

Ontologies

els reptes de les organitzacions

V. Richard BENJAMINS

Traducció: David BARRI

RESUM: *La web semàntica ha estimulat la creació de tecnologies que permeten afegir significat al contingut. Una de les àrees principals d'aplicació de les tecnologies de la web semàntica, avui dia (2005/2006), és la gestió del coneixement i, en particular, el tractament d'informació desestructurada com si fos estructurada. Les ontologies són una de les claus de qualsevol aplicació de la web semàntica: poden considerar-se els «cervells» de l'aplicació. S'expliquen les metes que les organitzacions afronten en construir una aplicació basada en una ontologia.*

MOTS CLAU: web semàntica, ontologies, costos de la informació, iSOCO.

1 INTRODUCCIÓ

«La web semàntica és una extensió de la web corrent en la qual la informació es dona amb un significat ben definit, i que permet millorar el treball conjunt entre humans i ordinadors.»

Tim Berners-Lee, James Hendler, Ora Lassila, «The semantic web», *Scientific American* (May 2001).

En aquests darrers anys, molts dels projectes en recerca i desenvolupament s'han endegat amb l'objectiu d'elaborar la infraestructura i la tecnologia necessàries per fer realitat la web semàntica. Entre els temes estudiats hi ha la representació del coneixement a la web (RDF, OWL), les ontologies, el processament del llenguatge natural i els mitjans multimèdia. El realment important per a la web semàntica és la combinació d'aquestes tecnologies. En aquest article ens centrarem en les ontologies,¹ i, particularment, en el

1. <http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html>.

Dr. V. Richard Benjamins. Director of Innovation and R&D. <rbenjamins@isoco.com>
iSOCO. Pedro de Valdivia, 10. 28006 Madrid. <www.isoco.com>
 Article rebut el 3 de març de 2006.

seu paper en l'elaboració d'aplicacions intel·ligents o en la gestió del coneixement.

Les aplicacions relacionades amb l'elaboració de tecnologies de web semàntica són un projecte de programari i, com a tal, tenen els inconvenients —i avantatges— de qualsevol projecte de programari. Qualsevol organització interessada, com a client, en un projecte de programari demana uns pressupostos i un control sobre factors tals com planificació i despeses. La pràctica demostra que això és difícil. En projectes de web semàntica, hi ha un factor addicional relacionat amb la planificació i amb el control d'una ontologia.

A la següent secció, descriurem breument els avantatges i inconvenients de les ontologies per elaborar aplicacions més intel·ligents. Llavors, presentarem alguns dels reptes a què s'han d'enfrontar les ontologies en el procés d'adopció per part del món empresarial. Finalment, per raons il·lustratives, descriurem breument una aplicació basada en una ontologia.

2 ELS PUNTS FORTS DE LES ONTOLOGIES

Una ontologia és un especificació formal i explícita d'una conceptualització compartida.² Una ontologia dona significat al contingut. Permet anar més enllà de simples seqüències de caràcters i paraules clau. Per exemple, en lloc de considerar *Picasso* com una sèrie de caràcters —*P, i, c, a, s, s, o*—, sap que Picasso és un exemple de pintor, que és, alhora, un tipus d'artista, i que és, també, una persona. Una persona té un nom i un lloc de naixement. Un artista ha creat obres d'art i pot haver estat inspirat per altres artistes, etc. Tot això podria ser el començament d'una ontologia cultural.³ Per tant, més que ser només capaç de recuperar documents que continguin la seqüència *Picasso*, usant ontologies, les aplicacions saben què «vol dir» Picasso i poden inferir noves coses i formular preguntes per completar llacunes, etc. Les ontologies, per tant, proporcionen un avantatge únic en termes de competitivitat sobre la tecnologia de recuperació de la informació tradicional.⁴ Vegeu el lloc web del projecte SEKT <<http://www.sekt-project.com>> per a una vi-

2. R. Studer; V.R. Benjamins; D. Fensel, «Knowledge engineering: principles and methods», *Data and knowledge engineering*, 25 (1-2)(March 1998), p. 161-197.

3. V.R. Benjamins; J. Contreras; M. Blázquez; *et al.*, «Cultural heritage and the semantic web», en: European Semantic Web Symposium (1r: 2004: Càndia, Grècia), *The semantic web: research and applications* (Berlin: Springer, 2004), p. 433-444.

