

Cap a l'ús efectiu d'ontologies en aplicacions informàtiques

Jordi Conesa i Caralt (jconesac@uoc.edu)

Estudis d'Informàtica, Multimèdia i Telecomunicació, Universitat Oberta de Catalunya

Joan Casas Roma (jcasasrom@uoc.edu)

Internet Interdisciplinary Institute, Universitat Oberta de Catalunya

Working paper

Working Paper Series WP11-002

Grup de Recerca en Enginyeria del Programari (GRES)
Coordinador del grup de recerca: Robert Clariso (UOC)

Data de recepció: Gener 2012
Data d'acceptació: Març 2012
Data de publicació: Març 2012



Internet Interdisciplinary Institute (IN3)

<http://www.uoc.edu/in3>
Parc Mediterrani de la Tecnologia
Av. Canal Olímpic, s/n.
08860 Castelldefels
Barcelona (Espanya)
Tel. 93 673 50 00

Universitat Oberta de Catalunya (UOC)

<http://www.uoc.edu/>
Av. Tibidabo, 39-43
08035 Barcelona
Espanya
Tel. 93 253 23 00



The texts published in this journal are – unless indicated otherwise – covered by the Creative Commons Spain Attribution 3.0 licence. You may copy, distribute, transmit and adapt the work, provided you attribute it (authorship, journal name, publisher) in the manner specified by the author(s) or licensor(s).

The full text of the licence can be consulted here:
<http://creativecommons.org/licenses/by/3.0/es/deed.en>.

Taula de Continguts

1. Introducció.....	6
2. Creació d'ontologies a partir d'etiquetes de llocs d'etiquetatge social	7
2.1. Feina Relacionada.....	8
2.2. Identificant Conceptes de Domini a partir d'Etiquetes	9
2.3. Implementació	13
2.4. Discussió	15
3. ITINER@: Ús d'ontologies en Sistemes d'Informació Geogràfica.....	16
3.1. Ontologies en l'àmbit del turisme	17
3.2. Ús d'Ontologies per la Creació de Rutes Turístiques: ITINER@.....	23
3.3. Discussió	31
4. Conclusions.....	32
Agraïments	33
Referències Bibliogràfiques	33

Cap a l'ús efectiu d'ontologies en aplicacions informàtiques

Jordi Conesa i Caralt (jconesac@uoc.edu)

IN3-UOC Resident Researcher

Joan Casas Romà (jcasasrom@uoc.edu)

Internet Interdisciplinary Institute, Universitat Oberta de Catalunya

Resum

Una ontologia proporciona una visió comú d'un domini i pot ser vista com un esquema conceptual consensuat que ha de ser entès tant per ordinadors, com per éssers humans. En l'actualitat, les ontologies s'utilitzen per resoldre problemes en els camps del disseny del programari, la representació del coneixement, la millora de la interoperabilitat, la millora de consultes web i la web semàntica, però majoritàriament en el món acadèmic. Podem dir que una ontologia s'utilitza efectivament en un programa quan el seu ús fa que el programa "sembli" més intel·ligent, millorant alguna de les seves facetes, tals com proporcionar més funcionalitats, millorar el seu temps de resposta o augmentar la compressió de dades.

La investigació realitzada estudia com utilitzar les ontologies de manera efectiva en aplicacions reals i, per tant, ajudar a promoure el seu ús en el món real. Amb aquest objectiu en ment s'han estudiat les metodologies i eines necessàries per identificar ontologies (o parts d'elles) rellevants per a una determinada tasca, l'estat actual de les ontologies, i l'aplicació efectiva de les ontologies en varis camps amb la finalitat de millorar les funcionalitats dels seus sistemes d'informació. En aquest article en particular es mostren dues experiències enfocades a entendre com usar efectivament ontologies per resoldre un problema concret de l'àmbit dels sistemes d'informació geogràfics i com crear ontologies que facilitin la seva reutilització efectiva. La primera de les experiències presenta un mètode per crear ontologies que reutilitza la informació existent a Internet. La segona utilitza un conjunt d'ontologies i especificacions de forma efectiva per crear un assistent turístic virtual mòbil amb funcionalitats avançades. A partir d'aquestes experiències es presenten algunes guies per millorar l'eficiència en la creació, la selecció i l'ús d'ontologies.

Paraules clau

Ontologies, semàntica, sistemes d'informació, *Ontology Driven Information Systems*, Sistemes d'Informació Geogràfica, sistemes basats en localització, web social, *knowledge of the crowd*

Citació recomanada:

CONESA, Jordi; CASAS, Joan (2012). "Cap a l'ús efectiu d'ontologies en aplicacions informàtiques" [online working paper]. (Working Paper Series; WP11-002). IN3 Working Paper Series. IN3 (UOC). [Accessed: dd/mm/yy].
<<http://in3wps.uoc.edu/ojs/index.php/in3-working-paper-series/article/view/n11-conesa-casas/n11-conesa-casas>>

1. Introducció

Internet i l'ús que se'n fa ha provocat un canvi global no tan sols en la manera de viure, sinó també en com concebre els negocis. Podem dir que estem davant d'un nou paradigma cultural, social i econòmic que ha obert noves possibilitats i perspectives en molts de sentits. No obstant, tot i la dràstica evolució d'Internet en els últims anys encara hi ha un punt que no acaba d'arrencar: la Web Semàntica. Les webs actuals contenen dades textuais que tan sols són fàcilment interpretables per a éssers humans. La web semàntica implica un canvi de paradigma on les aplicacions informàtiques també són capaces d'entendre la informació de les pàgines web i per tant de fer servir aquesta informació per oferir als internautes serveis que es comportin de manera intel·ligent. Imagineu un cercador web que en funció del nostre historial desambigui completament les nostres cerques i les estengui per obtenir millors resultats, o bé una aplicació que sigui capaç de preparar-nos les vacances a partir de les nostres preferències personals (tipus de viatge, ubicacions, interessos turístics, forma física...), el nostre pressupost i les dates disponibles. Doncs bé, tot això seria teòricament possible en el moment en que la web semàntica esdevingui una realitat. No obstant, això no passarà fins que les webs no continguin informació en un format que sigui entenedor pels programes informàtics. Un dels mecanismes per dotar de semàntica Internet, i per tant, que pot ajudar a solucionar aquest problema són les ontologies.

Una ontologia és una especificació d'un vocabulari de representació per un domini de discurs que permet descriure el món des d'un punt de vista concret. Actualment, hi ha un gran nombre de reptes associats amb el desenvolupament i ús d'ontologies. Cada ontologia és específica a un domini i és difícil crear-la perquè requereix d'un coneixement profund sobre el domini a ser descrit. Les ontologies més complexes i grans requereixen l'esforç col·laboratiu d'una comunitat d'experts en els dominis involucrats. Tot i que les ontologies han de ser compartides i reutilitzades, és difícil fer-ho quan diferents experts del domini desenvolupen diferents ontologies, i encara més quan el domini a representar canvia amb el temps. D'altra banda, algunes ontologies poden ser culturalment dependents i, per tant, no ser reusables entre aplicacions destinades a diferents cultures.

L'objectiu d'aquest treball és estudiar com es poden utilitzar les ontologies de forma efectiva des de totes les seves vessants, que són: la identificació d'ontologies potencialment rellevants per un domini concret, la creació de noves ontologies a partir d'informació existent (altres ontologies o informació del web), i l'ús efectiu d'ontologies (ja siguin noves o reutilitzades).

Aquest objectiu tan ambiciós s'ha abordat via la creació i aplicació efectiva d'ontologies en diferents àmbits d'aplicació. En aquest article es presenten dues d'aquestes experiències on s'han abordat els següents reptes:

1. Definir mecanismes per crear ontologies de forma que es pugui reutilitzar la informació que existeix al Web, ja vingui d'altres ontologies o es trobi integrada en diferents llocs webs.
2. Reutilitzar ontologies en diferents àmbits de forma efectiva. En particular, en aquest document es mostra com s'ha aplicat de forma efectiva una ontologia en l'àmbit dels sistemes d'informació geogràfics (SIG). Tal i com hem comentat, una ontologia s'utilitza de forma efectiva per una aplicació informàtica si el seu ús permet que l'aplicació es comporti de forma més intel·ligent.

Aquest document s'estructura de la següent forma: la segona secció presenta una aproximació que permet reutilitzar informació d'Internet, en particular de *Delicious*¹, per tal d'identificar conceptes potencialment rellevants per a una ontologia de domini. Després, la secció 3 presentarà una experiència on s'han utilitzat diferents ontologies en l'àmbit de sistemes d'informació geogràfica per tal d'oferir un assistent turístic virtual que excedeixi les funcionalitats dels actuals assistents. Finalment, la secció 4 discuteix la feina realitzada i presenta les conclusions del treball presentat.

2. Creació d'ontologies a partir d'etiquetes de llocs d'etiquetatge social

Internet és una font d'informació molt valuosa. La seva evolució ha fet què, cada vegada més, es presenti com una plataforma que permet que els ciutadans d'Internet, és a dir els internautes, es comuniquin, col·laborin i comparteixin continguts. Aquesta combinació d'emmagatzemament d'informació i col·laboració està canviant la nostra forma de treballar, així com la manera de dur a terme tasques específiques, com ara la recerca d'informació (Hendler and Golbeck 2008). Un element que ha contribuït significativament a aquest canvi són les etiquetes, també anomenades metadades febles. El seu resultat és una capa de dades de marcat social que conté un potencial significatiu. La feina presentada en aquesta secció explora com treure profit d'aquest potencial, reutilitzant les dades de marcadors socials per a l'extracció dels conceptes de domini, entenent per domini com qualsevol àrea d'interès: aprenentatge, viatges, oci en bicicleta,

¹ <http://delicious.com/>

L'objectiu d'aquest treball és desenvolupar, implementar i avaluar un mètode per extreure els conceptes que estan implícitament representats per etiquetes definides pels internautes en llocs d'etiquetatge social, com és per exemple Delicious. Les etiquetes s'han de netejar, estructurar, i agregar abans d'extreure i identificar els conceptes de nivell de domini que representen. En aquest treball es desenvolupen un conjunt d'heurístiques i procediments, s'apliquen a les etiquetes extretes d'un lloc web d'etiquetatge social, i s'analitzen els resultats obtinguts. La contribució d'aquesta investigació és demostrar la viabilitat d'extreure conceptes de nivell de domini de les etiquetes. També mostra que és possible reutilitzar el contingut generat per la web social per contribuir a la semàntica de la informació continguda a la World Wide Web.

2.1. Feina Relacionada

L'etiquetatge és l'acte de col·locar una etiqueta a un recurs que reculli una interpretació. Pot ser una etiqueta que evoqui el contingut de quelcom (per exemple, *productes ecològics*), una etiqueta que capti una impressió o acció que provoca quelcom a un usuari (per exemple: *important* o *per llegir*), o una combinació dels dos (per exemple, *productes ecològics importants*). La majoria de les etiquetes que trobem a Internet entren en la primera categoria, descrivint l'origen del element etiquetat (Furnas et al. 1987). Per tant, proporcionen una manera ràpida d'identificar aspectes clau d'un domini, o dels elements que conté. Cada etiqueta captura la semàntica sobre els recursos i conté el potencial per identificar i extreure conceptes importants en un domini (Weikum et al. 2009).

Una col·lecció d'etiquetes sobre un domini concret pot ser vist com el coneixement que té un col·lectiu de persones sobre aquell domini, fet conegut pel terme anglès *Knowledge of the Crowd*. Aquest coneixement reflecteix les opinions independents i diverses d'un grup de persones (Sunstein 2006). Les etiquetes d'aquesta *folksonomy* (Angeletou et al. 2007) estan sovint força relacionades entre si, ja que poden ser utilitzades per un usuari per etiquetar recursos relacionats, o per diferents usuaris per etiquetar el mateix recurs (Wu, Zhang, and Yu 2006; Campbell 2006). Els patrons que sorgeixen d'aquestes pràctiques d'etiquetatge proporcionen, no només una manera d'organitzar la informació per als usuaris (Kipp and Campbell 2006), sinó també una capa de semàntica feble (Dye 2006) que es pot explotar amb l'objectiu d'identificar i extreure conceptes de domini.

