

29.10.04-16.11.04

ENTORN INTERACTIU CONTROLAT PER AGENTS INTEL·LIGENTS PER RECOLLIR CONEIXEMENT POPULAR SOBRE PLANTES AROMÀTIQUES

ABSTRACT

En aquest article es presenta una perspectiva de les tecnologies de la informació que mostra el paper fonamental que poden tenir en recollir i preservar el coneixement popular que tenen les persones grans. Usuaris sense gaires coneixements tecnològics, poden cercar o deixar informació sobre algun tipus de coneixement. L'entorn es basa en el concepte de Swarm Intelligence, això és, l'ús de la metàfora del comportament de les colònies d'insectes sobre una plataforma d'intel·ligència artificial distribuïda (agents). En particular, s'utilitza la metàfora del comportament de les abelles. Gràcies a aquests agents abelles, i a partir de les aportacions dels usuaris, el coneixement emergeix del sistema a partir d'un procediment de validació. Tot i que el sistema és aplicable a la recollida de qualsevol coneixement popular, a l'hora de provar la validesa de l'entorn ens hem centrat en la temàtica de plantes aromàtiques i medicinals.

1. Introducció

La nostra imatge, és la de l'àvia que tornant de l'escola amb el net pel camí de casa, a pagès, el nen es queixa de mal de panxa i l'àvia recull unes herbes mentre amorosa li diu "quan arribem a casa, l'avia et farà una taceta d'aigua i ja veuràs que en passarà de seguida". I potser, en la resta de camí cap a casa, li explicarà que quan ella era petita, una germana de la seva àvia, feia uns recuits boníssims amb preona, una herba que es fa sobre aquell camp de girasols, on aleshores només hi sembraven l'ordi. Li sembla que encara se'n recorda, però no sap segur si feia bullir l'herba i posava una mica d'aquell suc a la llet, o la hi posava directament,... i,... li sembla que... Bé, el que és segur és que li quedaven molt bons, molt diferents, i certament molt més bons que els que ara es compren. El nen, que com tots els néts, és el més espavilat del món, li explica a l'àvia que navegant per una pàgina web, potser pot trobar aquell detall que no recorda, i moltes més coses d'aquestes. Ah, i que si vol, també pot afegir-hi, a la xarxa, totes aquestes coses que només ella sap. Potser, àvia i net negociaran el "jo t'explico aquelles coses més secretes que sé d'aquestes plantes si tu m'expliques com buscar el que ara no recordo per això que... com en dieu? web? internet?".

És ben conegut que "els grans" són dipositaris d'una gran quantitat de saber i coneixement i, que si bé és cert, que gran part d'aquest coneixement l'han tret de llibres o d'un entorn cultural accessible, no és menys cert que n'hi ha encara una part molt important que només s'ha tramés de boca a orella creant un substracte de coneixement popular. Aquest coneixement no està escrit en lloc i només es pot recollir escoltant el que poden explicar-nos els nostres "grans" (Carpó, 1994).

Com es pot intentar recollir i preservar el coneixement popular que tenen les persones d'edat? Des de la nostra perspectiva, creiem que les tecnologies de la informació han de jugar un paper fonamental en aquesta tasca social. Amb aquesta hipòtesi, la nostra recerca ens ha portat a la realització d'un entorn informàtic, on usuaris sense gaires coneixements tecnològics, puguin cercar o deixar informació sobre algun tipus de coneixement. Tot i que el sistema és aplicable a la recollida de qualsevol coneixement popular, a l'hora de provar la validesa de l'entorn ens hem centrat en una temàtica, la etnobotànica.

Pretenem que qualsevol persona, pugui publicar o recollir informació sobre plantes

RECER

GT - 1
Tecnologies i
Antropologia
Cultura, coneixement
i medi

IMPRIM

Comunicacions

Presentació

Fórum de

Conclusió

versió original

versió traduïda

Autor/

Martí Fàbregas
Beatriz López
Maria de los

Josep Mas

Paraules

cibercultura
coneixement | |
coneixement | | s
infor

Comunicacions
relacionades

Compartir audècia
competir. La inter
interoperativitat er
d?història n

Cibercultura, si
pensament cont
Explorant les pos
conceptuals d'un
emerge

Moviments Socials
en Xarxa

Narrativa hiperme
nous contingu
ciberno

aromàtiques o plantes medicinals deixant-la accessible o accedint a ella respectivament a través de la xarxa via pàgines web. La informació que es pot trobar, o deixar, és bàsicament informació textual o gràfica, però no defugim de posar-hi qualsevol altre tipus d'informació multimèdia més endavant. Aquesta informació es processa mitjançant tècniques utilitzades en Intel·ligència Artificial, concretament amb tècniques basades en el paradigma conegut com a *Swarm Intelligence*. Aquestes tècniques ens permeten validar la informació sobre la que es treballa, tan en la fiabilitat intrínseca de la mateixa informació com en la dels informadors que la proporcionen. D'aquesta manera, podem parlar d'un procés de transformació de la informació inicial en coneixement.

Aquest article està organitzat de la següent manera. Presentem inicialment els conceptes de *Swarm Intelligence* que són la base del nostre model. Seguim a la tercera secció amb un estudi de l'arquitectura que ens permet detallar pas a pas la construcció de la nostra aplicació. En la quarta secció fem una anàlisi de les tècniques que hem fet servir per a validar la informació sobre la que es treballa. En la cinquena secció, exposem els resultats experimentals del que hem fet. Acabem l'article amb un apartat de conclusions on recopilem el treball exposat, la seva realització, i els projectes de futur que hi tenim previst.

2.- Un model basat en 'Swarm Intelligence'

Els biòlegs han estudiat diferents teories relacionades amb l'experiència de milions d'anys d'evolució sobre les actuacions individuals i col·lectives de les colònies d'insectes, però fins que no s'han desenvolupat eines computacionalment potents no han pogut verificar els models ni fer simulacions de la complexitat biològica. Com a conseqüència dels resultats obtinguts en les simulacions, els informàtics tornem a girar els ulls vers la natura, per intentar entendre com és possible que unes formigues, que individualment mostren un comportament simple, tinguin les capacitats de, col·lectivament, cultivar fongs, organitzar camins o fins i tot fer una estructura tan complicada com pot ser un formiguer amb els problemes d'enginyeria que comporta la ventilació o l'evacuació d'aigües pluvials, per exemple.

L'adopció de la metàfora del comportament social dels insectes per resoldre problemes complexos és el que s'ha donat a conèixer com a *Swarm Intelligence*.

Swarm Intelligence podríem traduir-ho com a Intel·ligència d'Eixam, i és el nom que es dona a una de les línies de recerca amb més projecció dins del camp de la Intel·ligència Artificial (Muntaner, 2003; Bonabeau & Théraulaz, 2000; Bonabeau et al. 1999). *Swarm Intelligence* es refereix a la capacitat de resoldre problemes complexos amb senzilles interaccions. Les colònies d'insectes ho fan així. Sense supervisions generals, fan un treball distribuït on cada individu es preocupa d'una part de la feina, la seva, i només fa la seva. De la feina de tots els components en surten les solucions als problemes complexos de la comunitat. D'aquestes solucions col·lectives, on cada un dels individus fa només la seva part de feina, se'n diu que emergeix una intel·ligència col·lectiva. Interpretem la intel·ligència emergent com una capacitat intel·lectiva basada en interaccions senzilles.

El nostre sistema, inspirat en el model de *Swarm Intelligence*, vol obtenir coneixement emergent de la informació que proporcionen els usuaris del sistema. Per implementar el model proposem l'ús d'una tecnologia aplicada des de fa alguns anys en el camp de la informàtica distribuïda i coneguda com a Tecnologia Agent.