4. De fet, estem trivialitzant la feina d'associar contingut amb metadades ontològiques. Hi ha tota una àrea de recerca en PLN i web semàntica dedicada a aquesta qüestió.

sió general de les tecnologies de web semàntica per a la gestió del coneixement, i també <<http://seminars.ijis.si/sekt/training>> per als vídeos d'acompanyament; o també el número especial sobre gestió semàntica del coneixement del *Journal of knowledge management*.⁵

Quin és, però, el preu que cal pagar per aquest avantatge? La resposta és senzilla: cal elaborar i mantenir l'ontologia; i això, en l'actualitat, encara és més una art que una disciplina. Avui en dia, les ontologies són elaborades pels anomenats enginyers d'ontologies. El cost d'elaboració depèn de l'habilitat i experiència de l'enginyer, de la complexitat de l'ontologia i de la complexitat del procés d'elaboració (localitzacions diversos, públics amb interessos diversos...). La disponibilitat d'enginyers d'ontologies també és un aspecte que cal tenir en compte, perquè hi ha molts més enginyers informàtics que enginyers d'ontologies o de coneixement. S'ha avançat molt poc deixant fer aquesta feina a especialistes d'altres camps. Hi ha, però, una excepció interessant en el projecte Halo,⁶ que té com a objectiu el desenvolupament d'eines intel·ligents per tal que biòlegs, químics i físics siguin capaços de desenvolupar ontologies.

En els darrers anys, hem tingut l'oportunitat d'elaborar diverses ontologies, cosa que ens ha donat una experiència important que creiem que val la pena de compartir amb un públic més ampli. A continuació, llistem els àmbits de l'ontologia, amb el tipus d'aplicació per a la qual s'ha elaborat cadascun:

- àmbit jurídic
 - preguntes freqüents (en anglès, FAQ) per a jutges⁷
 - seguiment de casos legals a través de xarxes internacionals d'empreses legals⁸
- patrimoni cultural
 - accés intel·ligent⁹
- projectes, gent, habilitats, tecnologies
 - cercador de gent (no revelada)

5. *Journal of knowledge management*, vol. 9, issue 5 (2005). Número especial sobre gestió semàntica del coneixement.

6. <<http://www.projecthalo.com>>.

7. P. Casanovas; M. Poblet; N. Casellas; *et al.*, «Supporting newly-appointed judges: a legal knowledge management case study», *Journal of knowledge management*, vol. 9, issue 5 (2005), p. 7- 27.

8. J. Contreras; M. Poblet, «Netcase: an intelligent system to assist legal services providers in transnational legal networks», en V.R. Benjamins; *et al.* (ed.), *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications* (Berlin; Heidelberg: Springer, 2005). p. 218-232.

9. V.R. Benjamins; J. Contreras; M. Blázquez; *et al.*, «Cultural heritage and the semantic web».

- administració pública
 - cerca semàntica de serveis en línia (en curs de desenvolupament)
- relacions internacionals
 - motor de cerca semàntica¹⁰
- notícies
 - portal semàntic¹¹
- entitats financeres
 - adquisició i publicació de dades¹²
- àmbit financer
 - alerta de descobert¹³
 - agent de borsa (no revelada)
 - gestió de la innovació (no revelada)
- biologia, física, química
 - resolució de problemes (no revelada)
- graelles de càlcul
 - dades, serveis i procedència (en curs de desenvolupament)

En la següent secció, descriurem les nostres experiències, especialment pel que fa a l'apreciació que tenen les organitzacions clients sobre els projectes en els quals les ontologies tenen un pes important.

3 ELS REPTES QUE LES ONTOLOGIES HAN D'ENCARAR

En l'actual estat de coses, es té la idea que les ontologies:

- són difícils d'elaborar i mantenir,
- impliquen, potencialment, molts interessos de públics diferents, i
- presenten dificultats a l'hora d'estimar-ne i controlar-ne les despeses.

10. L. Rodrigo; V.R. Benjamins; J. Contreras; *et al.*, «Portal semàntic para relaciones internacionales», *Novática*, 178 (2005), p. 26–30. Número especial sobre web semàntica.