2.2. Identificant Conceptes de Domini a partir d'Etiquetes

El mètode desenvolupat en aquest treball consta de diferents fases, representades a la figura 1. Aquest article es centra en les dues primeres fases: la neteja de dades i la identificació de conceptes rellevants i les seves interrelacions. La resta de fases es poden veure descrites amb més detall a (Sugumaran et al. 2010).

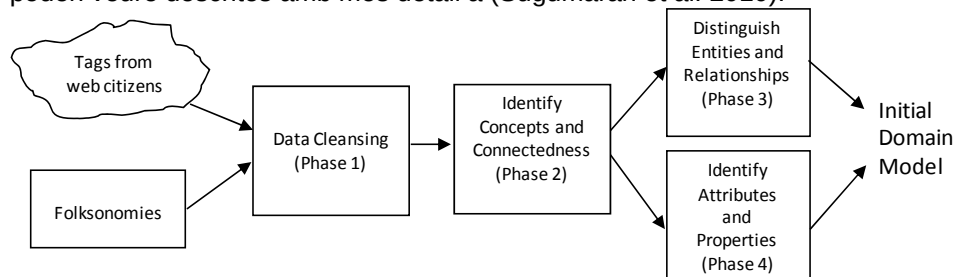


Fig. 1. Mètode per identificar conceptes de domini a partir d'etiquetes

2.2.1. Fase 1: *Data Cleansing*

Abans de començar a treballar amb les dades obtingudes d'un lloc web d'etiquetatge social, cal netejar les dades obtingudes. Aquesta primera fase, *data cleansing*, consisteix a detectar i corregir els registres corruptes o inexactes (Maletic and Marcus 2000; Tan, Han, and Elmasri 2002) obtinguts en el lloc web d'etiquetatge social. El terme *data cleansing* s'utilitza en el context de la mineria de dades, i es refereix a la identificació de les peces incompletes, incorrectes, inexactes o irrellevants de les dades per tal de corregir aquests errors mitjançant la substitució, modificació o supressió d'aquestes peces de dades.

La fase de neteja de dades és important perquè la majoria dels llocs web d'etiquetatge no restringeixen ni validen les etiquetes aportades pels usuaris. El resultat és una considerable quantitat d'etiquetes incorrectes, inexactes i de vegades duplicades. Per exemple, l'ús de puntuació en les etiquetes s'utilitza sovint per compensar la incapacitat de llocs web com *delicious.com* a gestionar paraules clau compostes. Etiquetes com ara *coolwebsites*, *cool_web*, *cool-site*, *coolsite* representen, per tant, el mateix. A més, diverses etiquetes representen variacions de la mateixa paraula (per exemple, *bank* i *banking*) o etiquetes diferents per a un mateix concepte, però en diferents idiomes (per exemple, el concepte *environment* representat per les etiquetes *meio_ambiente*, *meioambiente*, *Medio_ambiente*, *medi_ambient*). La fase de neteja de dades, per tant, utilitza múltiples tècniques per solucionar aquests problemes. Algunes de les operacions realitzades en aquesta fase inclouen:

- la combinació de formes singulars i plurals de les etiquetes (per exemple, *auctions* i *auction*),
- la conversió de totes les etiquetes en minúscules (per exemple, *static* i *Static*).
- substitució de caràcters especials, com ara [_,.,,;,.] (Per exemple, *web 2.0* i *web2_0*),
- la conversió de diferents conjugacions dels verbs a la seva forma en infinitiu (per exemple, *awaiting* o *awaits* per *to await*),
- l'expansió d'abreviatures (per exemple, *sw* per *SouthWest*),
- l'ús de sinònims (*achieved* per *attain*),
- l'esborrat dels recursos web que no han estat etiquetats amb cap etiqueta,
- l'esborrat de les etiquetes dels usuaris que han etiquetat pocs recursos: per exemple, descartant les etiquetes dels usuaris que han etiquetat menys de x recursos, on x és una variable configurable, i
- aplicant una llista de transformacions creades addhoc per cada domini. Aquesta llista de transformacions és útil per tractar errors terminològics, neologismes i abreviatures (per exemple, ampliar *diy* per "*do it yourself*").

Alguns exemples d'aquestes transformacions realitzades en el procés de *data cleansing* són el *e-commerce* -> *Ecommerce* (traient el guió), *Auktionen* -> *auction* (traducció d'etiquetes en diferents idiomes), *Banking* -> *bank* (retirar el gerundi), *Finacial* -> *finance* (eliminació el 'ial ', un adjectiu de formació de sufix), *supervise* -> *oversee* (mapatge de paraules amb significat semblant), *fig* -> *figure* (mitjançant l'expansió d' abreviatures), i *OCW*-> *OpenCourseWare* (mitjançant l'expansió de les abreviatures manualment indicades). El procés descrit fins aquí presenta una neteja de dades basada en informació lingüística, per tant, pressuposa que totes les etiquetes entrades són correctes (sense errors ortogràfics). Per tant, si una persona etiqueta un recurs amb l'etiqueta *banquing* el nostre sistema no serà capaç de detectar que aquesta etiqueta es refereix al mateix concepte que *banking*. Els errors ortogràfics són molt freqüents en les etiquetes de llocs web d'etiquetatge social, fent que sigui especialment rellevant el seu tractament en el nostre cas.

Per tal de solucionar aquest problema hem estès el *data cleansing* amb algunes operacions fonètiques. En particular, hem usat traductors fonètics. Els traductors fonètics són algorismes que permeten calcular el codi fonètic d'una paraula. Aquests codis fonètics indiquen com és pronuncia la paraula. Per tant, si donades dues paraules diferents obtenim el mateix codi fonètic, llavors podem dir que les dues paraules sonen igual. Això és especialment útil per detectar errors d'ortografia, ja que en la majoria dels casos el codi fonètic d'una paraula correctament expressada i la mateixa paraula amb errors d'ortografia sonen igual o de forma molt semblant (veure figura 2). Exemples d'això són *Auction* i *Auktion*, *Alicante* i *Alikante* o *Recommendation* i *Recomendation*.

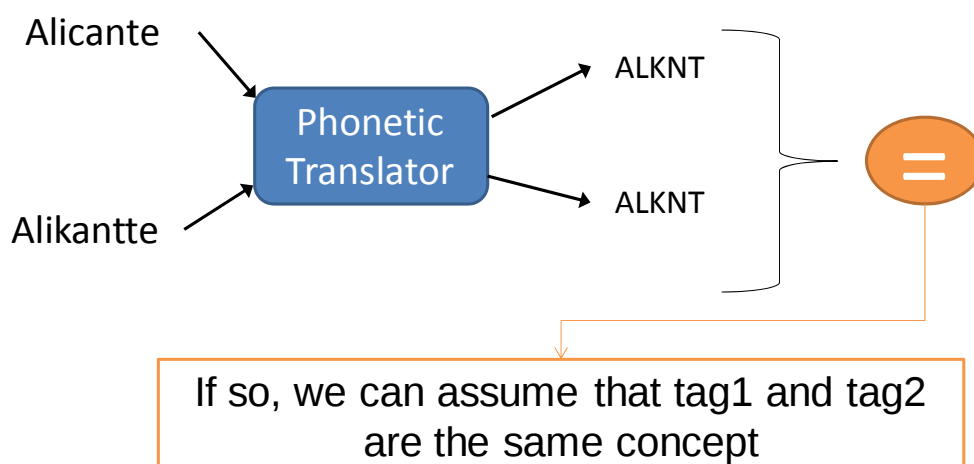


Fig. 2. Ús de traductors fonètics per detectar errors ortogràfics

Després de testejar diferents traductors fonètics, s'ha acabat escollint l'algoritme *Methapone* (Philips 2000), ja que dona molt bon resultat amb paraules de qualsevol longitud i té en compte les lletres veïnes a l'hora de crear el codi fonètic. Per tal de detectar la similitud entre diferents codis fonètics s'ha utilitzat la funció de distància de *Levenshtein* amb una variació *Damerau*. La funció de *Levenshtein* permet calcular com de diferents són dos codis fonètics. El resultat d'aquesta funció és un número anomenat distància de *Levenshtein* i indica el cost necessari per convertir un codi fonètic en l'altre. Per exemple la distància entre SGNL (codi fonètic de *Signal*) i SKNL (codi fonètic de *Signaling*) seria 1, ja que tan sols canvia la segona lletra. La variació de *Damerau* aplicada permet reduir la distància en el cas que s'intercanviï l'ordre de dues lletres, per exemple en *Signal* i *Singal*, que és un cas molt comú en l'etiquetatge web. En el cas anterior la distància de *Levenshtein* seria 2 (SGNL - SNGL), mentre que utilitzant la variació de *Damerau* la distància seria 1.

La Figura 3 mostra els resultats que es van obtenir en el procés de *data cleansing* en la consulta *corporate sustainability*. La consulta va recuperar 26,073 *bookmarks*. Per a cada *bookmark* es va registrar el recurs etiquetat, l'usuari que el va etiquetar, i totes les etiquetes amb que cada usuari el va etiquetar. Això va produir 10.030 etiquetes. La neteja de dades va retallar aquest nombre a 8,472 etiquetes. Cal dir que el rati de neteja obtingut en els casos testejats ha estat d'entre un 10 i un 15%.

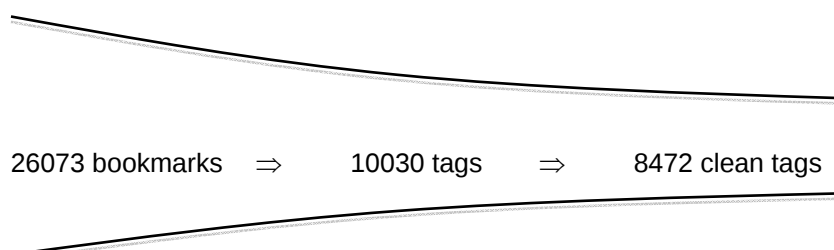


Fig. 3. Resultats de la fase de *data cleansing* per la consulta de *Corporate*

Sustainability

2.2.2. Fase 2: Identificar els conceptes de domini i les seves relacions.

Aquesta fase detecta els conceptes més rellevants del domini i com estan relacionats entre sí. Es recolza en sis heurístics per avaluar la probabilitat de que una etiqueta sigui un concepte de domini i per identificar les relacions entre els conceptes candidats. Al començar la fase, la probabilitat de que una etiqueta sigui un concepte de domini és zero, els diferents heurístics permeten augmentar progressivament aquesta probabilitat a mesura que l'etiqueta en qüestió vagi complint condicions favorables. Els heurístics utilitzats són:

Heurístic 1: Nombre d'usuaris que utilitzen una etiqueta. Com més gran sigui el nombre d'usuaris que utilitzen una etiqueta, major és l'acord de que l'etiqueta representa un concepte de domini. Aquest heurístic dona pes a les etiquetes utilitzades per multitud d'usuaris.

Heurístic 2: Nombre de recursos etiquetats amb una etiqueta. Aquest heurístic suggereix que el nombre de recursos etiquetats amb una etiqueta determinada indica la importància d'aquesta etiqueta per al domini.

Heurístic 3: Freqüència d'ús d'etiquetes. Aquest heurístic analitza la freqüència de les etiquetes i utilitza l'argument que com més gran sigui la freqüència, més important serà l'etiqueta.

Heurístic 4: Centralitat de l'etiqueta. Aquest heurístic interpreta la centralitat de cada etiqueta en el núvol d'etiquetes, tenint en compte la seva posició dintre del núvol d'etiquetes i la mida del núvol al que pertany.

Fins aquí els heurístics utilitzats per detectar els conceptes potencialment rellevants pel domini. Aquests passen dels usuaris a les etiquetes, i finalment als recursos. En cada pas, s'augmenta la confiança de que una etiqueta sigui un concepte de domini.

La relació entre les etiquetes s'avalua mitjançant la mesura de les freqüències de la co-ocurrència. La co-ocurrència pot indicar que les etiquetes són sinònimes, que representen conceptes relacionats que finalment estaran connectats per un tipus de relació, o bé que representen les construccions relacionades (per exemple, una és una entitat, i l'altre és un atribut). Aquesta heurística només considera les etiquetes identificades pels heurístics anteriors com a candidates per ser conceptes de domini. La possible relació entre els conceptes representats per les etiquetes seleccionades s'estudia utilitzant dos heurístics:

Heurístic 5: Ocurrencia: dues etiquetes estan connectades quan s'utilitzen per etiquetar el mateix recurs. Per exemple, les etiquetes *travel* i *flights*, identificades després d'executar la consulta "Air Travel", co-ocorren 31 vegades en els primers 100 resultats retornats.