És difícil trobar una definició d'agent, perquè se li ha donat diferents interpretacions depenent la precisió que se li ha volgut donar, dels límits que se li han assignat, inclòs de la visió específica que en té cada autor (Weiss, 1999). A nivell semàntic, ja es veu que el nom d'agent va referit a alguna cosa que actua, o ampliant el concepte, que pot actuar en representació d'altres (Wooldridge, 1999). Alguns agents existeixen en el món tangible, com per exemple els robots que actuen sols o representen algú en les seves actuacions. Però som en el món virtual de la informàtica i podem parlar d'agents software que existeixen en un món virtual, el món dels sistemes operatius, de les xarxes, de bases de dades, un món que només existeix en l'essència de la unitat central de procés d'un computador. Aquests agents

Cibercultur@ com de comunicació cc de la perit

Un Perfil Professi Coneixement en la la Informac documentalista

Impactes de l'Or Tractament de la en el Servei

L'Observatori Ci Ciutat Mediterràni: tecnologies de la i servei de la cor científic

El capital intel·lec suport a la gestió de l'organització desafiaments de la coneixem

Disseny i posada a la RIVI

'Societat del con identitat persona (im)materialitat, unitat/fragmentac prensió/constru subject

Indicadors ciber popularitat: anà revistes biomèdiq en l'Índex Mexicà Biomèdiq Latinoameri



de software actuen com a tasques, processos i usuaris o poden actuar en representació seva. És aquesta faceta dels agents software la que hem adoptat nosaltres: agents que recullen, validen i presenten informació sobre una temàtica.

En particular, nosaltres, al parlar d'agent de software ens basarem en la definició més clàssica del concepte proporcionada per Jennings i Wooldridge i Jennings (Jennings & Wooldridge, 1998). Aquests autors defineixen un agent com un sistema hardware o software amb les propietats d'autonomia, sociabilitat, reactivitat i proactivitat. La primera propietat, la d'autonomia, la refereixen als sistemes que tenen un cert grau de control de les seves activitats i del seu estat intern sense intervenció humana. La sociabilitat va referida bàsicament a la possibilitat de comunicació entre ells o amb els humans a través d'un llenguatge propi. La reactivitat vol tenir un sentit més sensorial, ve a ser la capacitat de percebre el seu entorn i poder actuar-hi quan rep algun estimul. Finalment, la proactivitat s'interpreta com la capacitat de l'agent de decidir les accions a realitzar o si cal prendre la iniciativa per a assolir els seus objectius.

El nostre model, doncs, pren com a partida el coneixement sobre el comportament dels eixams d'abelles i l'associació de l'eixam a un sistema d'agents. Cada un dels seus individus (abelles obreres, abellots, reines) mitjançant la tecnologia agent serà un element del nostre muntatge per a desenvolupar un sistema distribuït de recollida de coneixement popular (veure figura 1). L'entorn resultant l'hem anomenat En_C_Prou (**Entorn de Coneixement Popular Recollit a la Web** (Masanas, 2003).

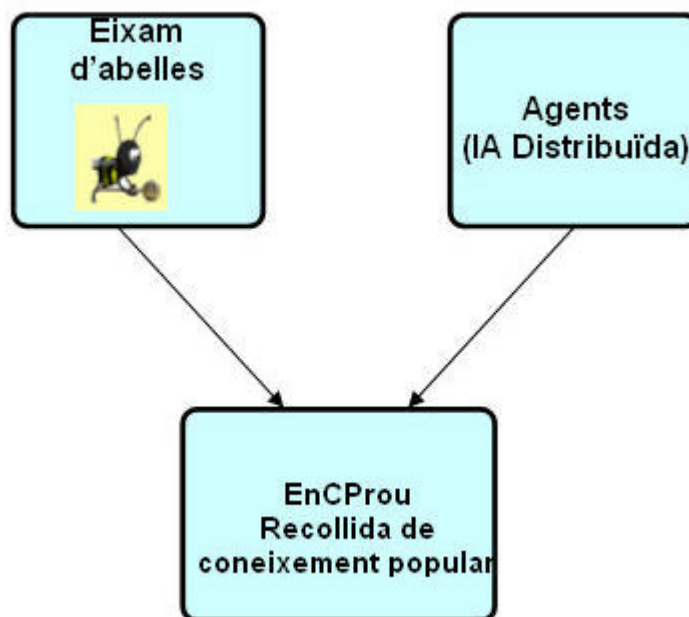


Figura 1. Tecnologia utilitzada en l'entorn d'adquisició de coneixement popular EnCProu

3.- Arquitectura

Situant-nos en l'àmbit de la Swarm Intelligence, hem generat un model propi basat en el comportament de les abelles. Definim una comunitat d'abelles (agents) que busquen nèctar (informació no confirmada, i susceptible d'ésser-ho) per a convertir-la en mel (la informació ja verificada) (veure figura 2).

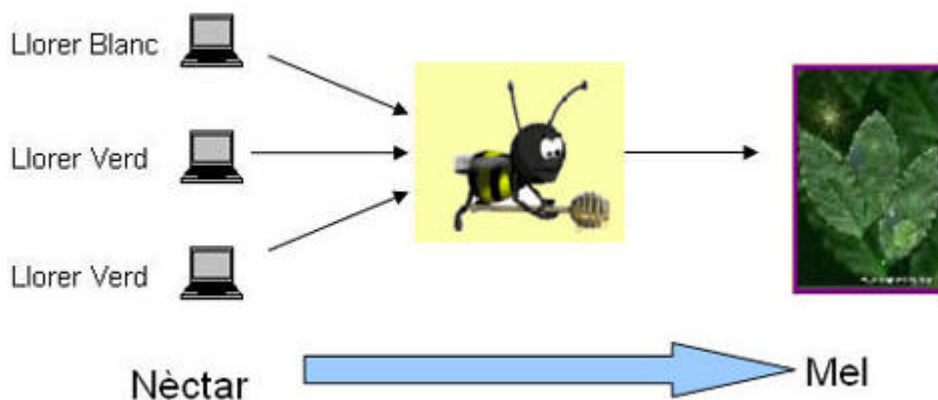


Figura 2. Símil del comportament de les abelles en la generació de coneixement sobre plantes aromàtiques i medicinals.

En el nostre sistema distingim els següents components bàsics:

- ⌘ Les abelles que són agents reactius simples que actuen segons un caràcter predefinit. Anàlogament que hi ha diferents tipus d'abelles (reina, obreres, abellots, ...), cadascuna amb una funció ben definida, també en EnCProu distingim diferents tipus d'agents.
- ⌘ L'apicultor que és l'usuari humà que actua sobre el sistema encarregats de proporcionar la matèria primera inicial, i confirmar la qualitat de la mel (o sigui el coneixement generat sobre les plantes)
- ⌘ L'agent relacions públiques, creat amb la finalitat de controlar el nivell de veracitat dels usuaris en funció de les aportacions pròpies i les validacions d'altres usuaris.
- ⌘ El rusc que és la base de dades on s'emmagatzema la informació.

En la figura 3 es mostra un diagrama de tots aquests components.