11. P. Castells; F. Perdrix; E. Pulido; *et al.*, «Semantic web technologies for a digital newspaper archive», en: European Semantic Web Symposium (1r: 2004: Càndia, Grècia), *The semantic web: research and applications* (Berlin: Springer, 2004).

12. J. Barrasa; O. Corcho; A. Gómez-Pérez, «Fund finder: a case study of database-to-ontology mapping», International Semantic Web Conference, *The semantic web* (2003), p. 6. Presentat al taller sobre integració semàntica (Sanibel Island, Florida).

13. <<http://dip.semanticweb.org/documents/D10.3.pdf>>.

14. <<http://www.aifb.uni-karlsruhe.de/WBS/cte/ontologyengineering/ontocom/ontocom.htm>>.

En altres paraules, les ontologies es consideren un mal menor necessari per poder gaudir dels avantatges que proporciona la web semàntica. Les organitzacions es pregunten coses com: «Disposo de personal capacitat per gestionar ontologies?», «Com podré mantenir l'ontologia?», «Com s'integrarà amb la meua pròpia infraestructura informàtica?», «Hauré de duplicar la informació que hi ha en les bases de dades existents en l'ontologia?», «Quin esforç i quin pressupost em representarà tot plegat?». Totes aquestes qüestions —i moltes d'altres— són perfectament legítimes. Al cap i a la fi, les organitzacions que s'embarquen en projectes d'aquesta mena estan fent una inversió en una tecnologia nova per elaborar sistemes que se suposa que han de funcionar molts anys. És a dir, que no els podem dir: «Deixeu-nos fer el projecte i oblideu-vos-en.»

Segons la nostra experiència, hi ha tres reptes principals que cal encarar abans que la web semàntica surti al mercat. Les organitzacions clients haurien de percebre que les ontologies:

- són factibles, controlables i gestionables, així com també assequibles,
- són un actiu que permet ser més competitius, i
- són un actiu que es pot vendre com a contingut de valor afegit.

El primer punt —«les ontologies són factibles, controlables i gestionables, així com també assequibles»— no és cap sorpresa, ja que això és aplicable a qualsevol projecte (de programari). Entre els factors importants per aconseguir-ho hi ha les metodologies per gestionar el cicle de vida de l'ontologia, les eines de gestió del projecte, i el reaprofitament de les ontologies ja existents. Quan no hi ha ontologies rellevants ja existents, el cost en personal i en esforços és important. Com hem dit, l'elaboració d'una ontologia és més que una art que una tècnica, i en són especialment difícils els primers passos. Hi ha, però, moltes eines que poden ser d'ajut per superar la síndrome de la pàgina en blanc (per on començo?), com ara un mapa mental, l'anàlisi de la freqüència d'aparició de paraules en les fonts, l'entrevista a experts per aprendre les *preguntes competencials* —aquelles que una ontologia en procés d'elaboració ha de ser capaç de respondre correctament—, tempesta d'idees (*brainstorming*), etc. Aquestes eines ajuden a elaborar una llista inicial de termes que cal encarar, així com un esquelet d'estructura taxonòmica. Posteriorment, això pot ser perfeccionat usant editors d'ontologies, revisors de consistència i, per suposat, amb la comunicació amb experts en el camp de què es tracti.

Pel que fa al segon punt, per a moltes organitzacions, el coneixement i la informació del seu negoci és un actiu que els ajuda a competir en l'economia global. És cert, també, que moltes organitzacions encara estan lluitant per aconseguir la manera més adequada de capturar, emmagatzemar, mantenir i reutilitzar aquest coneixement. Cada cop més aplicacions empresarials —com ara ERP i CRM— s'estan adoptant per tal que s'encarreguin d'aspectes concrets del coneixement del negoci. La proposta realment valuosa de les ontologies és que se centren en el coneixement rellevant —en comptes de l'aplicació—, el qual pot ser reutilitzat posteriorment per moltes aplicacions diferents. Les eines de gestió d'ontologies proporcionen una infraestructura que permet que l'ontologia evolucioni al mateix ritme que l'empresa. Les organitzacions que tenen una ontologia en el lloc adequat poden elaborar més ràpidament noves aplicacions, perquè part de l'aplicació (el coneixement del negoci) ja existeix.