Heurístic 6: Tendència: dues etiquetes estan connectades quan una quantitat significativa dels usuaris tendeixen a utilitzar les dues etiquetes per etiquetar el mateix recurs. Per al càlcul d'aquest heurístic utilitzem Anàlisi de Components Principals (PCA) (C. Ding and He 2004). El PCA s'ha executat amb els següents paràmetres: ús de la màxima versemblança, una rotació Varimax i els camps només amb un pes major o igual a 0.3 es tenen en compte.

L'anàlisi PCA identifica un conjunt de factors que conceptualment poden ser vistos com clusters, on les etiquetes representen un punt de vista des del que es pot contemplar el domini. Les etiquetes de cada grup poden estar semànticament connectades, ja que tracten amb el mateix concepte semàntic o sub-domini. Per tant, les etiquetes d'un clúster donat poden ser usades per inferir les etiquetes potencialment rellevants per a l'usuari (si un usuari decideix que una de les etiquetes d'un clúster és rellevant, totes les seves etiquetes són potencialment rellevants). Per exemple, després d'aplicar l'anàlisi de components principals amb els resultats obtinguts de la consulta de "online auction", alguns dels factors obtinguts (o clústers) són els següents: 1) *craft, handmade, art, design, gifts* i *diy*; 2) *daily, technology, gadgets, woot*, i *electronics* i 3) *paypal, money, finance* i *bank*. Es fàcil de veure que el primer cluster està relacionat amb l'artesania i venda de coses fetes a mà, el segon relacionat amb tecnologia i el tercer relacionat amb el pagament de les subhastes.

Al final d'aquesta fase, s'obté un conjunt d'etiquetes que conté els possibles candidats a conceptes de domini, i un conjunt de possibles relacions entre aquests conceptes. L'objectiu de les fases 3 i 4, explicades amb més detall a (Sugumaran et al. 2010), és identificar, per cada concepte de domini triat, sí es tracta d'una entitat, un atribut o una relació.

2.3. Implementació

El mètode presentat s'ha aplicat per analitzar les etiquetes obtingudes a partir de *Delicious.com*. Un usuari pot introduir un terme o una frase relacionada amb un domini per recuperar els documents i etiquetes de *delicious* que tracten amb aquest domini. El prototip implementat executa els processos de *data cleansing* de manera supervisada i aplica els heurístics de forma automàtica. El resultat del procés, els conceptes seleccionats i la connexió entre ells, els hem representat visualment en forma de grafs locals, com es mostra a la Figura 4. Actualment, un component

important dels experiments és l'anàlisi de components principals que s'utilitza per calcular la relació entre etiquetes. El cost computacional d'executar el PCA és assumible ja que s'executa al final de tot del procés de selecció de conceptes candidats i per tant sobre un nombre d'etiquetes relativament baix. Per exemple, en els casos descrits en aquest document, l'anàlisi de components principals involucrats ha estat del voltant de 100 etiquetes per cada un.

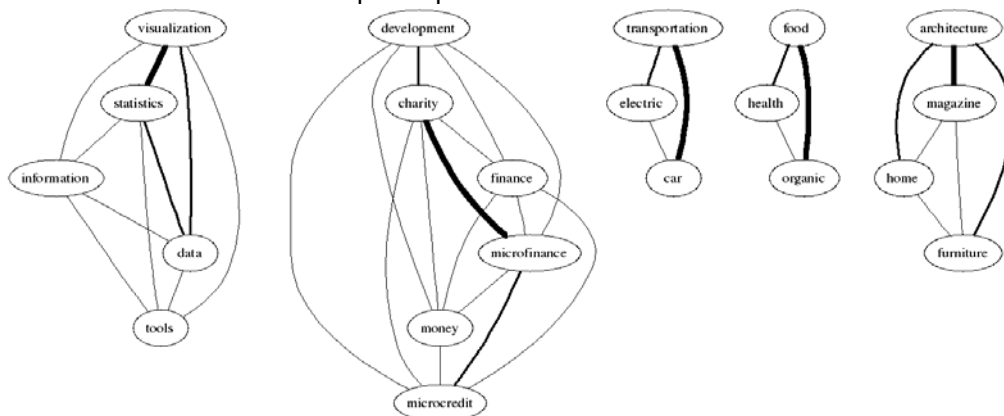


Fig. 4. Grafs locals parcials obtinguts de per la cerca de *Corporate Sustainability*²

A partir dels l'anàlisi PCA s'obtenen factors (també anomenats components). Cadascun d'aquests està compost per un conjunt d'etiquetes. Encara que el PCA no és una tècnica de clustering, el conjunt d'etiquetes en cada component pot ser considerat com un clúster en el que les etiquetes representen tendències d'anotació. Per exemple, en el cas de *corporate sustainability* podem interpretar un dels factors, com una indicació de que un nombre significatiu de persones tendeixen a utilitzar juntes les etiquetes *food*, *organic* i *health*. Les etiquetes dins de cada clúster poden, per tant, considerar-se semànticament connectades (vegeu Figura 4, on el gruix de la línia indica la correlació entre les etiquetes connectades). La Taula 1 mostra els primers factors (clústers) obtinguts a partir de l'anàlisi de PCA per a la consulta "*corporate sustainability*".

Un anàlisi de l'exemple de "*corporate sustainability*", mostra que cada component descriu un subdomini diferent. També mostra que el primer i cinquè component no estan relacionats amb el domini, sinó que provenen dels marcadors d'un lloc web tangencial al domini. Entenent aquests llocs web com llocs que descriuen altres dominis, però que poden contenir etiquetes rellevants pel domini cercat. Un exemple d'això seria una notícia que parli d'un accident de cotxe on un dels afectats és un cotxe elèctric, que podria respondre al tercer graf local de la figura 4 tot i no tractar amb "*corporate sustainability*". L'anàlisi mostra que un procediment basat en el nombre d'ocurrències pot esbiaixar els resultats. Seria útil estendre la fase de *data*

² En el gràfic tan sols es mostren els grafs locals dels cinc factors més rellevants, és a dir, els que expliquen més part de la variança.

cleansing per donar suport en la detecció d'aquests llocs web tangencials. Això permetria als usuaris decidir si cal tenir en compte aquests llocs web abans de començar a fer el anàlisi i per tant millorar els resultats obtinguts.

Taula 1. Etiquetes contingudes en els factors obtinguts per la consulta “*corporate sustainability*”.

Corporate Sustainability	
1 st	Visualization + statistics + information + data + tools
2 nd	Development + charity + finance + microfinance + money + microcredit
3 rd	Transportation + electric + car
4 th	Food + health + organic
5 th	Architecture + magazine + home + furniture
6 th	Trends + future + strategy + consulting
7 th	Inspiration + flash + portfolio + webdesign
8 th	Architecture + collaboration + opensource
9 th	Movie + film + documentary

Independentment de les possibles millores que pot patir el mètode proposat, els gràfics locals generats per a cada component presenten diversos sub-dominis interessants que podrien ser difícils de predir fins i tot per un expert en el domini. Exemples d'això són els sub-dominis relacionats amb microfinances (segon graf local de la figura 4) i arquitectura (a la dreta de la figura 4). Aquest mètode s'ha executat també a d'altres camps d'aplicació amb els següents resultats: subhastes en línia (102 conceptes de domini a partir de 4,635 recursos), viatges aeris (150 conceptes a partir de 11.760 recursos), i respostes d'emergència (225 conceptes de 1.664 recursos).

2.4. Discussió

Aquesta secció ha descrit un mètode per analitzar les etiquetes dels llocs d'etiquetatge social per extreure conceptes de domini. Els resultats dels experiments realitzats demostren la viabilitat de l'enfocament. La principal contribució d'aquesta investigació és la creació d'un mètode que permet identificar conceptes de domini i la seva interrelació a partir d'un enfocament de “*knowledge of the crowd*”. En particular, l'anàlisi PCA proporciona informació útil per ajudar en la identificació dels conceptes d'un domini. Els resultats mostren que la selecció (o rebuig) de les etiquetes no pot ser automàtica, sinó que requereix de retroalimentació per part d'un usuari. Els resultats també posen de manifest els reptes, com ara el caràcter genèric i l'ambigüitat de les etiquetes que requereixen neteja, i la necessitat d'agrupar etiquetes que representen el mateix concepte. A més, els components generats a partir de l'anàlisi de components principals (com *microfinance* i *architecture*) apunten a la possibilitat que

l'enfocament proposat té el potencial per superar amb escreix les limitacions de les perspectives d'experts en els models de domini, tenint en compte una gamma molt més àmplia de conceptes que poden ser part de la representació d'un domini.

La recerca futura se centrarà en fer experiments addicionals, millorar el procés de neteja de dades per augmentar l'eficàcia al descartar etiquetes irrelevantes i fusionar-ne d'equivalents, el desenvolupament de tècniques per extreure més idees de les etiquetes, i l'ús de variables difuses. També seria convenient refinar l'heurística, l'aplicació de les fases restants, i realitzar anàlisis empíriques.

3. ITINER@: Ús d'ontologies en Sistemes d'Informació Geogràfica

El turisme és una àrea que ha experimentat un gran creixement en els últims anys degut principalment a l'aparició de noves tecnologies (com ara Internet, la web social, els dispositius mòbils amb GPS, etcètera) i a l'aparició d'informació geogràfica accessible de forma gratuïta i l'abast de tothom (com per exemple la que podem obtenir de *Openstreetmaps*³, *Wikiloc*⁴, *Google Maps*⁵, i *Bing Maps*⁶). L'aparició i desenvolupament d'Internet ha provocat un gran impacte en la manera com les persones accedeixen a la informació sobre les seves destinacions turístiques, contracten els seus viatges, o reserven hotels i/o viatges en avió (Prantner et al. 2007). De fet, en els últims anys han aparegut nombroses aplicacions que proporcionen la capacitat d'oferir productes de major complexitat que satisfacin les demandes dels diferents viatgers depenent de les seves preferències (J. de la Flor, J. Borràs, Y. Pérez, A. Russo, S. Antón-Clave, A. Moreno, A. Valls 2011; Mínguez, Berrueta, and Polo n.d.; Niemann, Mochol, and Tolksdorf 2008).

La gran quantitat de dades turístiques a Internet, la seva descentralització i la seva falta de normalització fa molt difícil la tasca dels agents o portals de serveis turístics a l'hora d'oferir l'usuari la informació completa i actualitzada que demana. La selecció de les dades turístiques que són rellevants per a l'usuari encara guanya més importància quan entren en joc dispositius mòbils amb una capacitat de procés i emmagatzematge limitats, com poden ser dispositius *iPhone* o *GoogleAndroid*.

Per gestionar la informació disponible a la web és imprescindible afrontar el repte de localitzar, processar i integrar tota la informació rellevant. Com la majoria de la informació a la web ha estat generada sense cap tipus de control o organització, és

³<http://www.openstreetmap.org/>

⁴<http://es.wikiloc.com/wikiloc/home.do>

⁵<http://maps.google.com/>

⁶<http://www.bing.com/maps/>

necessari utilitzar tecnologies que permetin que els ordinadors processin aquesta informació des d'un punt de vista semàntic per a la seva classificació i posterior ús. D'aquesta manera, sorgeix el concepte de Web Semàntica (Lee, Hendler, and Lassila 2001). Les ontologies són el pilar bàsic de la web semàntica i tal com s'ha demostrat en diferents projectes, el món de les ontologies i el del turisme es poden barrejar, de manera que utilitzar ontologies amb informació turística permeti als usuaris obtenir rutes turístiques de manera senzilla.

El principal objectiu de la feina presentada en aquest apartat és estudiar l'ús d'ontologies en el context d'aplicacions turístiques. Amb aquesta finalitat, l'article presenta les principals ontologies, taxonomies i glossaris turístics utilitzats fins ara. Posteriorment, es proposa una ontologia turística per donar suport en la creació de rutes turístiques personalitzades.