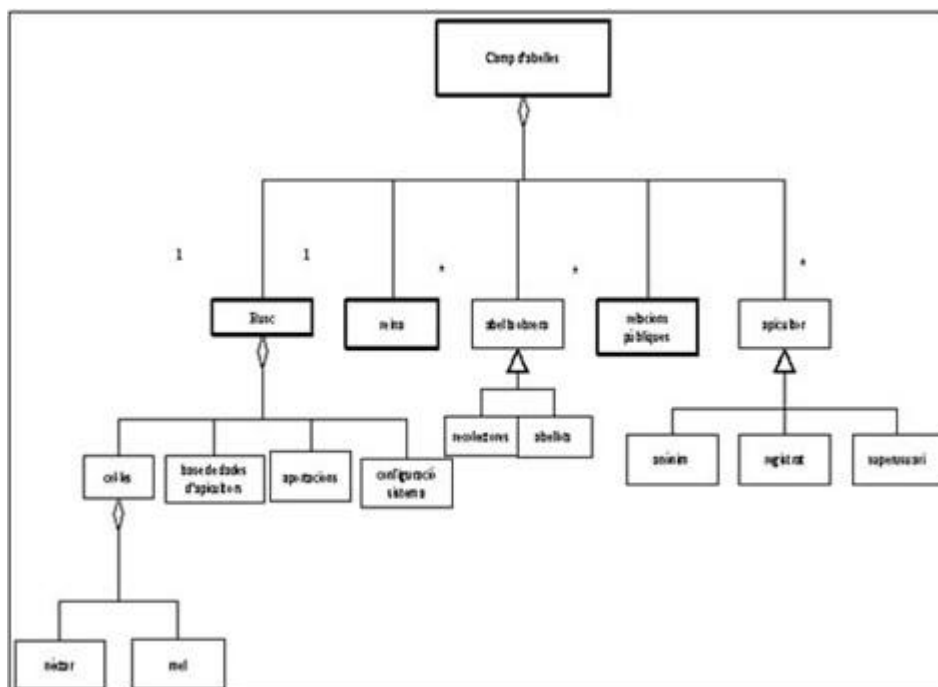


Figura 3. Diagrama dels components del sistema.

El sistema sempre parteix de la base de que els apicultors hi participen de bona fe a l'hora de mantenir la informació del rusc. El mecanisme d'enriquiment es basa en establir graus de similitud entre plantes per a relacionar o inferir nous coneixements. És a dir, el model parteix del supòsit que un usuari pot haver deixat una certa informació sobre una planta (timó), però que no ha estat capaç de relacionar amb una altra planta amb importants graus de coincidència (farigola) perquè, per exemple, desconeix la terminologia que ha fet servir l'altre usuari. Aleshores, el sistema busca plantes que tenen un grau elevat de coincidències i infereix un cert coneixement a partir de tota la informació rebuda. Aquest coneixement és validat a posteriori pels apicultors experts. Només el coneixement validat passa a formar part del cos del coneixement contingut en el sistema.

A més a més, el sistema manté enllaços entre plantes semblants per a facilitar -ne el creuament de coneixements. Això facilita l'accés dels usuaris al coneixement del sistema, estalviant temps en la revisió d'informacions donades per altres usuaris, afavorint la incorporació de la informació, i oferint la possibilitat de permetre als usuaris finals una consulta ràpida sobre determinades qualitats d'una planta. Cal remarcar, que el coneixement contingut en el sistema no és estrictament acadèmic, i per tant, les informacions proporcionades, s'han d'interpretar sempre amb la confiança i reserva que correspon a qualsevol tipus de coneixement popular.

El model d'agents s'integra dins d'un sistema d'informació que conté el model de presentació, desenvolupat sobre pàgines web, com es mostra a la figura 4. Els usuaris (apicultors) accedeixen a una pàgina web on poden realitzar les seves consultes o aportacions. L'usuari disposa d'un menú en el que pot triar entre fer una aportació sobre els temes existents en el sistema, entrar-ne de nous, o qualsevol altre de les accions que li estan permeses (consultar, verificar, etc.). En funció de l'acció escollida, l'usuari és guiat pas a pas, a través d'altres pàgines, i realitza l'activitat mitjançant eleccions amb el ratolí. El contingut de les pàgines web és dinàmic i representa la informació actual que hi ha en el sistema.

3.1 Agents abelles

La comunitat d'abelles constitueix un sistema multi-agent on cada abella presenta les característiques d'autonomia, sociabilitat, reactivitat i proactivitat pròpies d'un agent intel·ligent (veure la secció 2). Cada abella té una tasca assignada autònoma dins del sistema. Per exemple, un agent obrera té associada la tasca de recollir informació d'una planta definida en la base de dades. Així, a la planta romaní li associem l'abella obrera agent del mateix nom, romaní. Totes les abelles disposen d'un sistema senzill de comunicació per tal d'establir els contactes socials dins de la comunitat. El sistema ve coordinat i centralitzat per una abella reina. La reactivitat ve donada especialment en la reina quan desperta a una abella obrera davant de nova informació susceptible de ser validada en el sistema. La reactivitat en les obreres, s'estableix quan són despertades per l'abella reina i actua segons el rol que li ve donat pel context. Finalment, la característica de proactivitat l'exhibeix principalment l'abella reina quan dóna indicacions a les abelles obreres de la necessitat de cercar el nèctar mentre que en les obreres la trobaríem quan en un moment determinat, una obrera es desperta a instància de la reina, i abandona la seva parcel·la (registre de la base de dades) enduent-se amb ella la informació que disposa (nèctar) per anar a buscar informació comparable en les parcel·les de les companyes.

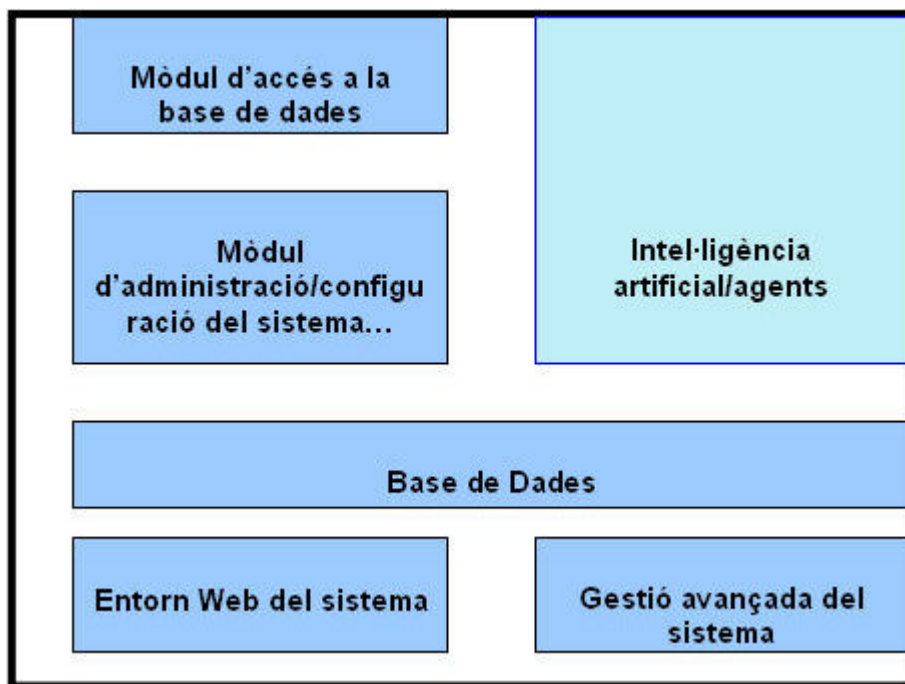


Figura 4. Model del sistema d'informació.