En relació amb el tercer punt, quan una ontologia s'ha creat, no només constitueix la columna vertebral de les aplicacions de l'organització, sinó que algunes parts (per exemple, les que no continguin informació confidencial) es poden vendre a altres organitzacions. Penseu, per exemple, en una ontologia del camp de l'administració electrònica o la sanitat. Per quin motiu la ciutat de Barcelona no podria «vendre» parts de la seva ontologia sobre serveis en línia a la ciutat de Madrid? Per descomptat, encara cal desenvolupar models adequats a cada negoci abans que tot això es pugui dur a la pràctica.

4 INDICADORS DE COSTOS EN LA CONSTRUCCIÓ D'UNA ONTOLOGIA

Pel fet que l'elaboració d'una ontologia és més una art que una tècnica, el seu cost és difícil de determinar. En el context del projecte SEKT <www.sekt-project.com>, s'està treballant per entendre millor aquest procés (*Ontocom: ontology cost model*, model de costos d'ontologies).¹⁴ És important —especialment per a ontologies grans— tenir una visió clara del cost dels factors que n'afecten directament el procés d'elaboració, així com una estimació de com l'ontologia concreta els supera.

Ontocom diferencia entre indicadors de costos de producte, de procés i de personal. El primer dóna compte de la influència de les *propietats de l'ontologia* en el cost total. El segon estableix les dimensions del procés d'enginyeria que són rellevants per a l'estimació del cost, mentre que el tercer posa l'èm-

14. <<http://www.aifb.uni-karlsruhe.de/WBS/cte/ontologyengineering/ontocom/ontocom.htm>>.

fasi en l'experiència, habilitat i continuïtat de l'equip per a l'esforç que implica el procés.

El fet de tenir una visió clara dels indicadors de costos no resol el problema, en el sentit que, actualment, tots els costos estan controlats i no tenen perquè produir-se desviacions. No obstant això, seria clarificador per a les organitzacions saber quins factors estan controlats i poden ser manipulats, quins són interns a l'organització i no són com hom voldria, quins factors es troben més enllà de llur control i són una amenaça potencial per al projecte, i, finalment, quins factors són externs, però no plantegen un perill especial. Això permet que l'organització se centri en els riscos principals, per tal de poder-los gestionar adequadament, així com planificar les contingències que puguin esdevenir-se. La taula inferior mostra aquestes quatre situacions. Cada factor ha de ser completat amb la informació detallada en els següents apartats.

Gestió de riscos per a factors rellevants per a l'elaboració d'ontologies		
	Interns	Externs
Desitjats	D'acord. Cal assegurar-se que continuï així.	D'acord. Cal, però, controlar-lo i elaborar plans d'acció.
No desitjats	Actuar si el factor és rellevant.	Mirar d'influir per tal que terceres parts canviïn o evitar la importància del factor.

En els següents apartats descriurem breument els factors de cost rellevants per a l'elaboració i manteniment d'ontologies, per tal com han estat identificats pel treball d'Ontocom en el si del projecte SEKT.

4.1 Factors relatius al producte

Factors relatius a la pròpia ontologia.

a) Complexitat de l'ontologia

Complexitat del camp. Quina ha de ser l'amplada de la cobertura? Visió estreta, sentit comú i baixa connectivitat o visió ampla, coneixement expert i alta connectivitat. Quantes fonts s'han de tenir en compte i s'han d'incloure per tal d'elaborar l'ontologia?

Complexitat conceptual. Quin és el grau de dificultat d'identificació dels conceptes, els atributs i les relacions del camp concret.

Complexitat d'implementació. La dificultat d'elaborar un mapa de les construccions ontològiques d'acord amb el llenguatge d'implementació. En el cas que s'usi una eina determinada, l'esforç té a veure amb el grau de dificultat que implica el modelatge de l'ontologia fent ús de l'entorn de l'eina (p. ex., Protégé).

b) Dades

Com és de difícil donar exemples a l'ontologia? Provenen aquests de fonts altament estructurades o de dades no estructurades en forma lliure de llenguatge natural?

c) Reutilització

S'hauria de reutilitzar l'ontologia per a altres aplicacions diferents de l'aplicació per a la qual es va dissenyar? A major grau de reutilització, major preu.

d) Documentació

Quina documentació necessita l'ontologia? N'hi ha prou amb l'ontologia en si mateixa —pel que fa a conceptes, atributs, relacions, exemples— o cal, també, el procés complet del cicle de vida, incloent-hi l'elaboració i el manteniment?

e) Avaluació de l'ontologia

Quina quantitat de tests són necessaris o desitjables per provar l'ontologia? Com més esforços s'hi dediquin, major serà el preu.

f) Integració de l'ontologia

Aquest factor té a veure amb la integració de les ontologies en el cas que l'ontologia concreta s'hagi d'elaborar a partir d'altres ja existents.