Aquesta secció s'estructura de la següent forma: la següent subsecció presenta diferents ontologies que contenen informació turística, i per tant amb un gran potencial en aplicacions de l'àmbit del turisme. Per treure el màxim rendiment a la informació, una aplicació turística no només cal que inclogui informació turística, sinó també informació sobre el perfil de l'usuari, els seus amics i les seves preferències, informació sobre els punts d'interès, etcètera. En la subsecció 3.2 es presenta l'arquitectura de l'aplicació Itiner@, un assistent turístic que permet oferir rutes turístiques personalitzades, i es descriu l'ontologia que s'ha creat per donar suport a aquest sistema. Finalment, la subsecció 3.3. presenta les conclusions i el treball futur previst.

3.1. Ontologies en l'àmbit del turisme

Les ontologies es poden classificar en: ontologies de domini, que capturen conceptes sobre un domini d'aplicació particular com per exemple viatges, reserves, futbol o menjar, i ontologies d'alt nivell o propòsit general, que contenen gran quantitat d'informació sobre el món real i són independents del domini (Guarino 1998). L'ontologia d'alt nivell més coneguda és el Cyc (del terme anglès *encyclopedia*), que té origen en un projecte que pretenia capturar el coneixement que es té sobre el món i codificar-lo en una base de coneixement. La majoria d'ontologies d'alt nivell contenen informació turística i geogràfica i, per tant, podrien ser usades per assistir als sistemes informàtics turístics, tal com es pot veure a la figura 5. El principal problema d'utilitzar aquest tipus d'ontologies és la seva dificultat d'ús (Conesa, V. C. Storey, and Sugumaran 2010), que provoca que tinguin una menor adopció i dificulti la interoperabilitat.

Respecte a les ontologies de domini, en l'àmbit del turisme s'han desenvolupat diferents catàlegs de dades i taxonomies per facilitar la gestió de la informació als

agents turístics. Actualment existeix un gran nombre d'ontologies que estan en un estat de maduresa molt avançat. Aquestes ontologies han estat representades segons diferents nivells de formalisme (des de tesaurus a ontologies formals), d'acord a diferents objectius (des d'aspectes genèrics de l'àmbit turístic fins a subdominis més específics per descriure escenaris més concrets, com ontologies d'àmbit regional) i d'acord a diferents llicències d'ús (des de públiques fins a propietàries). Les ontologies propietàries, com *Mondeca* i *Hi-Touch*, han estat descartades en l'estat de l'art d'aquest document a causa de la impossibilitat de poder consultar el seu codi i per tant no poder realitzar una anàlisi crítica sobre això.

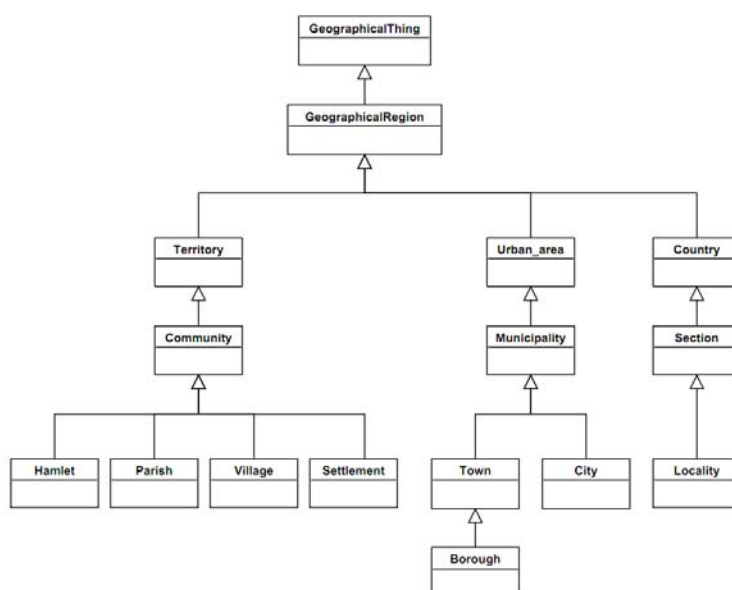


Fig. 5. Fragment de l'ontologia Cyc que conté informació geopolítica

A continuació es fa un repàs d'algunes ontologies existents utilitzades en l'àmbit del turisme. Començant amb l'estudi de tesaurus i especificacions turístiques i continuant amb informació sobre algunes de les ontologies turístiques més importants. Finalment, donarem algunes pinzellades de *OpenLinkedData* i de com utilitzar-la en l'àmbit del turisme.

3.1.1. Tesaurus i Especificacions

Hi ha una sèrie d'especificacions que han servit de base per al desenvolupament de la majoria de les ontologies que veurem. Les dues més importants són el tesaurus de l'Organització Mundial de Turisme (*World Tourism Organization* o *WTO*) i l'especificació de l'Open Travel Alliance (*OTA*).

L'Organització Mundial de Turisme ofereix un tesaurus de conceptes turístics i activitats de lleure per ajudar a la classificació de la informació relacionada amb les activitats turístiques. Aquest diccionari pot ser utilitzat com una guia terminològica relacionada amb el turisme i té l'objectiu de normalitzar aquests termes per aconseguir una classificació estàndard de qualsevol document de tipus turístic. Conté més de 1500 conceptes expressats en diferents idiomes (anglès, francès i espanyol)⁷. Aquest tesaurus ha servit de base per al desenvolupament de les principals ontologies existents en l'actualitat.

La *OpenTravel Alliance* (OTA) és una organització sense ànim de lucre formada per organitzacions que representen a tots els segments de la indústria del turisme (línies Aéreas, empreses hoteleres, de lloguer de cotxes, etc.), així com també proveïdors de tecnologia i serveis. El seu objectiu és facilitar la interoperabilitat en el desenvolupament de programari entre companyies de l'àmbit turístic. La seva activitat principal és el desenvolupament i manteniment d'una biblioteca d'esquemes XML per al seu ús en la indústria del turisme. Aquests esquemes XML conformen l'especificació *OpenTravel XML*. Aquesta especificació conté més de 200 documents XML agrupats en paquets que defineixen conceptes i entitats del sector turístic, i permeten normalitzar el format dels missatges utilitzats per facilitar la comunicació entre sistemes de diferents companyies⁸.

3.1.2. Ontologies Turístiques

Harmonise: És l'element central de *HarmoNET* (*Harmonisation Network for the Exchange of Travel and Tourism Information*), una xarxa internacional per a la normalització i l'intercanvi de dades en la indústria turística (DE NORMALISATION and NORMUNG 2009).

L'objectiu de l'ontologia *Harmonise* és proporcionar a les organitzacions turístiques la capacitat d'intercanviar informació sense necessitat de modificar les seves estructures internes ni els seus sistemes d'informació. A causa del gran nombre d'agents turístics, no és factible construir un mediador genèric, i per tant es va crear una ontologia que contingues un conjunt mínim dels conceptes més rellevants dins de la indústria del turisme, anomenada *IMHO* (*Interoperable Minimum Harmonization Ontology*). *Harmonise* està implementada en el llenguatge *RDF*⁹ i conté al voltant de 200 conceptes i propietats, centrats fonamentalment en el camp de l'allotjament (e.g. hotels, bed & breakfast, turisme rural, càmpings, etc.), dels esdeveniments i activitats (e.g. festivals, conferències, esdeveniments esportius, etc.), de la gastronomia, i dels monuments i llocs d'interès.

⁷ <http://wtoterm.wto.org/multiterm/>

⁸ <http://www.opentravel.org/Specifications/default.aspx?spec=2011A>

⁹ <http://www.w3.org/RDF/>

Actualment, l'ontologia *Harmonise* s'ha utilitzat amb èxit en nombrosos projectes. En l'actualitat, la xarxa *HarmoNET* està composta per més de 20 companyies i continua oferint els seus serveis com a mediador entre sistemes.

QALL-ME: L'ontologia *QALL-ME* (Ou et al. 2008) apareix per donar suport a un sistema de resolució de preguntes dins del sector del turisme, que permeti respondre les preguntes dels usuaris en llenguatge natural i en diferents idiomes.

L'ontologia *QALL-ME* proporciona un model conceptual que cobreix multitud d'aspectes de l'àmbit del turisme, com a destinacions turístiques (e.g. ciutats, pobles, etc.), llocs turístics (e.g. allotjament, llocs d'interès, gastronomia, etc.), esdeveniments (e.g. representacions, esdeveniments esportius, etc.) i mitjans de transport. Està més centrada a representar informació turística estàtica que activitats (e.g. viatges, rutes turístiques, etc.). *QALL-ME* està escrita en OWL-DL i està formada per classes, categoritzades en 15 apartats i 107 propietats.

DERI e-Tourism: L'ontologia *DERI e-Tourism* va ser desenvolupada en el context del projecte *OnTour*. L'objectiu d'aquesta és la creació d'un portal web de suport a les recerques turístiques utilitzant tecnologies de Web Semàntica (Siorpaes and Bachlechner 2006).

L'ontologia està centrada en la descripció d'allotjaments i infraestructures turístiques, amb l'objecte de proporcionar tota la informació que pugui ser rellevant en la recerca de paquets turístics. Inclou a més conceptes per descriure activitats d'oci i dades geogràfiques. Per maximitzar la interoperabilitat, l'ontologia va ser creada a partir del thesaurus de la WTO, afegint propietats i relacions als conceptes escollits. L'ontologia inclou aspectes geogràfics (i.e. adreces postals o coordenades GPS), per facilitar el càlcul de distàncies entre objectes turístics, i temporals, per oferir més personalització.

EON Travelling Ontology: Va ser una de les primeres ontologies turístiques a desenvolupar-se i descriu conceptes turístics classificats en diverses categories, com reserves de viatges i allotjament, mitjans de transport, llocs d'interès i allotjaments. En l'actualitat sembla haver caigut en desús atesa la poca informació existent sobre la mateixa.

cDott: Té l'objectiu de donar suport a la interoperabilitat dels actors turístics en operacions de baix nivell. Les seves sigles provenen de (*the Core Domain Ontology for Travel and Tourism*), i s'ha desenvolupat a partir de les ontologies *Harmonise* i *EON*.

cDott (Barta et al. 2009) proposa una estructura modular que integra diferents ontologies de domini específiques. Com a node central de l'estructura es troba una ontologia que conté dades genèriques sobre el sector turístic, definint objectes, esdeveniments i destinacions turístiques. Al voltant d'aquesta es connecten diferents

ontologies de propòsit més específic i altres independents del domini turístic per estendre el seu abast i funcionalitat com ara, ontologies relacionades amb l'allotjament o gastronomia, o ontologies relacionades amb dades temporals, climatològiques, monetàries, de perfil de usuari, etc. La seva estructura modular permet una millor personalització de les dades proporcionades a l'usuari, permetent escollir diferents models turístics en funció del tipus i les preferències del viatger.

Aquesta ontologia presenta un gran avenç dins de l'àmbit turístic, però malauradament sembla estar encara en fase de desenvolupament.

TAGA Travel Ontology: Està centrada en l'àmbit dels viatges i escrita en OWL. Forma part del projecte *TAGA*, un agent experimental per a la simulació d'un mercat mundial de viatges a Internet (Zou et al. 2003).

GETES: Va ser desenvolupada fa més de 10 anys per donar suport a una eina de recerca turística a la web i a la posterior classificació i interpretació semàntica de les seves respostes, que permeti a l'usuari escriure les seves consultes mitjançant llenguatge natural. L'ontologia GETESS conté més de 1000 conceptes i 200 relacions, proporcionant termes bilingües (en anglès i alemany) per a cadascun dels seus conceptes (Staab et al. 1999).

CRUZAR: S'ha desenvolupat dins del projecte "*Un visitante, una ruta*", finançat per l'Ajuntament de Saragossa. El seu objectiu és la construcció d'una aplicació per al càlcul de rutes turístiques a la ciutat de Saragossa sobre la base del perfil i context de l'usuari. L'ontologia permet modelitzar l'organització general del domini turístic i garantir la seva interoperabilitat amb altres sistemes similars, i assegurar també la compatibilitat semàntica futura amb altres extensions, en cas que sigui necessari afegir informació sobre nous recursos relacionats amb el turisme. S'ha creat a partir de l'ontologia d'alt nivell *DOLCE*.

L'ontologia CRUZAR captura la semàntica de tres tipus d'entitats: recursos turístics de la ciutat de Saragossa (i.e. monuments, restaurants, parcs, allotjaments, activitats d'oci, etc.), perfils d'usuari (e.g. atributs de situació i de preferències d'usuari) i rutes turístiques, descrivint les característiques d'una visita turística (i.e. punts d'interès o els esdeveniments relacionats) (Mínguez et al. n.d.).