Abella reina és un agent "estàtic" que media entre la resta d'agents "dinàmics". Les funcionalitats bàsiques són les següents:

- ✗ Enviar senyals i rebre'n, pot sincronitzar les altres abelles.
- ✗ "Fer néixer" noves abelles, és a dir, crear noves plantes sota la instància d'un apicultor.
- ✗ "Despertar" i "adormir" les obreres, en funció de les aportacions rebudes i la necessitat de generar nou coneixement.
- ✗ Actualitzar el rusc amb la informació sobre els usuaris ("apicultors").
- ✗ Mantenir llistes d'aportacions rebudes dels usuaris per a que siguin consultades per les obreres.

Abella obrera. Cada abella obrera és l'agent representant d'una planta. Pot tenir dos rols diferents:

- ✗ Cercadora/recol·lectora de nèctar, amb la missió de "volar" per tota la base de dades (rusc) buscant informació que podria ser important (nèctar) en relació a la planta que controla.
- ✗ Abellot (fecundador): té la missió de "confirmar" el seu nèctar.

Les tasques que realitza, són doncs les següents:

- ✗ Quan es desperta, recollir la informació corresponent a la seva planta i que està emmagatzemada en una bresca del rusc (veure figura 5).
- ✗ Consultar el seu rol
- ✗ Buscar nèctar o convertir nèctar en mel, en funció del rol actual.

3.2 Apicultors

El sistema dona cabuda a usuaris aficionats, curiosos, experts i professionals en la

matèria, però sempre s'intentarà de captar l'opinió "no professional" o popular ja que és aquesta la que no sol estar recollida enlloc. Ara bé, el procés d'obtenció del coneixement (mel) a partir de les aportacions dels usuaris, estarà condicionat precisament per la categoria d'aquest. Per aquesta raó s'han definit tres tipus d'usuari en el sistema a part de la de l'administrador:

- Usuari anònim: accedeix a la web esporàdicament i dona o consulta algun tipus d'informació
- Usuari registrat: usuari que té un cert interès en el sistema i que d'alguna manera o altra s'ha confirmat la seva vàlua.
- Usuari expert: usuari coneixedor de la matèria. Es el responsable final de que no es fabriqui mel de mala qualitat (informació falsa), exercint d'ajuda als abellots.
- Administrador (super-usuari): responsable del manteniment del sistema.

Cada usuari té una fiabilitat base, fruit de la seva categoria. Aquesta fiabilitat va evolucionant en el temps degut a la seva interacció amb el sistema i al resultat de les seves aportacions com a coneixement validat. La fiabilitat de l'usuari s'assigna en un valor dins l'interval [0,1].

3.3 L'agent "relacions públiques"

Aquest agent és l'encarregat de controlar el nivell de veracitat dels usuaris. Fa les següents tasques

- ⌘ Degradar un usuari quan la informació que ha donat o bé no està en conformitat amb els criteris dels experts, o bé entra en contradicció amb la majoria d'aportacions d'altres usuaris.
- ⌘ Ascendir un usuari d'acord amb el seu número d'aportacions i la confirmació d'elles.
- ⌘ Eliminar algun usuari que s'ha degradat al mínim nivell. S'elimina un usuari del sistema quan facilita informació falsa sistemàticament i se li prohibeix l'entrada al verificar les dades de la comunicació.

3.4 Rusc

El rusc l'assimilem a la base de dades del sistema on s'emmagatzema la informació. Cada concepte (planta), pot tenir un nombre diferent d'atributs, segons com l'hagi caracteritzat l'usuari que l'ha definit, o les aportacions que hi han afegit altres usuaris a posteriori. En un principi, per un domini concret, com és el de la etnobotànica, es defineixen a priori un conjunt d'atributs que són els que inicialment s'han assignat en aquest domini. La estructura interna de la base de dades és prou flexible per poder reflectir la variabilitat dels components d'una planta en funció de la informació disponible. Així mateix, és possible afegir i eliminar atributs en el domini en funció de les peticions dels apicultors. Per qüestions de simplicitat, remetem al lector a (Masanas, 2003) per veure el detall de l'estructura. Per entendre el funcionament del sistema en tenim prou veient el contingut de la base de dades en bresques o conjunt de registres relatius a una planta. Podem entendre la bresca com si fos una fitxa de la planta (veure figura 5). Cada bresca conté cel·les que recullen les aportacions sobre atributs, juntament amb la confiança de les dades contingudes.

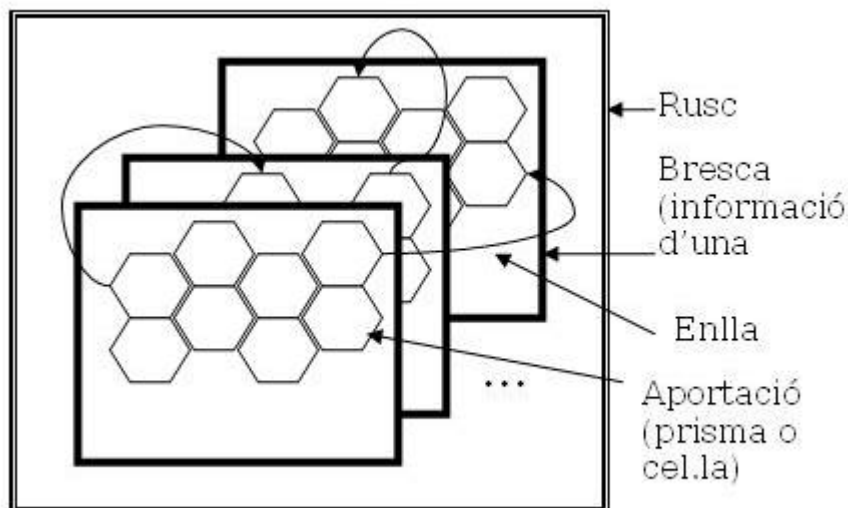


Figura 5. Representació visual de la base de dades del sistema.

Una aportació està caracteritzada per un atribut d'un tema i un valor definit en l'interval $[-1,1]$ que reflexa la fiabilitat de l'usuari transformada en confiança en l'atribut i que correspon a la confiança que l'usuari ofereix sobre la vinculació de l'atribut al tema. El signe negatiu indica que l'usuari nega una informació (atribut), conseqüentment la confiança de la informació serà el negat de la fiabilitat de l'usuari. Si el signe és positiu, indica que l'usuari confirma l'atribut com a característica del tema, i l'atribut prendrà el mateix valor de la fiabilitat de l'usuari. Per exemple, en etnobotànica podem tenir l'atribut "aromàtic" per la planta "llorer" amb les aportacions següents:

Usuari anònim. Fiabilitat 0.1. Opina Si. Confiança aportació = 0.1
 Usuari registrat. Fiabilitat 0.3. Opina No. Confiança aportació = -0.3
 Usuari expert: Fiabilitat 0.8. Opina Si. Confiança aportació = 0.8

Això vol dir que l'usuari anònim pensa que el llorer és aromàtic, però com la seva fiabilitat és baixa, 0.1, la confiança en aquesta aportació és també baixa: 0.1. Anàlogament l'usuari registrat, pensa que el llorer no és aromàtic, i aquesta aportació rep una confiança de -0.3. Finalment, l'usuari expert creu que sí ho és, i aleshores li aporta una confiança de 0.8. Aquesta confiança que els usuaris ofereixen en un atribut és combina per determinar-ne el seu valor final, és a dir la validesa de la informació.

4. Inferència de coneixement

A partir de les aportacions realitzades pels usuaris, el sistema aplica un procés de contrastació a través del qual diem que infereix coneixement. No és pot parlar d'un procés centralitzat i seqüencial, ja que cada abella en el sistema desenvolupa la seva tasca concurrentment amb la resta d'individus. El coneixement, doncs, emergeix com a conseqüència de les activitats individuals realitzades pels agents en el sistema.

