

Pràctica de Sistemes Basats en el Coneixement Intel·ligència Artificial

Bernat Farrero Badal
Jordi Romero de Villalonga

12 d'Abril de 2010

Índex

1	Introducció	2
1.1	Metodologia per l'anàlisi del problema	2
2	Identificació del problema	4
2.1	Identificació del problema	4
2.1.1	El SBC és viable	4
2.1.2	Fonts de coneixement	4
2.1.3	Evidències, hipòtesis i abstraccions heurístiques extretes .	5
2.1.4	Objectius del sistema	5
3	Conceptualització	6
4	Formalització del problema	7
5	Implementació	8
6	Ontologia	9
6.1	Fases de desenvolupament de l'ontologia	9
7	Descripció dels mòduls	10
8	Casos de prova	11
9	Conclusions	12
9.1	Secció	12

Capítol 1

Introducció

En aquesta pràctica es planteja el problema de crear un Sistema Basat en el Coneixement per recomanar llibres adients als usuaris segons les seves preferències i característiques. En l'actualitat hi ha una enorme oferta de llibres de tots els tipus i gèneres, precisament per això resulta difícil per un usuari desinformat destriar entre tots ells quin és el llibre perfecte. Evidentment, el procés de recomanació d'un llibre és quelcom que va més enllà d'elegir gènere, tema, popularitat del llibre, etc. Per recomanar adequadament un llibre cal conèixer bé l'usuari i emular el procés de recomanació d'un llibreter que ens coneix de tota la vida.

En el nostre cas el domini de cerca es restringeix a les novel·les de ficció, tot i que dins d'aquest gènere hi ha molts subgèneres i llibres de molt diferent naturalesa. El Sistema Basat en el Coneixement que construïm recollirà eficientment informació de l'usuari i restringirà progressivament el domini. El sistema ha de comptar amb una base de dades d'instàncies de llibres significativa per poder donar respostes adequades.

En primer lloc, el sistema començarà identificant l'usuari, li realitzarà preguntes comunes i després preguntes específiques pel seu perfil, traurà conclusions esborrant opcions no adequades, realitzarà l'associació heurística segons els criteris extrets, es refinarà el resultat (categoritzant entre opcions parcialment recomanables, recomanables o molt recomanables) i, finalment, es mostrarà a l'usuari les opcions recomanades ordenades per adequació. En cas de tractar-se d'opcions parcialment recomanables s'indicarà quins criteris no compleixen.

1.1 Metodologia per l'anàlisi del problema

El problema que afrontem és un problema d'anàlisi i la metodologia que hem seguit es basa en un cicle de vida en cascada que també encaixa alhora amb la metodologia basada en el prototipat ràpid i el disseny incremental. Consta de les següents fases:

- Identificació del problema: On es determina la viabilitat de la construcció del SBC i la disponibilitat de les fonts de coneixement.
- Conceptualització: Descripció semiformal del coneixement del domini del problema i descomposició

en subproblemes, segons la visió d'un expert. - Formalització: Cal definir el mecanisme adequat de representació del coneixement, en aquest cas segons la visió de l'enginyer de coneixement.

Seguidament a aquestes fases, ve la implementació i les proves del sistema. Com que es tracta d'un procés incremental, s'ha començat amb un prototipat senzill, que s'ha anat extenent i refinant a mesura que s'ha anat provant, amb els canvis que això ha suposat al model final d'ontologia, la descomposició de subproblemes i el desplegament de casos.

Capítol 2

Identificació del problema

2.1 Identificació del problema

La identificació del problema és la primera fase de l'anàlisi del problema. En primer lloc, consisteix en analitzar la viabilitat del sistema de coneixement (SBC) proposat. Cal justificar el sistema, demostrar que és adequat i que és capaç d'arribar a solucions satisfactòries. És també necessari especificar quines seran les fonts de coneixement que tindrem i quin d'aquest coneixement serà necessari pel sistema. Determinarem quins són els objectius del sistema, i quins tipus de sortides serà capaç d'oferir. També detallarem les hipòtesis prèvies que assumim.