4.2 Factors relatius al personal

Els factors de personal posen l'èmfasi en l'experiència de l'equip, les habilitats i la continuïtat del personal involucrat en el procés d'elaboració.

a) Capacitat i experiència dels enginyers d'ontologies

A major grau de capacitació i d'experiència dels enginyers d'ontologies, menor esforç en l'elaboració de l'ontologia.

b) Capacitat i experiència dels experts de camp

A major grau de capacitació i d'experiència dels enginyers d'ontologies, menor esforç en l'elaboració de l'ontologia. L'experiència es relaciona tant amb el camp d'interès, com en el procés d'elaboració d'una ontologia.

c) Continuïtat del personal

Quan els equips encarregats del projecte —mentre aquest existeix— són estables, l'esforç esmerçat en l'elaboració de l'ontologia acostuma a ser inferior —si els altres factors són els mateixos.

4.3 Factors relatius al procés

Aquests factors tenen a veure amb el procés d'enginyeria d'elaboració de l'ontologia.

a) Eines

Quan hi ha eines sofisticades disponibles per a l'equip i el projecte, l'esforç d'elaboració de l'ontologia serà menor. Les eines es poden utilitzar per a anàlisi de camp, conceptualització, implementació, comprensió i avaluació de l'ontologia, exemplificació, modificació, traducció, integració i documentació.

b) Ubicació

Si el procés d'elaboració implica gent amb ubicacions diferents, el procés serà més lent. La possibilitat de tenir trobades cara a cara és un factor de re-

ducció del cost. El telèfon i el correu electrònic són útils, però no són la millor opció.

Tenir en compte cadascun d'aquests factors i ser conscients de quins poden causar problemes, així com saber si estan sota control o no, ajuda significativament a tenir una idea fiable del cost del projecte.

5 APLICACIÓ: UN MOTOR DE CERCA SEMÀNTIC PER A LES RELACIONS INTERNACIONALS

El Real Instituto Elcano <<http://www.realinstitutoelcano.org>> és una institució que proporciona —a través de la seva pàgina web— informes de la situació política mundial, amb un interès especial en els esdeveniments rellevants per a Espanya. Aquests informes s'organitzen en diferents categories, com ara economia, defensa, societat, orient mitjà, etc. En col·laboració amb l'institut, iSOCO ha desenvolupat una ontologia de relacions internacionals que descriu coses com ara països, governs, fronteres, guerres, exèrcits, etc., així com llurs relacions. L'ontologia consisteix en 70 conceptes i 120 atributs. Més de 100.000 exemples de l'ontologia s'han incorporat automàticament agregant i consolidant sis llocs web rellevants, com ara el *CIA world factbook*,¹⁵ el ministeri espanyol d'affers exteriors i el lloc web israelià sobre terrorisme. Amb l'ús de l'ontologia —i dels seus exemples—, automàticament inseríem enllaços als informes de la institució, que permet navegar dels documents als exemples de l'ontologia. També proporcionàvem una interfície en llenguatge natural per interrogar l'ontologia. El resultat és un registre ontològic —o diversos— que constitueix la resposta a la interrogació. Per exemple, la pregunta «quins països limiten amb els Països Baixos?» dóna com a resultat els registres d'Alemanya i Bèlgica.¹⁶ En aquesta aplicació, l'ontologia conforma el cervell de l'aplicació: si la qüestió es troba dins del camp de relacions internacionals, la resposta sobrepasa significativament els tradicionals motors de cerca (formulació de preguntes enfront de recuperació documental). L'experiència que hem acumulat en l'elaboració d'ontologies és la següent:

1. La construcció de l'esquema ontològic ha estat relativament senzilla, ja que vam poder començar a partir del *CIA World Factbook* <<http://www.cia.gov/cia/publications/factbook/>>.