Inicialment, els apicultors introdueixen informació sobre un tema (plantes) des de l'entorn web. L'abella reina, a mesura que té constància de les aportacions, crea les obreres corresponents. Per exemple, si l'apicultor ha introduït informació sobre el "timó", aleshores crearà una obrera "timó". Paral·lelament, si un altre usuari ha introduït informació sobre "farigola", crearà una obrera amb aquest nom.

Cada obrera, decideix el seu rol (recol·lector o abellot) en funció del nombre d'aportacions que hi ha hagut. Una abella recol·lectora, accedeix a la base de dades per obtenir tota la informació corresponent a la seva bresca. La seva tasca

important serà convertir la informació recent (nèctar) en coneixement (mel). Per aconseguir-ho, cerca sobre les referències immediates a la llista de referències proporcionada per la reina i sobre les referències relacionades. Aleshores computa els valors de confiança a les aportacions.

Quan la recol·lectora retorna l'informe de la seva actuació, la reina desperta les abelles més relacionades amb ella, perquè revisin també la seva mel. Al mateix temps es desperta un abellot per esbrinar si hi ha informació que necessita ser verificada i per tant requereix el vist i plau de l'expert.

L'abellot retorna un informe de necessitat de verificació o bé d'un tema o bé de tot el sistema. Aleshores, la reina actualitza el registre de la informació amb la data de revisió. L'abellot també és el responsable de proporcionar informació a la reina de les plantes que tenen un alt grau de coincidència (semblants).

La reina també pot crear un agent relacions públiques quan un usuari ha fet prou intervencions per merèixer una revisió del grau de fiabilitat que té en el sistema. Si la fiabilitat varia, aleshores es crida a les recol·lectores de les plantes on ha intervingut l'usuari, per tal d'actualitzar la validesa de la informació continguda en el sistema.

A continuació es descriu amb més detall els mètodes de determinació de la semblança de bresques, càlcul de la confiança de la informació, valoració de la necessitat de verificació, i de la fiabilitat de l'usuari.

4.1 Semblança de bresques

Una de les funcions de més valor que té el sistema per a l'usuari és el de la comparació de bresques, és a dir, en el cas de l'Etnobotànica, saber en quin grau una planta és més semblant a una altra. Cal recordar, que tota la informació d'una planta a la base de dades o rusc es pot visualitzar en una bresca. Com s'ha dit abans, no tots els atributs per una bresca poden estar definits i contenir informació. Això depèn de les aportacions realitzades fins al moment pels usuaris. Per la qual cosa, no és gens senzill definir un mètode de semblança entre les bresques.

Els mètodes de semblança han estat estudiats de molts investigadors en Intel·ligència Artificial (Wilson & Martínez, 1997). Tanmateix, no hi ha un guia universal que orienti sobre quin mètode és més adequat. La selecció depèn molt estretament del domini d'aplicació.

Després de consultar a experts en etnobotànica, s'ha corroborat que resulta molt complicat definir un criteri únic de semblança entre plantes. Per algú, una planta A pot semblar-se a una planta B si fa una funció semblant; mentre que per algú altre la semblança pot venir determinada per l'hàbitat o la forma que té. El camp de l'etnobotànica, és doncs, un clar exemple per entendre la complicació implícita que pot comportar el tractament del coneixement popular. Després de diversos estudis, finalment hem determinat que dues plantes són semblants si tenen força atributs comparables respecte als atributs totals i que a més a més en aquests atributs hi tenen el mateix valor. Això ha estat formalitzat en la següent metodologia.

Donades dues plantes P1 i P2, on

#ccomu: número d'atributs en comú
 #ctotals: número d'atributs diferents entre els dos (unió d'atributs)
 #ciguals: número d'atributs del mateix valor

Aleshores, la semblança total entre P1 i P2 es determinarà com segueix:

$$semblança_total = 0.5 * mesura_semblança + 0.5 * caract_semblants$$

on:

$$mesura_semblança = \frac{\#ciguals}{\#ccomu}$$

$$caract_semblants = \frac{\#ccomu}{\#ctotals}$$

A la taula 1 es mostren quatre exemples d'aplicació de la mesura sobre dos hipotètiques plantes P1 i P2.

Taula 1. Exemples de semblança.

Planta1	Planta2						
Número d'atributs informats	Número d'atributs informats	#ccomu	#ctotals	#ciguals	Mesura _semblança	Caract _semblants	Semblança total
6	10	4	8	3	0.75	0.5	0.63
16	20	4	28	3	0.75	0.14	0.45
1	1	1	1	1	1	1	1
4	10	4	6	3	0.75	0.67	0.71

4.2 Confiança de la informació

La informació que roman com a coneixement és la composició de totes les aportacions donades pels usuaris i ponderades per la seva fiabilitat. Existeixen diferents mètodes per a realitzar la composició (Valls, 2003). Tanmateix, el procés que interessa és un procediment incremental, que vagi variant la confiança de la informació a mesura que arribin noves aportacions. Es tracta, doncs, de variar la informació a partir de les evidències o aportacions que es disposen. Un mètode adequat per propagar evidències és el mètode dels factors de confiança tipus MYCIN (Shortliffe, 1976) que és el que ha estat escollit per la seva simplicitat.

Donades un conjunt d'aportacions d'usuaris sobre un atribut d'un tema dins de l'interval [-1,1], per obtenir un valor final de la informació:

1. S'ordena ascendentment i per valor absolut les aportacions.
2. S'aplica les fórmules de combinacions de factors de certesa de MYCIN que es mostren a continuació. Donades dues aportacions **x** i **y**, aleshores:

i) Si les dues fonts d'informació confirmen l'esdeveniment ($x = 0, y = 0$):

Confiança resultant = $x+y-xy$

ii) Si les dues fonts d'informació neguen l'esdeveniment ($x = 0, y = 0$):

Confiança resultant = $x+y+xy$

iii) Si hi ha discrepàncies entre els informadors ($x = 0, y = 0$ o $x = 0, y = 0$):

Confiança resultant = $x+y$

Es comença aplicant la confiança més petita en valor absolut, i es va agregant cada vegada una nova aportació.

Per il·lustrar el mètode amb un exemple, suposem que tenim les aportacions que es mostren en la figura 6:

Tema: Llorer
 Atribut: Aromàtic
 Aportacions: SI: 0.1,0.8
 No: -0.3

Després d'aplicar el primer pas, l'ordenació resultant (amb valor absolut) és: 0.1, -0.3, 0.8. Aleshores, es realitza la primera combinació entre 0.1 i -0.3. Atès que les dues aportacions són discrepants cal aplicar la tercera regla (iii). El resultat obtingut és -0.2. Aleshores, cal combinar aquest resultat -0.2 amb l'evidència de l'usuari expert 0.8. Com també són discrepants, es torna a aplicar la regla (iii), obtenint el valor final de confiança sobre l'atribut aromàtic del llorer de 0.6. Com es pot comprovar, l'aportació proporcionada per l'expert és la que finalment té més influència. Actualment, s'està treballant per cercar altres mètodes (Tena, 2004).