ANOTA: Desenvolupada sota el projecte *ANOTA*, l'objectiu principal és permetre l'ús d'anotacions semàntiques a Internet. L'ontologia reutilitza parts d'altres ontologies turístiques existents, així com els esquemes XML de l'especificació de la *OTA*, per definir conceptes en el camp turístic (Murua, Llado, and Llodr 2006).

Com a conclusió, podem dir que hi ha nombroses ontologies turístiques que s'utilitzen de manera efectiva en l'àmbit del turisme. A part de les presentades aquí podem trobar altres, com ara *OnTourism* o *e-Tourism*, que hem decidit no tractar a causa de

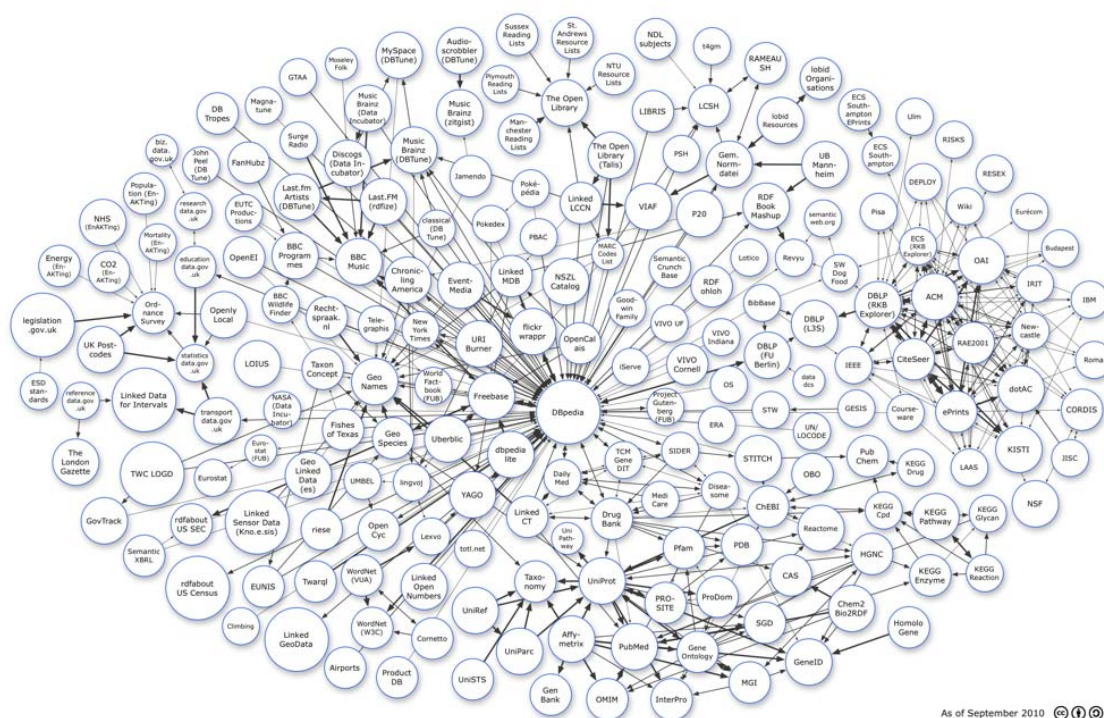
restriccions d'espai i al seu reduït àmbit d'aplicació (centrades a Àustria i Madeira). Potser, el principal problema de les actuals aplicacions turístiques és la interoperabilitat. Els tesaurus i ontologies creats fins al moment permeten una millor interoperabilitat entre aplicacions, però la solució dista molt d'estar a prop. Una possible solució als problemes d'interoperabilitat podria ser la utilització d'ontologies ben conegudes i adoptades per la majoria de la comunitat com ara les del projecte *OpenLinkedData*.

3.1.3. OpenLinkedData: Què és i quin suport pot oferir en el turisme?

OpenLinkedData¹⁰ (OLD a partir d'ara) és una iniciativa que busca interconnectar la informació de la web d'una manera que sigui comprensible per a programes informàtics. La forma de realitzar això és definint un conjunt d'ontologies en RDF, on cada ontologia descriu l'estructura de la informació d'un lloc web i conté informació sobre el lloc web en forma d'instàncies de l'ontologia. D'aquesta manera, un programa és capaç d'interpretar la informació que conté el lloc web a partir de l'estructura de la seva ontologia i obtenir informació sobre les seves dades a partir de les seves instàncies. Però el poder de OLD no està en les seves ontologies, sinó en les seves interrelacions. Aquestes ontologies relacionades entre si, possibiliten que els programes naveguin d'un concepte a altres conceptes relacionats, ja sigui en un mateix domini o en dominis diferents. Per exemple aquest sistema permetria que l'ontologia *LinkedGeoData*, que conté informació sobre *OpenStreetMaps*, estigués relacionada amb *DBpedia*, que conté informació sobre *Wikipedia*, permetent que punts d'interès de *LinkedGeoData* (com ara la ciutat de Girona) es puguin relacionar amb la seva entrada en *Wikipedia* via *DBpedia*, o fins i tot amb fotos de Girona via l'ontologia de *Flickr*.

A la figura 6 podem veure el núvol d'ontologies incloses en OLD i les seves interrelacions. *DBpedia* és l'ontologia central que manté i relaciona totes les ontologies. En el context del turisme i la informació geogràfica hi ha diferents ontologies a tenir en compte, com són: *LinkedGeoData* (ontologia de *OpenStreetMaps*), *GeoLinkedData(es)* (conté dades geoespacionals del territori nacional espanyol), *Geonames* (ontologia de topònims), *Eurostat* i *Eurostat(FUB)* (contenen informació sobre estats i províncies de la Unió Europea), *Telegraphis* (conté dades de continents, països, capitals i monedes utilitzades a cada país), *UN /Locode* (conté informació sobre més de 70.000 localitzacions de 243 països) i *LinkedSensorData* (informació sobre sensors - malauradament només sensors dels EUA).

¹⁰ <http://linkeddata.org/>



As of September 2010

Fig. 6. Ontologies incloses a la iniciativa *Linked Data* i les seves interrelacions

És cert que les ontologies de OLD tendeixen a ser menys formals i a tenir una estructura de menys qualitat que les ontologies de domini presentades anteriorment, però el gran nombre d'ontologies disponible, la gran quantitat de dades que contenen, el seu nivell de consens i la gran quantitat d'interrelacions entre elles fa que siguin molt més convenientes d'utilitzar que una aproximació més formal però menys interoperable. Aquest és, potser, el motiu pel qual cada dia més gent decideix utilitzar les ontologies d'OLD en comptes de reutilitzar o crear altres ontologies de domini o d'alt nivell.

3.2. Ús d'Ontologies per la Creació de Rutes Turístiques: ITINER@.

Les aplicacions SIG en l'àmbit del turisme són aplicacions sensibles al context. Això vol dir que la informació que han de subministrar en cada moment és diferent depenent del context de l'usuari: quin usuari la demana i les restriccions temporals i espacials, entre d'altres. Això implica que si volem afegir semàntica a una aplicació cal tenir en compte la semàntica de diferents dominis: tants com factors condicionin el context en l'aplicació a desenvolupar.

En el cas que ens ocupa (el projecte Itiner@) s'ha creat un assistent de rutes turístiques per a la zona d'Esterrri d'Àneu que funcioni en un dispositiu mòbil. Aquest

assistent haurà de proposar rutes turístiques als usuaris en funció de les seves preferències, del lloc a visitar, de la franja horària de la visita i de les limitacions del dispositiu mòbil que l'usuari utilitzarà per realitzar la ruta. Si volem utilitzar ontologies per donar suport a l'aplicació, necessitarem afegir ontologies per als següents dominis: turístic (informació sobre els punts d'interès, la seva localització i documents associats, com ara fotos, descripcions i vídeos), temporal (definició del factor temps i algunes funcions per tractar informació temporal), personal (informació sobre l'usuari, les seves preferències i els seus hàbits en les rutes realitzades fins al moment), i dispositius mòbils (informació sobre dispositius mòbils i les funcionalitats i limitacions de cada un d'ells). Altres ontologies que podríem necessitar en funció dels requisits necessaris serien: ontologies de topònims per facilitar la recerca d'elements geogràfics (*Geonames* és un ontologia de OLD que conté informació sobre topònims), informació relativa als usuaris i les seves comunitats socials (especificacions com *FOAF* en conjunció amb *SIOC* podrien ser útils per a tal aspecte) o informació de tercers sobre punts d'interès (*DBPedia* i *Flickrwrappr* són exemples d'ontologies que podrien utilitzar per enllaçar informació o fotos a punts d'interès).

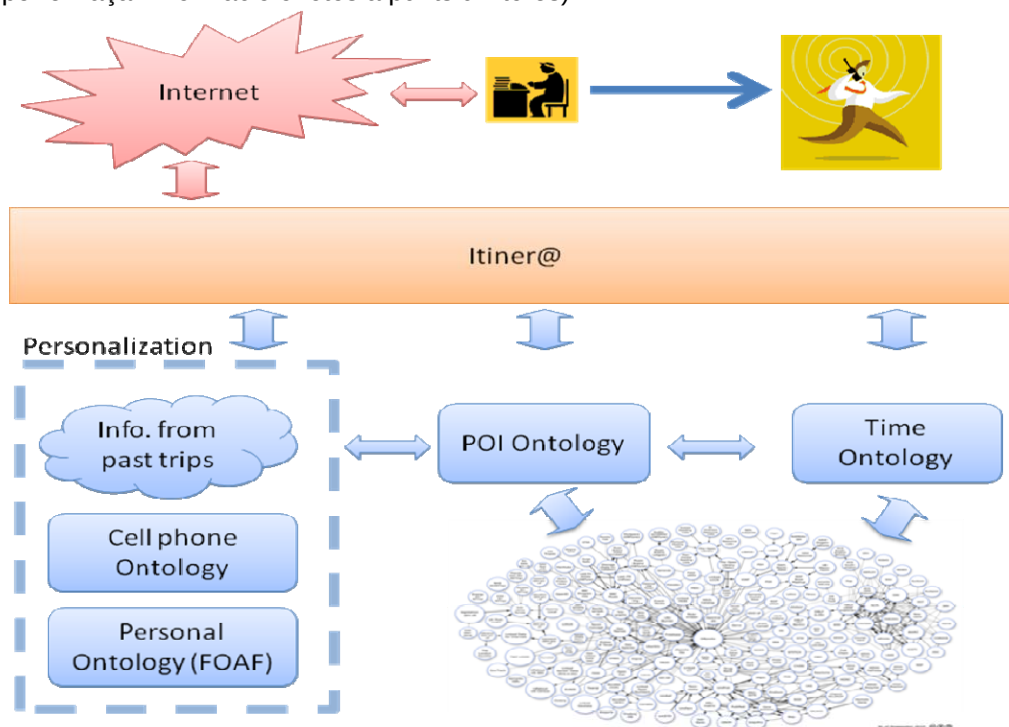


Fig. 7. Esquema de l'arquitectura del sistema Itiner@

La figura 7 presenta l'esquema de l'arquitectura del sistema Itiner@. Com podem veure, hi haurà diverses ontologies de suport a la personalització, que contindran informació sobre l'usuari, sobre el dispositiu mòbil i sobre els viatges realitzats fins al moment. Aquesta informació s'enllaçarà amb l'ontologia de punts d'interès, que al seu

torn estarà enllaçada amb una ontologia temporal que donarà suport a la gestió del temps. Notar que l'ontologia de punts d'interès i la de temps estaran relacionades amb les ontologies d'OLD per maximitzar la seva interoperabilitat. En particular, l'ontologia de punts d'interès estarà enllaçada amb les ontologies *LinkedGeoData* i *Geonames*. El sistema *Itiner@* utilitzarà la informació de les ontologies per presentar rutes personalitzades a l'usuari via una interfície web. Un cop l'usuari hagi escollit la ruta a realitzar, la ruta i tota la seva informació associada (Punts d'interès¹¹ a visitar, Pol alternatius, Pol d'emergències, restaurants, descripcions, fotos, vídeos ...) es traslladarà al dispositiu mòbil perquè l'usuari pugui realitzar la ruta sense estar connectat a Internet.