15. <<http://www.cia.gov/cia/publications/factbook/>>.

16. Els detalls de la implementació es poden trobar a Rodrigo, Benjamins, Contreras *et al.*

2. L'adquisició dels exemples ha estat barata, ja que s'ha generat automàticament, amb una mínima supervisió humana.
3. El manteniment dels exemples de l'ontologia requereix una reexecució periòdica del programari d'adquisició automàtica i d'hipervincles.
4. El manteniment de l'esquema de l'ontologia demana la intervenció d'experts de camp de l'institut, així com una reexecució periòdica del programari d'adquisició automàtica i d'hipervincles.

6 CONCLUSIONS

En aquest article, hem confirmat que les tecnologies de web semàntica —i, particularment, les ontologies— són unes candidates prometedores per conduir el web fins al màxim del seu potencial. Molts projectes de recerca que hi estan relacionats ja han proporcionat la tecnologia suficient per fer el pas al mercat i a la societat. Generalment, les organitzacions no estan interessades a adoptar noves tecnologies, a no ser que hi vegin un clar avantatge per al seu negoci. Però, un cop se n'adonen d'aquest avantatge, cal que també vegin que la nova tecnologia és comprensible i manejable. Les ontologies són un dels ingredients tecnològics de les tecnologies de web semàntica que cal explicar a les organitzacions. Això es pot fer, però no es pot donar per suposat. En aquest article, hem explicat breument la nostra experiència a l'iSOCO en l'elaboració de diferents ontologies per a organitzacions.

7 AGRAÏMENTS

Part del treball descrit en aquest article ha estat realitzat parcialment en el context del projecte SEKT (IST-FP6).

8 BIBLIOGRAFIA

- BARRASA, J.; CORCHO, O.; GÓMEZ-PÉREZ, A. «Fund finder: a case study of database-to-ontology mapping». INTERNATIONAL SEMANTIC WEB CONFERENCE. *The semantic web* (2003), p. 6. Presentat al taller sobre integració semàntica (Sanibel Island, Florida).
- BENJAMINS, V.R.; CONTRERAS, J.; BLÁZQUEZ, M.; *et al.* «Cultural heritage and the semantic web». En: EUROPEAN SEMANTIC WEB SYMPOSIUM (1r: 2004: Càndia, Grècia). *The semantic web: research and applications*, Berlin: Springer, 2004), p. 433-444.

CASANOVAS, P.; POBLET, M.; CASELLAS, N.; *et al.* «Supporting newly-appointed judges: a legal knowledge management case study». *Journal of knowledge management*. Vol. 9, issue 5 (2005), p. 7-27.

CASTELLS, P.; PERDRIX, F.; PULIDO, E.; *et al.* «Semantic web technologies for a digital newspaper archive». En: EUROPEAN SEMANTIC WEB SYMPOSIUM (1r: 2004: Cànida, Grècia), *The semantic web: research and applications*. Berlin: Springer, 2004).

CONTRERAS, J.; POBLET, M. «Netcase: an intelligent system to assist legal services providers in transnational legal networks». En: BENJAMINS, V.R.; *et al.* (ed.). *Law and the semantic web: legal ontologies, methodologies, legal information retrieval, and applications*. Berlin; Heidelberg: Springer, 2005. p. 218-232.

EUROPEAN SEMANTIC WEB SYMPOSIUM (1r: 2004: Cànida, Grècia). *The semantic web: research and applications*. J. Davies, D. Fensel, C. Bussler, R. Studer, eds. Berlin: Springer, 2004 (Lecture notes in computer science; v. 3.053).

Journal of knowledge management. Vol. 9, issue 5 (2005). Número especial sobre gestió semàntica del coneixement.

RODRIGO, L.; BENJAMINS, V.R.; CONTRERAS, J.; *et al.* «Portal semántico para relaciones internacionales». *Novática*, 178 (2005), p. 26-30. Número especial sobre web semàntica.

STUDER, R.; BENJAMINS, V.R.; FENSEL, D. «Knowledge engineering: principles and methods». *Data and knowledge engineering*, 25 (1-2)(March 1998), p. 161-197.