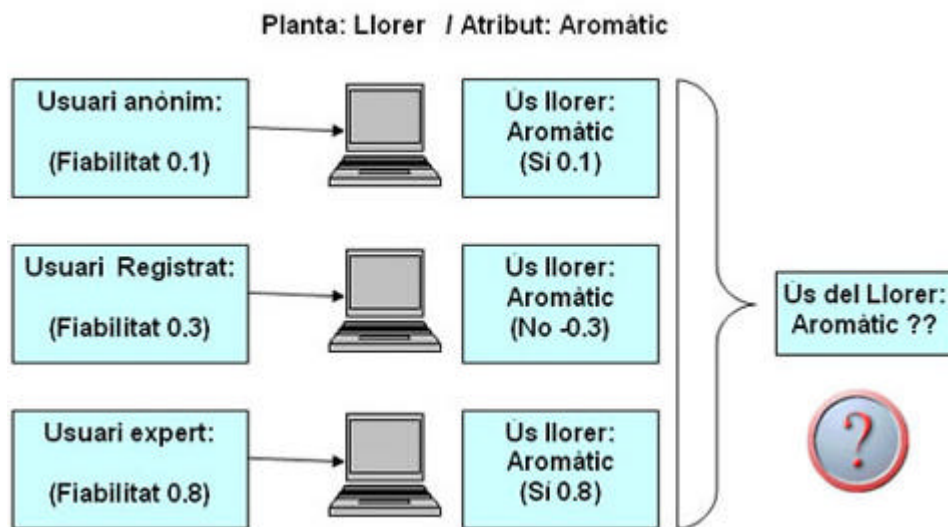


Figura 6. Problemàtica de la determinació de la confiança en un atribut.

4.3 Valoració de la necessitat de verificació

En l'apartat anterior, s'ha vist com un agent decideix sobre la confiança d'un atribut a partir de les aportacions dels usuaris. No obstant això, no és el mateix que un atribut tingui un valor de confiança de 0.4 a partir de 2 aportacions, o que tingui 0.4 i que compti amb més de 40 aportacions. És natural pensar que en el segon cas cal que un expert verifiqui la informació. Dues aportacions són molt poques, i segurament serà millor que s'espera la intervenció de més usuaris abans de fer intervenir un apicultor expert. En canvi amb més de 40 aportacions ja es pot dir que hi ha una clara diversitat d'opinions i pot ser molt necessari el vist-i-plau d'un expert. Aquesta condició de la informació és el que anomenem necessitat de verificació i que calcula l'agent abellot (veure figura 7).

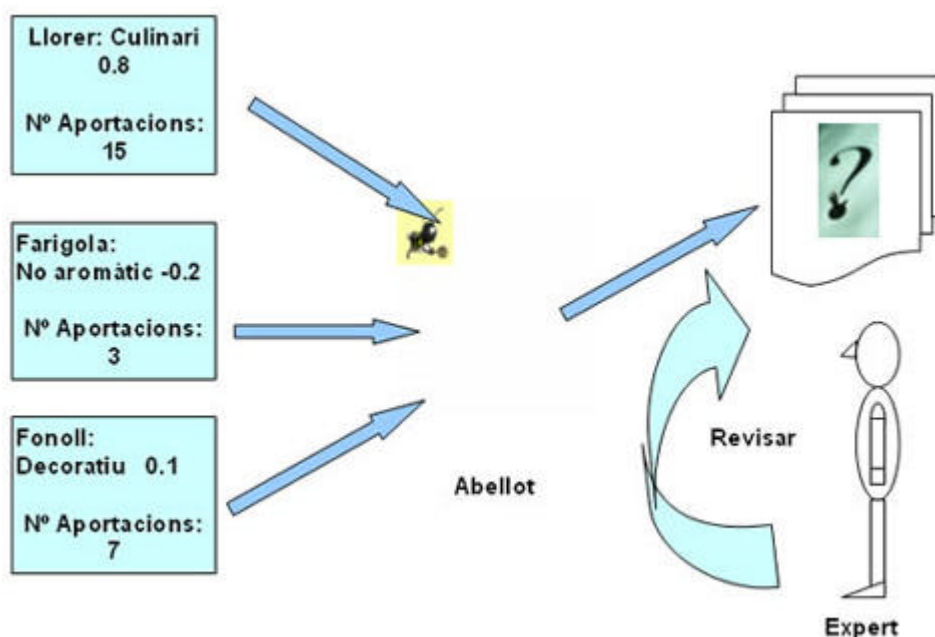


Figura 7. Necessitat de verificació.

La necessitat de verificació és un concepte imprecís de definir. Es pot parlar de una "alta" necessitat de verificació, o una "baixa" necessitat. Una de les tècniques que ens permet modelar aquest tipus de concepte són els mètodes de la lògica borrosa (Klir & Folger, 1992). A partir d'unes variables d'entrada, com pot ser la fiabilitat de l'usuari i el nombre d'aportacions, podem obtenir la variable de sortida grau_facunda, que expressa la necessitat de verificació.

Les variables s'han modelat amb les etiquetes següents:

- ⊗ Fiabilitat: mínima, baixa, normal i alta.
- ⊗ Número d'aportacions: molt poques, poques, mitjana, moltes
- ⊗ Grau fecunda: res, baixa, normal, alta.

S'ha establert una funció de pertinença borrosa per a cada etiqueta en funció dels valors numèrics de les variables als quals representa. A partir de les etiquetes s'han definit les regles de deducció del valor de grau fecunda segons es mostra a la figura 8.

<p>Si fiabilitat=m�nima i aportacions=molt poques llavors grau_fecunda = res</p> <p>Si fiabilitat=m�nima i aportacions= poques llavors grau_fecunda = normal</p> <p>Si fiabilitat=m�nima i aportacions=mitjana llavors grau_fecunda = alta</p> <p>Si fiabilitat=m�nima i aportacions=moltes llavors grau_fecunda = alta</p> <p>Si fiabilitat=baixa i aportacions=molt poques llavors grau_fecunda = res</p> <p>Si fiabilitat=baixa i aportacions= poques llavors grau_fecunda = baixa</p> <p>Si fiabilitat=baixa i aportacions=mitjana llavors grau_fecunda = normal</p> <p>Si fiabilitat=baixa i aportacions=moltes llavors grau_fecunda = alta</p> <p>Si fiabilitat=normal i aportacions=molt poques llavors grau_fecunda = res</p> <p>Si fiabilitat=normal i aportacions= poques llavors grau_fecunda = res</p> <p>Si fiabilitat=normal i aportacions=mitjana llavors grau_fecunda = baixa</p> <p>Si fiabilitat=normal i aportacions=moltes llavors grau_fecunda = normal</p> <p>Si fiabilitat=alta i aportacions=molt poques llavors grau_fecunda = res</p> <p>Si fiabilitat=alta i aportacions= poques llavors grau_fecunda = res</p> <p>Si fiabilitat=alta i aportacions=mitjana llavors grau_fecunda = res</p> <p>Si fiabilitat=alta i aportacions=moltes llavors grau_fecunda = baixa</p>

Figura 8. Regles fuzzy per deduir la necessitat de verificaci  (grau fecunda).

L'agent abellot aplica aquestes regles amb la finalitat de determinar si cal la intervenci  d'un expert i passa el resultat a la reina perquè actuï en conseq ncia.

4.4. Actualitzaci  de la fiabilitat d'un usuari.

El tractament de la confiança en la informaci   s un dels temes claus en el projecte i dep n dels usuaris que hi ha en el sistema, tots ells amb coneixements diferents i amb diferents graus de fiabilitat. Cal doncs, desenvolupar un mecanisme de control de la fiabilitat dels usuaris, que en el projecte EnCProu s'obte a partir de les aportacions correctes. El terme de correctesa d'una aportaci  pot semblar una incongru ncia quan s'est  dient que tota la informaci  que cont  la base de coneixement no t  mai una certesa absoluta. Per  si ens basem en un principi de "bona fe" de la majoria d'usuaris, es pot assumir que les aportacions correctes seran les que coincideixin amb el valor actual de la base de coneixement. Parlant en termes de n ctar i mel, una aportaci  (n ctar) ser  correcta si t  el mateix valor que la mel on est  contribuint.

