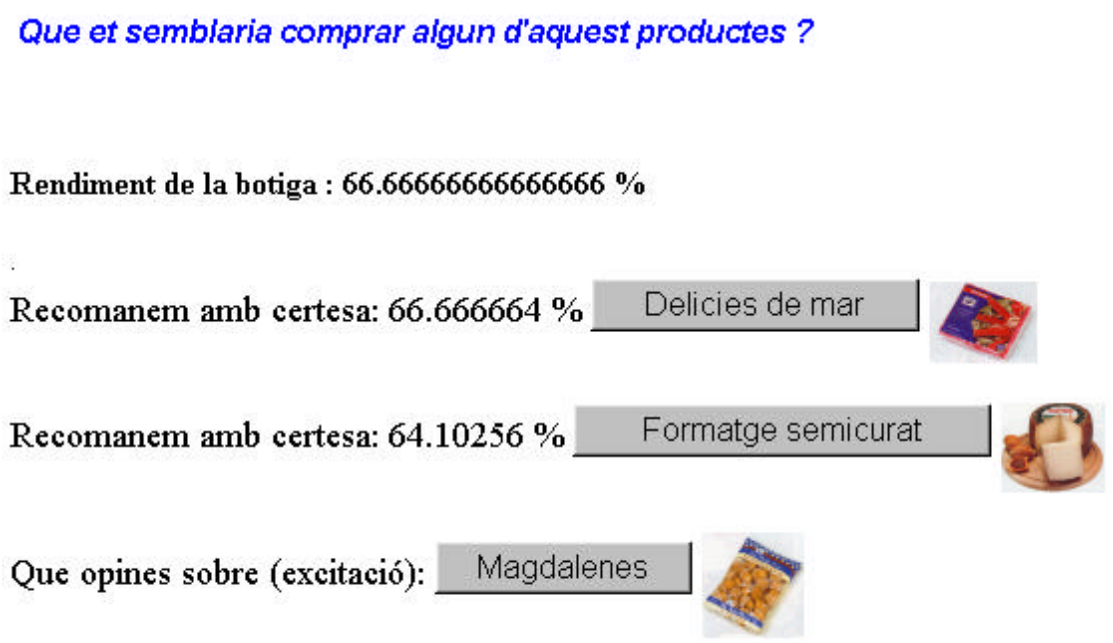

Fig5.25.Oferta de productes i serveis per la botiga virtual.

Després de moltes simulacions i proves canviant el llindar i la funció d'excitació podem concloure que per aquest model obtenim un bon funcionament establint la següent funció d'excitació:

- Si la semblança del cas està per sota del 65%
- Si la semblança del cas recuperat es més petita que el 90% de la mitja dels casos que han sigut acceptats.
- Si no es dona cap situació anterior proposem aleatòriament un producte en un 30% de vegades.

Amb el primer punt arreglem la situació que es dona quan és un cas molt dolent, per tant és millor donar una solució aleatòria abans que donar una solució dolenta. Hem considerat que un cas és dolent quan la seva funció de similitud està per sota del 65%.

Amb el segon punt el que estem fent és dibuixar una linea de la funció d'excitació que està per sota de l'evolució de la mitja de les similituds. Si el valor de la similitud del cas recuperat està per sota d'aquesta línia excitarem el sistema. Si és que està per sobre llavors estem dient que el cas recuperat té una similitud que és prou bona com per ser proposat.

Aquest punt és molt important ja que si la mitja de les similituds baixa, baixa el llindar d'excitació, per tant el nostre sistema no és tant exigent amb el valor de la similitud del cas. En canvi quan la mitja de les similituds va pujant, o sigui que el nostre sistema cada vegada va recuperant solucions millors, la funció és molt més exigent. La funció d'excitació farà que només es proposin els productes amb una gran qualitat de certesa de que serà acceptat.

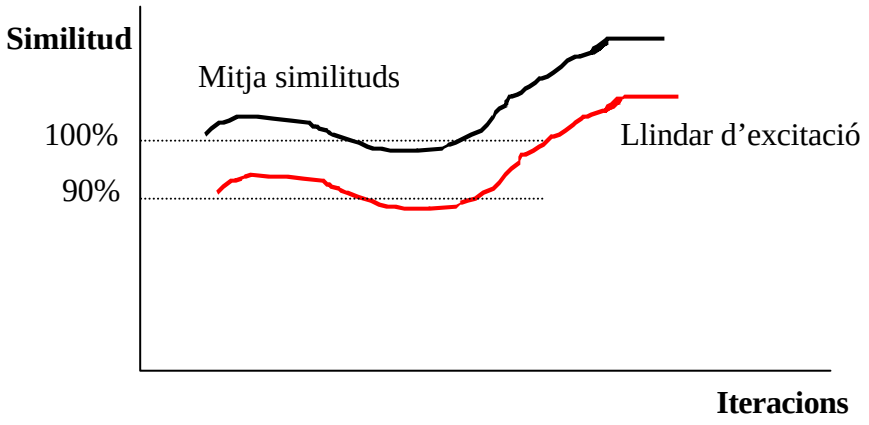

Fig5.26.Representació del llindar de la funció de similitud.