Actualment *Itiner@* està en fase de desenvolupament, i fins ara s'ha creat un primer prototip que ofereix una personalització bàsica i conté una ontologia de Pol enllaçada amb *LinkedGeoData* de OLD. S'ha decidit no reutilitzar directament l'ontologia *LinkedGeoData* perquè conté elements no necessaris en el nostre context i perquè no disposa d'alguns elements necessaris, com és el cas de conceptes de tipus "zona històrica", "museu etnogràfic", etc. Per aquest motiu, s'ha creat una ontologia que usa els conceptes rellevants de *LinkedGeoData* i els estén amb els conceptes necessaris per *Itiner@*. Com les noves classes s'han creat com a extensió de les classes existents de *LinkedGeoData* es garanteix la interoperabilitat en els dos sentits i per tant ens permet accedir directament a tots els punts d'interès disponibles a *OpenStreetMaps*.

3.2.1. L'Ontologia d'Itiner@

En aquesta secció és descriu breument l'ontologia creada pel projecte *Itiner@*, identificant el coneixement que permet representar, alguns exemples de la seva estructura i les seves interrelacions amb d'altres ontologies existents. Per més informació sobre l'ontologia desenvolupada veure (L. Descamps, J. Casas, J. Conesa 2012).

L'objectiu de l'ontologia és representar la informació necessària per tal de donar suport a un servei basat en la localització de l'usuari en les seves rutes turístiques. Per tant, l'ontologia contindrà informació tant sobre els punts d'interès, com sobre les preferències turístiques dels usuaris i com informació sobre les rutes visitades. Tot i que l'ontologia s'ha creat específicament per aquest projecte, l'objectiu és alinear-la amb les ontologies turístiques més reconegudes. Amb aquesta idea en ment, s'ha alineat l'ontologia presentada amb *LinkedGeoData*¹² (per representar els Pol),

¹¹ Anomenats Pol a partir d'ara.

¹² <http://linkedgeodata.org/About>

WGS_84¹³ (per representar les coordenades geogràfiques) i amb la *Time Ontology*¹⁴ (per emmagatzemar i tractar informació temporal).

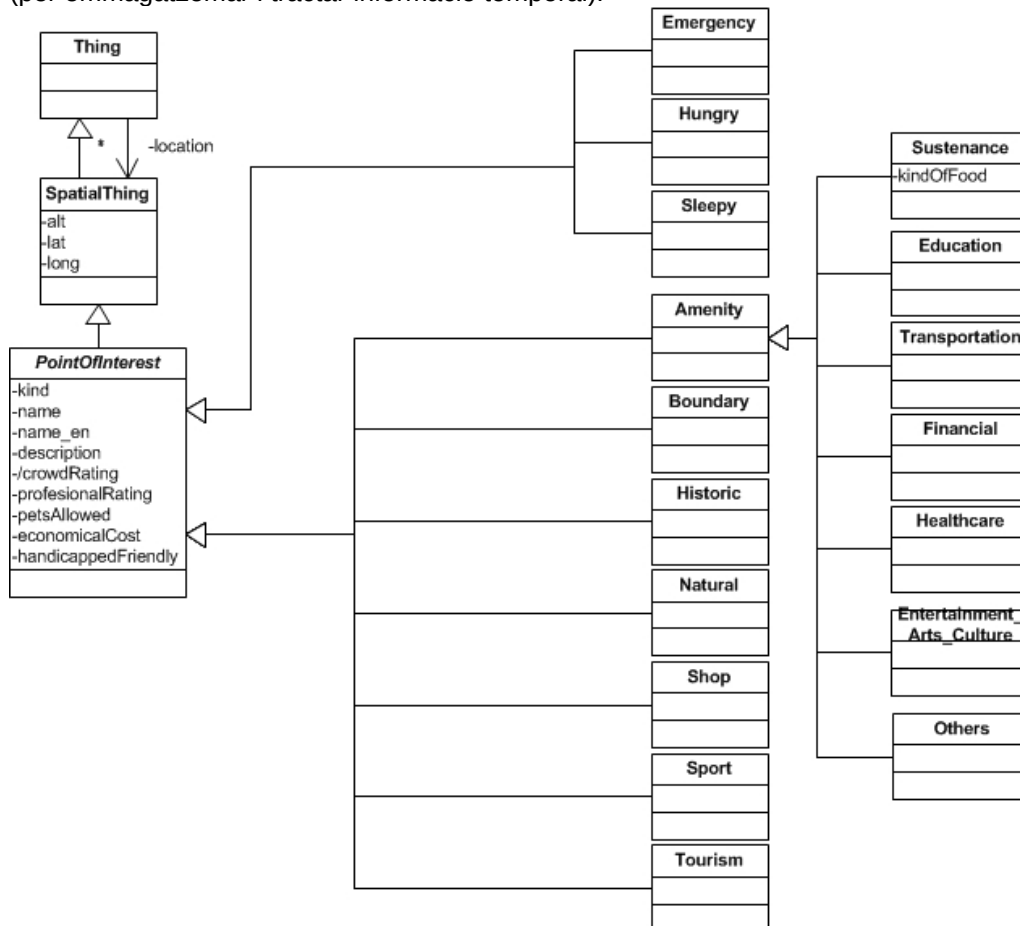


Fig. 8. Fragment de l'ontologia relacionat amb la classe *Punt d'Interès*

Probablement l'element més cèntric de l'ontologia és la classe "punt d'interès", també anomenada PoI (veure figura 8). Un punt d'interès representa qualsevol element espacial que mereix ser esmentat per la seva rellevància en un sistema de rutes turístic i/o comercial (per tant inclou tant monuments o atractius turístics com botiges, restaurants i hospitals). Algunes de les classes relacionades amb PoI s'han reutilitzat d'altres ontologies, com per exemple la classe *SpatialThing*, que ve de l'ontologia WGS_84. A més, l'especialització de classes de PoI s'ha basat en les categories de punt d'interès definides en l'ontologia *OpenLinkedGeodata*, és a dir, pel seu tipus i segons la necessitat que permet cobrir. Això permet que la nostra ontologia sigui directament compatible amb *Open Street Maps*. Com que l'especialització de PoI

¹³

<https://www1.nga.mil/ProductsServices/GeodesyGeophysics/WorldGeodeticSystem/Pages/default.aspx>

¹⁴ <http://www.w3.org/TR/owl-time/>

permet classificació múltiple, un punt d'interès pot ser instància de més d'un subtipus de Pol, permetent que un edifici *històric* sigui també una *botiga* per exemple.

Un punt d'interès pot tenir recursos associats (com ara imatges, vídeos, informació textual sobre el lloc, mapes, etc.). Aquests recursos no estaran emmagatzemats localment sinó que s'utilitzaran les seves URI per accedir-hi a través d'Internet. D'altra banda, un dels factors crítics en les aplicacions turístiques és el factor temps: cal tenir en compte aspectes com l'horari d'obertura dels punts d'interès, el calendari d'obertura, el temps necessari per visitar-lo, etc. Per tal d'emmagatzemar i operar eficientment amb variables temporals s'han reutilitzat conceptes de l'ontologia *Time Ontology*.

Un altre element rellevant a tenir en compte en l'ontologia són les preferències dels usuaris. Aquestes preferències han de poder ser emmagatzemades per ser capaços d'oferir diferents rutes en funció de les preferències de l'usuari i el context en que es trobi. Cal dir que les preferències són dependents del context, per tant podran variar en funció de si l'usuari viatja sol, en parella o amb nens. La figura 9 mostra un fragment de l'ontologia que s'encarrega d'emmagatzemar les preferències d'usuari. Un usuari pot definir un paquet de preferències d'acord als seus interessos sobre el tipus de punt d'interès que desitja visitar, els moments del dia en que prefereix fer la visita o la durada de la mateixa, les restriccions econòmiques, les preferències geogràfiques d'acord a la zona que li agrada a l'usuari o a la que acostuma a visitar, i en funció de la seva companyia, és a dir, si viatja amb nens, amb amics, amb mascotes o amb persones amb discapacitats. D'altra banda, el sistema també permet definir preferències compostes, que combinin dos o més d'aquest tipus de preferències per delimitar els seus desitjos més específicament. Això permetria representar coses com "quan viatjo sol m'agrada visitar punts d'interès històrics, però quan vaig amb nens prefereixo anar a punts d'interès infantils i acabar la ruta no més tard de les 5 de la tarda".

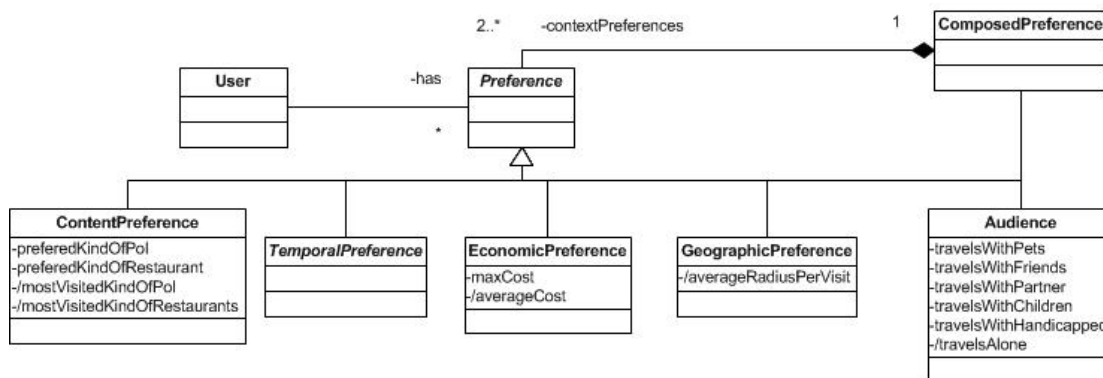


Fig. 9. Fragment de l'ontologia relacionat amb les preferències d'usuari

L'ontologia permetrà representar preferències explícites i implícites. Les preferències explícites són aquelles que el usuari ha introduït en el sistema. Les preferències implícites són les que s'extrauran després d'analitzar totes les rutes realitzades per cada usuari. Per exemple, algú pot dir que li agrada visitar esglésies romàniques i a l'hora de la veritat sempre visitar esglésies gòtiques. Ambdós tipus de preferències es tindran en compte per personalitzar les rutes dels usuaris. Per tal de tenir en compte les preferències implícites, la ontologia permetrà representar informació històrica de les rutes fetes pels usuaris (veure figura 10). Per tant, per cada ruta realitzada, es permetrà representar la ruta original (la que va planificar l'usuari), la ruta final (la que va acabar fent l'usuari) i els recursos consultats dels punts d'interès de la ruta. Quan l'usuari tingui activat el GPS també es podrà enregistrar els temps d'entrada i sortida de cada PoI, permetint inferir les desviacions de l'usuari respecte els temps mitjos de visita i viatge. Aquesta informació ens serà molt útil també per detectar recursos poc rellevants o errors en els punts d'interès.