L'actualitzaci  de la fiabilitat de l'usuari en funci  de les seves aportacions correctes la duu a terme l'agent relacions p bliques (veure figura 9). Per tal de veure la progressi  de cada un dels usuaris, l'abellot treballa amb les aportacions noves que

s'han fet des de l'última revisió. Primer es resten el número d'aportacions que hi havia en el moment de l'última revisió de les aportacions totals que hi ha ara i s'obté una variable amb el nombre d'aportacions noves (aportacions_noves) . Es fa el mateix amb les aportacions encertades, per obtenir la variable aportacions_encertades. Amb aquestes dades s'obté un coeficient d'encert, coef-encert, d'acord amb la següent expressió:

$$\text{coef_encert} = \frac{\text{aportacions_encertades}}{\text{aportacions_noves}}$$

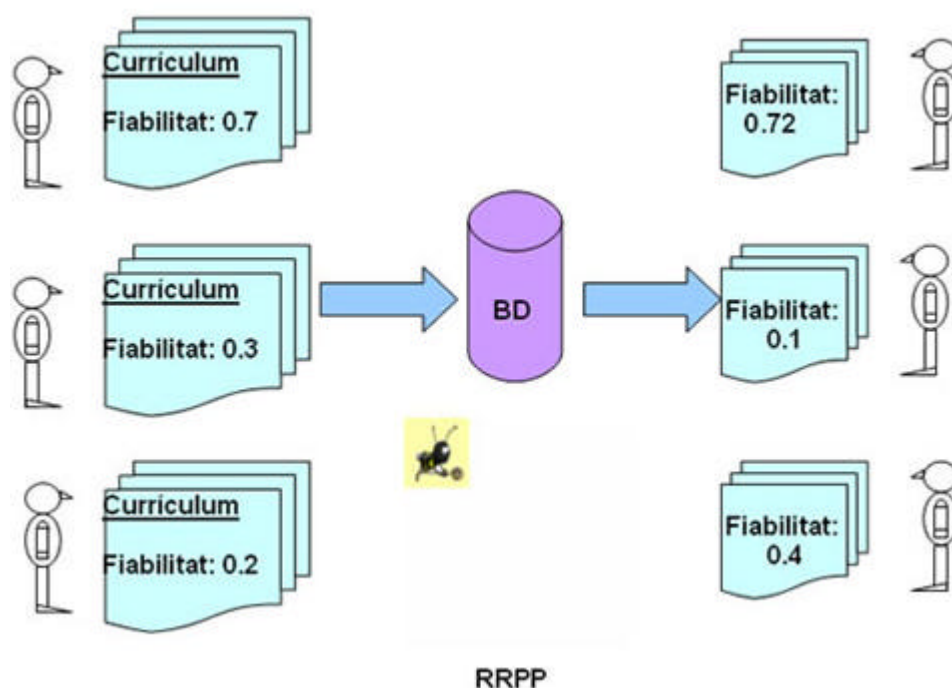


Figura 9. Actualització de la fiabilitat de l'usuari per part de l'agent relacions públiques.

Aquest coeficient seria 1 si s'hagués fet una sola aportació i aquesta fos encertada. Tanmateix, atès que es revisa la fiabilitat de l'usuari quan hi ha un cert nombre d'aportacions, això és bastant difícil que succeeixi.

El coeficient d'encert es compara amb el número d'aportacions normal que s'haurien d'encertar. De nou, aquest darrer concepte refereix a una quantitat imprecisa, i s'ha escollit la lògica difusa per modelar-lo. Les variables d'entrada són ara:

- ≠ Coeficient d'encert: mínim, baix, normal, alt
- ≠ Aportacions correctes: molt poques, poques, mitjana, moltes.
I la variable a calcular serà canvi_fiabilitat, que expressarà si cal baixar la fiabilitat de l'usuari, pujar-la o deixar-la com està:
- ≠ Canvi_fiabilitat: baixa, res, alta.

Les regles per deduir canvi_fiabilitat a partir de les entrades coeficient d'encert i aportacions correctes són les que es mostren a la figura 10.

```

Si coef_encert=mínim i aportacions_correctes=molt poques llavors
canvi_fiabilitat = res
Si coef_encert=mínim i aportacions_correctes= poques llavors canvi_fiabilitat
= res
Si coef_encert=mínim i aportacions_correctes=mitjana llavors canvi_fiabilitat
= res
Si coef_encert=mínim i aportacions_correctes=moltes llavors canvi_fiabilitat
= res

Si coef_encert=baix i aportacions_correctes=molt poques llavors
canvi_fiabilitat = res
Si coef_encert=baix i aportacions_correctes= poques llavors canvi_fiabilitat =
res
Si coef_encert=baix i aportacions_correctes=mitjana llavors canvi_fiabilitat =
res
Si coef_encert=baix i aportacions_correctes=moltes llavors canvi_fiabilitat =
baixa

Si coef_encert=normal i aportacions_correctes=molt poques llavors
canvi_fiabilitat = res
Si coef_encert=normal i aportacions_correctes= poques llavors canvi_fiabilitat
= baixa
Si coef_encert=normal i aportacions_correctes=mitjan llavors canvi_fiabilitat
= normal
Si coef_encert=normal i aportacions_correctes=moltes llavors canvi_fiabilitat
= alta

Si coef_encert=alt i aportacions_correctes=molt poques llavors canvi_fiabilitat
= baixa
Si coef_encert=alt i aportacions_correctes= poques llavors canvi_fiabilitat =
normal
Si coef_encert=alt i aportacions_correctes=mitjana llavors canvi_fiabilitat =
alta
Si coef_encert=alt i aportacions_correctes=moltes llavors canvi_fiabilitat =
alta

```

Figura 10. Regles per determinar el canvi de la fiabilitat d'un usuari

5.-Resultats experimentals

Els sistema s'ha implementat mitjançant la plataforma multi-agent JADE, sobre un servidor de pàgines web Apache, amb llenguatge d'scripts de pàgines web PHP, i el sistema gestor de base de dades MYSQL. Un exemple d'una pàgina web que configura lla interfície gràfica del sistema es mostra a la figura 11. A l'esquerra, l'usuari té un menú on pot escollir l'activitat a realitzar. Si escull "consultar plantes" es carregarà una pàgina web que li permetrà accedir a la planta sobre la qual estigui interessat. La informació que visualitzarà sobre la planta és tota la disponible en el sistema (bresca). Si escull "aportacions", podrà entrar informació relativa a una planta existent en el sistema. L'aportació consistirà en clicar sobre un atribut disponible en el sistema i opinar sobre ell. An'alogament es procedirà per afegir una nova planta.



Figura 11. Interfície gràfica del sistema.

El prototipus actual és accessible a la URL següent: <http://xixi.udg.es>, i funciona de forma experimental. Les proves de validació s'han fet amb l'ajuda de persones voluntàries, tant externes al projecte com vinculades a aquest.

Quant als mètodes de validació proposats, s'ha analitzat en primer lloc l'evolució dinàmica de la confiança de la informació (veure figura 12), i s'ha comprovat que mostra un comportament de mantenir les confiançaes altes quan s'arriba a un valor proper a 1.