En el tercer punt només està implementat per a la botiga virtual, ja que els usuaris que compren en aquesta són de moment pocs i per tant les excitacions que es produiran seran poques. Aquest llindar el que ens permet es donar una dinàmica més gran d'ofertes per a un conjunt reduït d'usuaris.

6 Conclusions

Basat en els experiments realitzats i resultats observats podem concloure el següent:

- Consultant en una base de casos d'observacions de compres anteriors i comparant una nova compra d'un usuari fent servir l'algoritme del veí més proper, es pot ofertar un nou producte a l'usuari sabent que existeixen unes grans possibilitats que aquest sigui acceptat.

- La creació d'un nou procés en el cicle de vida en la tecnologia del Raonament Basat en Casos que ens fa evolucionar la base de casos. Entre el procés de *Reuse* i *Revise* s'ha introduït el procés d'*Excitació*.

- Respondre a una de les grans preguntes que es fan actualment en els sistemes CBR. Extrapolat en aquest projecte es tracta de saber el número de compres que necessita emmagatzemar un sistema CBR per a que una oferta tingui èxit. Concloem que estudiant les mesures de rendiment que en aquest sistema es proposen, el número de compres que calen és el que s'observa quan la mesura del rendiment s'estabilitza.

- Establir punts sistemàtics per a la construcció de tot sistema CBR aplicat al comerç electrònic: El simulador de la botiga virtual i les mesures de rendiment. El simulador és totalment necessari degut a l'estudi i modificacions que calen fer en el sistema abans de que aquest tingui un funcionament òptim.

- El simulador construït és una de les primeres eines que existeix que és capaç de mostrar interactivament el funcionament intern del CBR dintre del comerç electrònic. A més té la capacitat de proporcionar els fitxers estadístics on hi consten les mesures de rendiment del sistema resultants de les execucions de cada simulació.

- El sistema presentat pot solucionar altres punts dintre del camp del comerç electrònic: Publicitat personalitzada als diferents tipus de clients. Promoció de nous productes, productes en oferta o rebaixa per als perfils de clients potencialment compradors. Servei d'ajuda personalitzat al client per seleccionar els productes que més s'adapti a ell.

Referències

[Aamodt – Plaza] Aadmot A. , Plaza (1994). Case Base Reasoning: Foundational issues, methodological variations, and system approaches. AI Communications.

[Klaus – Brigitte, 1996] Klaus-Dieter Althoff, Briggite Bartsch-Sporl (1996). Decision Support for Case Base Applications. Centre for Learning Systems and Applications, Department of Computer Science. University of Kaiserslautern, Germany.

[Wolfgang – Ralph – Stefan] Wolfgang Wilke, Ralph Bergmann , Stefan Wess. Negotiation During Intelligent Sales Suport with Case Base Reasoning.TECINNO GmbH and Centre for Learning Systems and Applications, Department of Computer Science. University of Kaiserslautern, Germany.

Agraïments

En primer lloc voldria donar les gràcies al meu director de projecte per haver confiat en mi a l'hora de proposar-me el treball i pel seu punt de vista professional.
Voldria agrair el suport tècnic rebut pel personal d'Acnosoft especialment al Dr. Erci Auriol i al Stephane loubry. A la Carme d'administració de la EPS per facilitar-me el tràmits de papers.
Gràcies també a la meva xicota Sussi a l'hora d'escanejar les fotos de la botiga, i pel seu gran suport. Un milió de gràcies als meus pares i al meu germà per ser com son.

Annex A – Contingut del CD-ROM

El contingut del CD-ROM és una còpia exacta del que es pot trobar en el meu espai web: http://eia.udg.es/~tmartinz , però s'ha creat aques disc per facilitar l'accès als fitxers.

En el directori arrel del disc hi ha les planes en format html, per a la plana principal, per a la botiga virtual i per al simulador.

A continuació es detalla en contingut dels directoris:

Codi: Conté el codi del l'aplicació.
Model : Conté el model de dades del CBDR
Model\Images\Productes: Conté les imtatges dels productes ofertats.
Images: Conté les imatges que es fan servir en les planes web.
Images\Productes: Conté les imtages dels productes a la venda de la botiga virtual.