2.1.1 El SBC és viable

Si es vol que el sistema sigui pràctic de cara a l'usuari, les bases de dades de llibres amb les que treballa han de ser molt grans, parlariem de l'ordre de milions de llibres. Evidentment, si el sistema fos algorímic, basat en qualsevol algorisme de cerca exhaustiva sobre tots aquests llibres, els resultats serien dolents i es trigaria molt en arribar a una solució. Com que el problema es defineix com un conjunt de restriccions i preferències, deduïm fàcilment que es pot resoldre amb un sistema basat en regles, com ara el que ofereix CLIPS.

En quant a la viabilitat econòmica del projecte, considerem que el mercat de les ventes de llibres per Internet es troba en constant creixement, on cada dia es presenta una nova plataforma de venda de llibres en línia. La majoria d'aquestes plataformes ofereixen avançats sistemes de cerca i molta informació a l'usuari, però no ofereixen sistemes intel·ligents de recomanació. Per tant, seria una aportació útil i innovadora pel mercat.

2.1.2 Fonts de coneixement

La informació disponible dels llibres l'hem extret de diverses llibreries en línia. En particular, hem usat Amazon.com, BarnesNoble i La Casa del Llibre. Les

instàncies particulars les hem extret (mitjançant la programació d'un programa *scraper*) dels llocs web esmentats.

Evidentment, podem comptar només amb aquella informació que és pública: la descripció dels llibres i tots els seus camps característics i de l'autor, tot i que no podem conèixer a priori detalls personals de l'autor ni quines influències ha tingut per comparar-les amb l'usuari. Tampoc tenim la possibilitat que el sistema aprengui noves regles a força d'ús. Per exemple, seria interessant que el sistema pogués rebre una valoració positiva quan una recomanació és adequada o negativa quan no ho és, i corregir o reescriure automàticament algunes de les seves regles (sistema conegut com aprenentatge basat en explicacions).

Les hipòtesis i abstraccions heurístiques (definicionals, qualitatives i de generalització) sobre el problema les hem obtingut contactant amb un expert, que en aquest cas és un lector habitual. En particular, hem entrevistat a Jordi Romero (entrevista a l'annex) que publica un blog de lectura. També hem obtingut informació de la valoració que ens han fet companys i amics i de consells que hem trobat en llibreries en línia.

El coneixement necessari pel sistema es divideix en aquell coneixement.

2.1.3 Evidències, hipòtesis i abstraccions heurístiques extretes

En procés de discussió a /heuristics.md

Evidències:

Abstraccions heurístiques:

2.1.4 Objectius del sistema

L'objectiu del sistema és generar una llista de **tres** llibres recomanats ordenat per grau d'adequació. Si algun d'aquests llibres és parcialment recomanat (perquè no compleix algun criteri) s'informarà l'usuari del criteri que no compleix. Però en principi i amb un domini suficientment gran, haurien d'existir almenys tres llibres recomanables per qualsevol combinació de preferències i restriccions de l'usuari.

Capítol 3

Conceptualització

En aquesta fase s'intenta adquirir la perspectiva de l'expert. De la interacció amb l'expert o experts en la recomanació de llibres n'hem d'extreure la dimensió final del domini i l'ontologia. A la vegada, hem de descomposar el problema en subproblemes i realitzar un anàlisi per refinaments successius amb l'objectiu d'obtenir la relació jeràrquica entre les diferents fases de resolució fins als operadors de raonament més elementals. Cal, doncs, identificar quin és el flux de raonament necessari en la resolució del problema des del punt de vista de l'expert.

3.0.5 Elements del domini

El domini del nostre SBC es basa en:

- Els llibres: El seu títol, el preu, l'any d'edició, el número d'edicions, el format, l'idioma, el número de pàgines, la mesura de la lletra, si és un best-seller, l'enquadernació, l'edat a la qual està orientat i el *ranking* de ventes.

Capítol 4

Formalització del problema

Capítol 5

Implementació

Capítol 6

Ontologia

6.1 Fases de desenvolupament de l'ontologia

(pàg 93 llibre)

Capítol 7

Descripció dels mòduls

Capítol 8

Casos de prova

Capítol 9

Conclusions

9.1 Secció