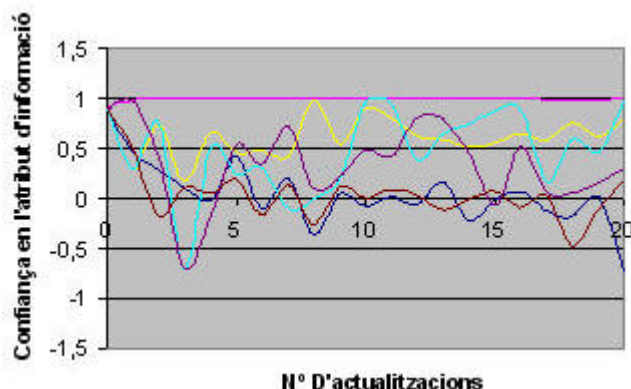


Figura 12. Evolució de la confiança d'informació a partir d'un usuari de fiabilitat alta (0.9).

En segon lloc, s'ha vist que el mètode proposat per a determinar la necessitat de verificació manté un comportament equilibrat quant al nombre de les aportacions i la fiabilitat dels usuaris (veure figura 13a). Un usuari amb una fiabilitat baixa i amb moltes aportacions, genera una necessitat de verificació significativa. Mentre que un usuari amb una fiabilitat elevada, amb moltes aportacions, la necessitat de verificació que genera és moderada. Finalment, el mètode de modificació de la fiabilitat dels usuaris també ha mostrat un comportament equilibrat en funció del coeficient encertat de l'usuari i el nombre d'aportacions encertades (veure figura

13b). Un usuari amb un coeficient d'incert elevat i moltes aportacions encertades, és susceptible de que la seva fiabilitat sigui revisada a l'alça. Mentre que un usuari amb un coeficient d'incert baix, és susceptible que la seva fiabilitat es vegi disminuïda. Aquest resultats corresponen als darrers mètodes que estem provant en el sistema (Tena, 2004).

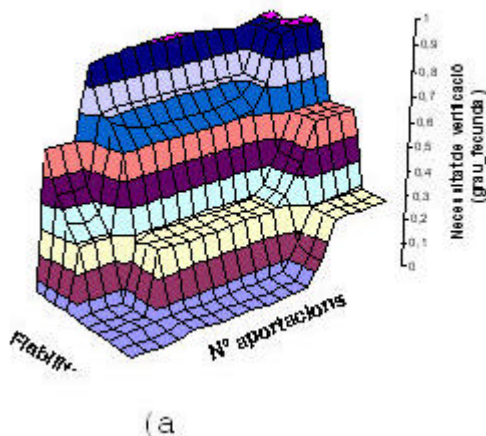


Figura 13. (a) Evolució de la necessitat d'informació.

(b) Evolució de la fiabilitat dels usuaris.

Tanmateix, en aquest prototipus del sistema s'han fet, de moment, les proves necessàries per comprovar les funcionalitat mínimes. Caldria garantir la seva consistència i seguretat amb proves addicionals en les que molt més usuaris interaccionessin amb el sistema.

6.- Conclusions

El coneixement popular, és un patrimoni que no ens podem permetre la vel·leïtat de deixar perdre. La importància de conservar el coneixement del que en són dipositaris el nostres "grans" mereix que dediquem el temps, medis i esforços necessaris per a aconseguir-ho. Des de la nostra aportació, em presentat una experiència que demostra com les tecnologies de la informació poden jugar un paper important en la recollida i preservació del coneixement popular. Usuaris sense gaires coneixements tecnològics, poden cercar o deixar informació sobre algun tipus de coneixement.

El sistema EnCProu està definit seguint els principis de la "Swarm Intelligence". Concretament s'ha implementat la metàfora de comportament d'una colònia d'abelles sobre una plataforma d'agents intel·ligents. Cada agent implementa el comportament d'un individu de la colònia: una abella reina que coordina totes les demés, una abella recol·lectora que assigna un valor confiança a la informació recollida, una abella abellot que cerca semblances entre temes a la vegada que determina la necessitat de verificació del coneixement acumulat, i una abella relacions públiques que actualitza la fiabilitat dels usuaris del sistema. El coneixement emergeix del sistema a partir de l'activitat de cada individu del rusc portat pel vol de les nostres agents abelles.

En aquest article s'han presentat els primers resultats a partir del prototipus, actualment disponible a Internet. Tot i que el sistema és aplicable a la recollida de qualsevol coneixement popular, alhora de provar la validesa de l'entorn ens hem centrat en la temàtica de plantes aromàtiques i medicinals.

Cal remarcar que existeix a Internet una aproximació col·laborativa de recollida d'informació anomenada wikipedia (<http://www.wikepeida.org/>). Es tracta d'una

enciclopèdia en línia escrita per centenars de voluntaris. L'eina informàtica que dona suport és un senzill programa d'edició anomenat Wiki, que segueix el concepte de codi lliure. Un dels elements més originals dels wiki és que les entrades ja publicades poden ser comentades i ampliades per altres usuaris. Tothom és lliure de proposar noves definicions de termes. Es tracta d'una edició totalment controlada per l'usuari. Tothom té dret a modificar les definicions sense cap limitació, només la d'estar registrat. La nostra proposta difereix de les wikipèdies ja que proposem mecanismes de contrastació (confiança, fiabilitat) i auto-organització (semblances, lligams) de la informació.

Com a treball futur, ens plantegem ampliar l'experimentació, instal·lant el sistema en equips més potents, fent difusió del sistema per a que usuaris puguin introduir informació. En conseqüència, caldrà avaluar l'evolució del sistema davant de grans volums de dades procedents de diverses fonts d'informació. Però el primer pas, un primer desenvolupament de l'entorn informàtic, ja està en marxa.

Bibliografia.

- ≠ Bonabeau, E., Théraulaz, G. Swarms Smarts. Scientific American, March 2000.
- ≠ Bonabeau, E., Dorigo, M., Thereaulaz, G. Swarm Intelligence. From Natural to Artificial Systems. Oxford University Press, 1999.
- ≠ Carpó, B. Costumari València 2. Edicions del Bullent, 1994.
- ≠ Jennings, N.R., Wooldridge, M. J. editors. Agent Technology. Foundations, Applications, and Markets. Springer-Verlag, Berlin, 1998.
- ≠ Klir, G. J., Folger, T. A. Fuzzy Sets, Uncertainty, and Information. Prentice Hall, 1992.
- ≠ Masanas, J. "Entorn interactiu de coneixement popular controlat per agents intel·ligents". Projecte Final de carrera dels estudis d'Enginyeria en Informàtica a la Universitat de Girona.Girona 2003
- ≠ Muntaner, E. Swarm Intelligence: Las hormigas nos ayudan a solucionar problemas complejos. IQ Magazine, January 2003.
- ≠ Shortliffe, E.H. MYCIN: Computer-based Medical Consultations. Artificial Intelligence Series, Tomo 2, Elsevier, NewYork, 1976.
- ≠ Tena Ligeró, M.LL. "Manteniment de la fiabilitat dels usuaris i de la confiança en la informació d'una base de coneixements populars autoorganitzada a través d'un sistema multiagent." Projecte Final de carrera dels estudis d'Enginyeria Tècnica en Informàtica de Sistemes a la Universitat de Girona. 2004
- ≠ Valls Mateu, A. . ClusDM: A Multiple Criteria Decision Making Method for Heterogeneous Data Sets. PhDThesis, Institut d'Investigació en Intel·ligència Artificial, CSIC, Bellaterra, Spain, 2003.
- ≠ Weiss, G. (editor). Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT Press, 1999.
- ≠ Wilson, D. R., Martínez, T. R. Improved Heterogeneous Distance Functions. Journal of Artificial Intelligence Research, 6(1997).
- ≠ Wooldridge, M. Intelligent Agents. In: Weiss, G. (editor). Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence. MIT Press, 1999. Chapter 1.