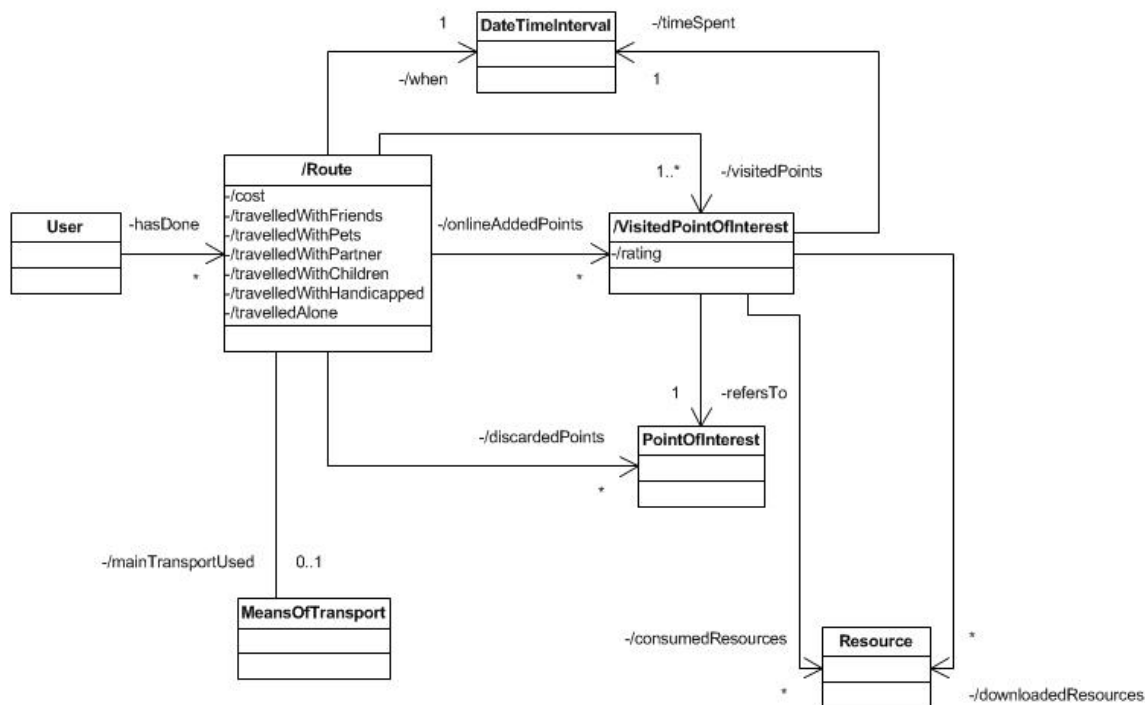
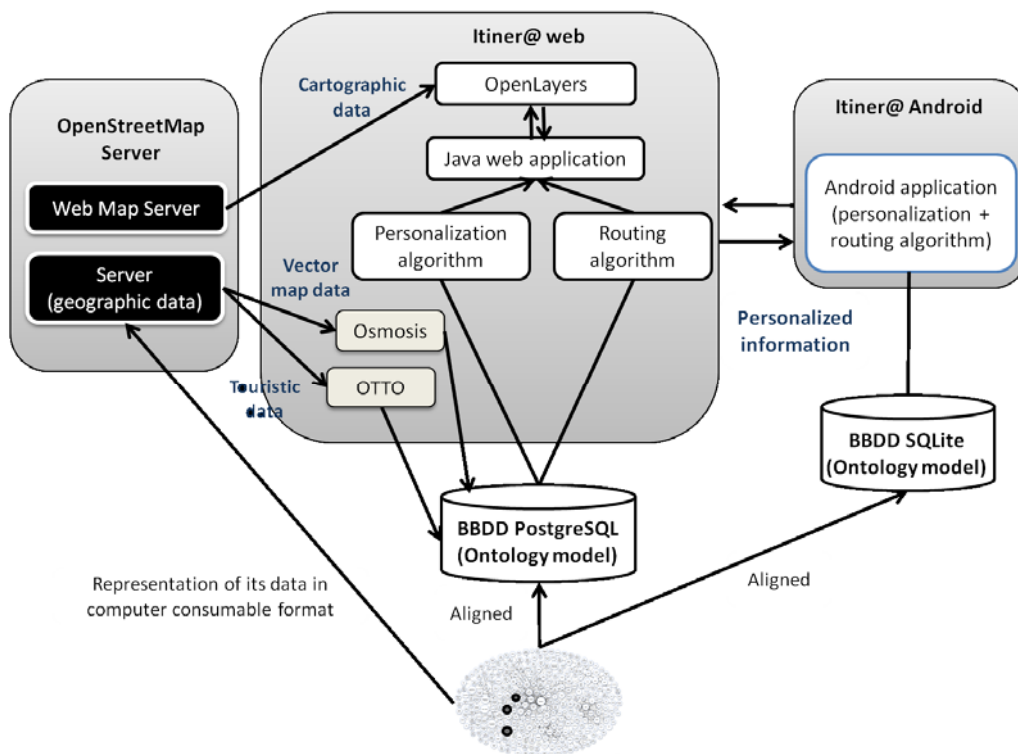


Fig. 10. Fragment de l'ontologia relacionat amb l'històric de rutes

3.2.2. Implementació

La figura 11 mostra l'arquitectura final del prototipus d'Itiner@. Degut a les limitacions dels dispositius mòbils (tant d'eficiència, com tecnològics, com la manca d'aplicacions pel tractament d'ontologies i informació geogràfica eficientment) hem hagut de

La informació geogràfica s'obté de *OpenStreetMap*. Les imatges dels mapes s'obtenen en línia a través del servei de mapes d'OSM i es processen i es mostren en el ordinador del client utilitzant la biblioteca de *OpenLayers*. D'altra banda, l'aplicació obté del servidor de OSM informació turística i vectorial sobre els punts d'interès. Totes les dades turístiques s'emmagatzemen en *PostgreSQL*, fent servir l'esquema de base de dades que correspon a l'ontologia desenvolupada. Les dades són carregades en la base de dades fent servir un programa anomenat *OTTO - OSM To Touristic Ontology*.



Més tard, l'algorisme de personalització utilitza les dades turístiques, les preferències dels usuaris, i la data i l'hora en que es vol fer la ruta per proposar els punts d'interès disponibles més atractius per a cada usuari, sigui quin sigui el seu context. Aquests punts s'introdueixen en l'algorisme d'encaminament, que proporciona una ruta d'accés adequat a través dels carrers per anar des d'un punt a un altre. Tota

aquesta informació s'envia de tornada a OpenLayers i es dibuixa al mapa presentat a l'usuari.

Degut a les limitacions dels dispositius mòbils i a la poca connectivitat en molta part del territori, el servidor només els envia la informació clau, que és: la ruta personalitzada, les dades turístiques dels punts d'interès de la ruta, les imatges dels mapes, i la informació geogràfica (en format vectorial) dels mapes de la regió de la ruta.

L'aplicació mòbil s'ha implementat sobre el sistema operatiu Android i té l'objectiu de mostrar les rutes personalitzades i donar suport a l'usuari en la seva ruta, ja sigui informant de punts d'interès alternatius, del compliment o no del horari de la ruta, o bé refent la ruta quan sigui necessari. Aquesta aplicació té la majoria de funcionalitats que ofereix el servidor, però sense necessitat d'estar connectat a Internet. De fet, l'aplicació mòbil tan sols requereix connexió a Internet per descarregar les rutes turístiques. La figura 12 mostra un exemple en el que l'aplicació mòbil presenta una ruta en el poble d'Esterri d'Aneu. Les xinxetes blaves representen els punts d'interès a visitar, la línia blava la ruta a seguir i la línia vermella el tram actual. Per mostrar els mapes obtinguts de OSM l'aplicació utilitza la biblioteca Osmroid¹⁵, que proveeix funcions similars a OpenLayers.



Fig. 12. Pantalla d'exemple de l'aplicació d'Itiner@ pel mòbil

¹⁵ <http://code.google.com/p/osmdroid/>

3.3. Discussió

Tal i com s'ha vist en aquesta secció, actualment, existeixen multitud d'ontologies turístiques i geogràfiques que es poden utilitzar en aplicacions SIG turístiques. Potser el més difícil és triar quina utilitzar en cada cas concret, ja que una mala elecció pot malbaratar o complicar excessivament un projecte. En l'elecció cal tenir en compte si les dades que ha d'oferir l'aplicació són pròpies i relativament estàtiques (en tal cas serà més interessant triar una ontologia de domini que s'adapti a les nostres necessitats), o si es vol accedir a dades públiques ubicades a la xarxa (en aquest cas l'ús d'ontologies del núvol de OLD és aconsellable ja que garanteix tenir accés a gran quantitat de dades públiques actualitzades). Les ontologies d'alt nivell són una opció quan es necessiten ontologies molt formals, però la seva magnitud i dificultat de maneig les desaconsellen en gran mesura.

Un altre dels condicionants a l'hora de decidir quina ontologia usar és l'arquitectura en la que correrà l'aplicació a desenvolupar. Si, com en el nostre cas, es vol crear una aplicació per dispositius mòbils, llavors caldrà tenir en compte el grau de formalisme de l'ontologia i la quantitat de dades que conté. Això és important perquè les eines disponibles en dispositius mòbils que permeten tractar amb ontologies són poques, parcials i amb poques funcionalitats, tal i com es pot veure en l'anàlisi realitzat a (L. Descamps, A. Perez-Navarro, J. Conesa 2011). Algunes de les principals limitacions d'aquestes eines són que: 1) la majoria d'elles no permeten tractar ontologies en format OWL, 2) poques eines permeten treballar amb consultes SPARQL, i cap d'elles de forma eficient, 3) no hi ha eines que permetin l'emmagatzematge i processament eficient d'informació geogràfica (en particular no hi ha bases de dades geogràfiques), i 4) hi ha poques eines d'encaminament que permetin recalculer rutes sense connexió a Internet, i l'eficiència de les que hi ha podria millorar-se.

El treball presentat en aquesta secció estudia com utilitzar informació semàntica per millorar aplicacions SIG en el sector de turisme. Per això s'enumeren diferents ontologies de caràcter general, turístiques i del projecte *Open Linked Data* que podrien utilitzar-se de suport en una aplicació SIG de l'àmbit del turisme. També es presenta l'arquitectura d'una aplicació SIG que permet oferir rutes personalitzades en funció de les preferències d'usuari (horari de la visita, dispositiu mòbil i lloc a visitar) i el paper que juguen les ontologies en aquesta arquitectura, i és mostra una ontologia que s'ha creat per especificar informació de punts d'interès, de preferències i hàbits d'usuaris i d'informació temporal. Aquesta ontologia s'ha alineat amb l'ontologia *Open Linked Geodata* per facilitar la importació de dades des de i cap a ella. Per mostrar la utilitat de l'ontologia, s'ha integrat en una aplicació mòbil turística anomenada Itiner@, que utilitza la informació de l'ontologia per crear una ruta personalitzada per a cada usuari en funció de: preferències, situació personal, mitjà de transport que fa servir, zona que visita i data i durada de la visita. La completesa i utilitat de l'ontologia s'han validat mitjançant l'ús de Itiner@ al poble d'Esterri d'Àneu.

La principal contribució del treball presentat és mostrar com utilitzar una ontologia per incrementar les funcionalitats dels sistemes d'informació turístics via l'ús d'informació semàntica. Val a dir que tot i que s'ha utilitzat la informació de l'ontologia de manera efectiva en el sistema Itiner@, es preveu en el futur permetre una major personalització a partir d'un estudi més profund de les seves dades i una millora de l'algoritme de encaminament utilitzat.

4. Conclusions

Aquest article presenta algunes de les experiències portades a terme amb l'objectiu de definir metodologies, tècniques i bones pràctiques per a donar suport a un ús més eficient de les ontologies en aplicacions informàtiques. Els treballs presentats en el document permeten abordar dos problemes fonamentals en l'ús eficient d'ontologies, que són la reutilització de coneixement preexistent per crear una nova ontologia des de zero que s'adapti a les necessitats dels usuaris (secció 2) i l'adaptació i extensió d'ontologies preexistents per donar suport a un sistema d'informació geogràfic mòbil amb funcionalitats avançades (secció 3).

Les ontologies, si s'usen de forma efectiva, poden ser importants per reutilitzar coneixement i promoure la interoperabilitat (Conesa, V. C. Storey, et al. 2011). No obstant, les aproximacions actuals que permeten desenvolupament i reutilització d'ontologies no han tingut massa èxit. Aquest article ha presentat un petit estat de l'art de com reutilitzar coneixement existent al crear noves ontologies i en la reutilització d'ontologies en l'àmbit de sistemes d'informació geogràfica. Les propostes presentades en aquest treball presenten guies per millorar la creació i la reutilització d'ontologies, promovent la reutilització i la interoperabilitat entre les diferents ontologies i especificacions actuals.

D'acord amb la nostra experiència, creiem que la reutilització de les ontologies depèn de les pròpies ontologies (com està representada i quin coneixement conté) i en les eines disponibles necessàries per a la seva gestió. En particular, es necessiten eines que permetin superar les mancances en la documentació de les ontologies, normalment inexistent, i en la seva representació. Per exemple, les ontologies grans, com el Cyc o com el núvol de *Open Linked Data*, necessiten eines per navegar, consultar i representar gràficament l'ontologia (Conesa et al. 2010). Per tant, eines d'aquest tipus poden ser molt útils per tal de, en primer terme, identificar les ontologies que poden ser rellevants per un problema concret i, en segon terme, quin és el subconjunt de coneixement d'aquestes ontologies a reutilitzar. D'altra banda, altres aspectes que condicionen fortament la reutilització d'una ontologia són: la seva mida,

la seva documentació, els llenguatges que utilitza, la seva capacitat per descriure diversos dominis, l'entorn on s'aplicarà, i la seva completesa, qualitat i redundància.

Finalment, en línia de les propostes presentades en aquest document, actualment estem treballant en la manera d'estendre el mètode de creació d'ontologies a base d'incrementar més l'ús del *crowd knowledge*. En particular, en addició als mètodes lingüístics i fonètics s'està treballant per utilitzar informació de webs 2.0 en l'etapa de *data cleansing* i en la creació de nous heurístics. L'ús d'aquest tipus d'informació ens permetrà identificar neologismes o paraules molt específiques d'un domini determinat, identificar relacions entre conceptes que són tan sols certes en determinats dominis d'aplicació i desmbiguar de forma semi-automàtica. Tot això és factible ja que, actualment, el coneixement que contenen els llocs web col·laboratius, com per exemple Wikipedia, és més extens i està més adaptat a diferents contextos d'aplicació que les ontologies formals existents, com per exemple WordNet o Cyc. D'altra banda, s'està treballant en l'extensió del projecte Itiner@ per aconseguir un grau de personalització més elevat i reaprofitar encara més els recursos disponibles al núvol de *Open Linked Data*. En particular, es vol estendre la informació lligada als punts d'interès a partir d'identificar els recursos de *Open Linked Data* que hi estiguin relacionats. Altres àmbits on els autors estan treballant per reutilitzar el coneixement de les ontologies existents de forma efectiva són els àmbits d'e-Learning, de modelització conceptual i de la millora de consultes web.

Agraïments

Aquest treball ha estat realitzat amb el finançament del Internet Interdisciplinary Institute (IN3), en el seu Programa d'Investigadors Residents del curs 2010-2011.

Referències Bibliogràfiques

- Angeletou, S, Sabou, M, Specia, L and Motta E. 2007. "Bridging the gap between folksonomies and the semantic web: An experience report." in *Workshop: Bridging the Gap between Semantic Web and Web*, vol. 2.
- Barta, R, Feilmayr, C, Pröll, B, Grün, C and Werthner, H. 2009. "Covering the semantic space of tourism: an approach based on modularized ontologies." Pp. 1-8 in *Proceedings of the 1st Workshop on Context, Information and Ontologies*.
- Campbell, D G. 2006. "A phenomenological framework for the relationship between the semantic web and user-centered tagging systems."

- Conesa, J, Storey, V C, Sugumaran, V and Caballé, S. 2011. "Knowledge Reuse: Its Role and Implications for Upper Level Ontologies." Pp. 233-240 in *Intelligent Networking and Collaborative Systems (INCoS), 2011 Third International Conference on*.
- Conesa, J, Storey, V C and Sugumaran, V. 2010. "Usability of upper level ontologies: The case of ResearchCyc." *Data & Knowledge Engineering* 69(4):343-356.
- Conesa, J, Caballé, S, Gañan, D and Prieto, J. 2011. "A Framework for Dealing with Knowledge about On-line Collaborative Learning Sessions." in *Third International Conference on Intelligent Networking and Collaborative Systems*.
- Descamps, L, Perez-Navarro, A, Conesa, J, Casas, J. 2011. "Cómo introducir semántica en las aplicaciones SIG móviles: expectativas, teoría y realidad." in *V Jornadas SIG Libre de Girona*.
- Descamps, L, Casas, J, Conesa, J, Pérez-Navarro, A. 2012. "Itiner@: Ontología para la personalización de rutas turísticas en dispositivos móviles." in *VI Jornadas SIG Libre de Girona*.
- Ding, C, and He, D. 2004. "K-means clustering via principal component analysis." P. 29 in *Proceedings of the twenty-first international conference on Machine learning*.
- Dye, J. 2006. "Folksonomy: A game of high-tech (and high-stakes) tag." *EContent* 29(3).
- Furnas, G W, Landauer, T K, Gomez, L M and Dumais, S T. 1987. "The vocabulary problem in human-system communication." *Communications of the ACM* 30(11):964-971.
- Guarino, Nicola. 1998. "Formal Ontology and Information Systems." Pp. 3-15 in. IOS Press.
- Hendler, J, and Golbeck, J. 2008. "Metcalf's Law Applies to Web 2.0 and the Semantic Web." *Journal of Web Semantics* 6(1):14-20.
- de la Flor, J, Borràs, J, Pérez, Y, Russo, A, Antón-Clave, S, Moreno, A, Valls, A, Isern, D. 2011. "SigTur/E-Destino: fusión de un SIG con métodos de Inteligencia Artificial para la gestión de regiones turísticas complejas." in *V Jornadas SIG Libre de Girona*.
- Kipp, M and Campbell, D G. 2006. "Patterns and inconsistencies in collaborative tagging systems: An examination of tagging practices." *Proceedings of the American Society for Information Science and Technology* 43(1):1-18.
- Lee, T B, Hendler, J and Lassila, O. 2001. "The semantic web." *Scientific American* 284(5):34-43.

- Maletic, J I and Marcus, A. 2000. "Data cleansing: Beyond integrity analysis." Pp. 200-209 in *Proceedings of the Conference on Information Quality*.
- Murua, Idoia, Esteve Llado, and Bel Llodr. 2006. "The Semantic Web for Improving Dynamic Tourist Packages Commercialisation." *Planta*.
- Mínguez, I, Berrueta, D and Polo, L. n.d. "CRUZAR: An Application of Semantic Matchmaking to e-Tourism."
- DE NORMALISATION, C E, and NORMUNG, E.K.F.Ü.R. 2009. "Harmonization of data interchange in tourism."
- Niemann, M, Mochol, M and Tolksdorf, R. 2008. "Enhancing hotel search with semantic web technologies." *Journal of theoretical and applied electronic commerce research* 3:82-96.
- Ou, S, Pekar, V, Orasan, C, Spurk, C and Negri, M. 2008. "Development and alignment of a domain-specific ontology for question answering." *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.
- Philips, L. 2000. "The double metaphone search algorithm." *C/C++ Users Journal* 18(6):38-43.
- Prantner, K, Ding, Y, Luger, M, Yan, Z and Herzog, C. 2007. "Tourism ontology and semantic management system: state-of-the-arts analysis." in *IADIS International Conference WWW/Internet*.
- Siorpaes, K, and Bachlechner, D. 2006. "OnTour: tourism information retrieval based on YARS." *Posters and Demos* 31.
- Staab, Steffen et al. 1999. "GETESS - Searching the Web Exploiting German Texts." Pp. 113-124 in *Proceedings of the 3rd International Workshop on Cooperative Information Agents III/Workshop on Cooperative Information Agents*. London, UK: Springer-Verlag.
- Sugumaran, V, Purao, S, Storey, VC and Conesa, J. 2010. "On-Demand Extraction of Domain Concepts and Relationships from Social Tagging Websites." *Natural Language Processing and Information Systems* 224-232.
- Sunstein, C. 2006. *Infotopia: How Many Minds Produce Knowledge*. Oxford University Press.
- Tan, K W, Han, H and Elmasri, R. 2002. "Web data cleansing and preparation for ontology extraction using WordNet." Pp. 11-18 in *Web Information Systems Engineering, 2000. Proceedings of the First International Conference on*, vol. 2. IEEE.

- Weikum, G, Kasneci, G, Ramanath, M and Suchanek, F. 2009. "Database and information-retrieval methods for knowledge discovery." *Communications of the ACM* 52(4):56-64.
- Wu, X, Zhang, L and Yu, Y. 2006. "Exploring social annotations for the semantic web." Pp. 417-426 in *Proceedings of the 15th international conference on World Wide Web*.
- Zou, Y, Finin, T, Ding, L, Chen, H and Pan, R. 2003. "Using semantic web technology in multi-agent systems: a case study in the taga trading agent environment."

Abstract

An ontology provides a common view of a particular domain and can be seen as a conceptual schema that is supposed to be shared and understood by computers as well as humans. Nowadays, ontologies are used for solving problems in the fields of software design, knowledge representation, interoperability enhancement, web query improvement and semantic web among others, but mostly in the academic world. We can say that an ontology is used effectively for a program when it helps the program to behave more intelligently, improving some of its facets, such as providing more functionalities, improving the time response or increasing data compression.

The presented research studies how to use the ontologies effectively in real applications and therefore promote their use in real world. With this objective in mind the research has dealt with the methodologies and tools used to identify relevant ontologies for a given task, the current state of the ontologies, and the effective application of ontologies to several fields in order to improve the functionalities of their Information Systems. This paper presents two experiences focused to understand how to use ontologies effectively to solve a given problem in the field of geographic information systems and how to create ontologies in order to improve their reusability. The first experience presents a method to create ontologies that nourish from the information from Internet. Second experience reuses a set of ontologies and specifications effectively in order to create a touristic virtual assistant for mobile devices with advanced functionalities. From these experiences some guidelines to improve the effectiveness in the creation, selection and reutilization of ontologies are presented.

Keywords

Ontologies, semantics, information systems, ontology driven information systems, geographic information systems, location based services, social web, knowledge of the crowd

Resumen

Una ontología proporciona una visión común de un dominio y puede ser vista como un esquema conceptual consensuado que debe ser entendido tanto por ordenadores, como por seres humanos. En la actualidad, las ontologías se utilizan para resolver problemas en los campos del diseño del software, la representación del conocimiento, la mejora de la interoperabilidad, la mejora de consultas web y la web semántica, pero mayoritariamente en el mundo académico. Podemos decir que una ontología se utiliza efectivamente en un programa cuando su uso hace que el programa “parezca” más inteligente, mejorando alguna de sus facetas, tales como proporcionar más funcionalidades, mejorar su tiempo de respuesta o aumentar la compresión de datos.

La investigación realizada estudia cómo utilizar las ontologías de manera efectiva en aplicaciones reales y, por lo tanto, ayuda a promover su uso en el mundo real. Con este objetivo en mente se han estudiado las metodologías y herramientas necesarias para identificar ontologías (o partes de ellas) relevantes para una determinada tarea, el estado actual de las ontologías, y la aplicación efectiva de las ontologías en varios campos con el fin de mejorar las funcionalidades de sus sistemas de información. En este artículo en particular se muestran dos experiencias enfocadas a entender cómo usar efectivamente ontologías para resolver un problema concreto del ámbito de los sistemas de información geográficos y como crear ontologías que faciliten su reuso efectivo. La primera de las experiencias presenta un método para crear ontologías que reutiliza la información existente en Internet. La segunda de las experiencias utiliza un conjunto de ontologías y especificaciones de forma efectiva para crear un asistente turístico virtual móvil con funcionalidades avanzadas. A partir de estas experiencias se presentan algunas guías para mejorar la eficiencia en la creación, la selección y el uso de ontologías.

Keywords

Ontologies, semántica, sistemas de información, Ontology Driven Information Systems, Sistemas de Información Geográficos, sistemas basados en localización, web social, knowledge of the crowd

Jordi Conesa Caralt

jconesac@uoc.edu

IN3- Universitat Oberta de Catalunya

Jordi Conesa és doctor en software per la Universitat Politècnica de Catalunya des de l'any 2008 i professor de l'àrea de Bases de Dades a la Universitat Oberta de Catalunya des del 2007. Els seus interessos de recerca estan relacionats a les àrees de modelització conceptual, Ontologies, Web Semàntica, sistemes d'informació intensius en coneixement i eLearning. El seu objectiu a llarg termini és el desenvolupament d'eines i metodologies que utilitzin ontologies de manera efectiva en diversos dominis d'aplicació, com ara l'enginyeria de programari, l'eLearning o els sistemes d'informació geogràfica.

Joan Casas Roma

jcasasrom@uoc.edu

IN3-Universitat Oberta de Catalunya

Joan Casas ha estat ajudant investigador a l' *Internet Interdisciplinary Institute* (IN3) de la Universitat Oberta de Catalunya durant els anys 2010 i 2011. Els seus interessos de recerca són la Lògica, les Ontologies, la Web Semàntica, la representació del coneixement i la Filosofia de la Ciència. És enginyer tècnic en informàtica per la Universitat de Vic i actualment està fent un màster en lògica i filosofia de la ciència a la Universitat de Salamanca.

