

# UNIVERSITÀ IUAV DI VENEZIA

SCUOLA DI DOTTORATO IUAV

DOTTORATO DI RICERCA IN

NUOVE TECNOLOGIE E INFORMAZIONE TERRITORIO E AMBIENTE

CICLO XXVI

## SERVIZI AVANZATI PER IL TURISMO

*RACCOMANDAZIONE CONTESTUALE BASATA SU MODELLI ONTOLOGICI*

**Vincenzo Giannotti**



**Relatore: prof. Andrea Prati**

**Correlatore: prof. Domenico Patassini**

**Coordinatore: prof. Luigi Di Prinzio**



## RINGRAZIAMENTI

Molte persone hanno contribuito, a vario titolo e in diversi modi, alla redazione di questa tesi, con consigli, discussioni, “raccomandazioni”, progetti, domande e risposte, tempo e lavoro. Li ringrazio tutti: i miei relatori e correlatori Andrea Prati, Domenico Patassini, Luigi Di Prinzio; i miei colleghi Ketty, Giovanni, Stefano, Antonella, Anna, Niccolò, Max, Andrea, Nuccio, Nicola, Stefano, Francesco. E poi Lara, Mauro, Adriano, Dario, Clara, Sabrina e Giorgio.

Un grazie particolare a Ivano, con cui ho lavorato per tanti anni e con cui ho condiviso buona parte di questo percorso e ad Alberto che nell'ultimo periodo mi ha aiutato con la parte applicativa della tesi.

Certamente avrò dimenticato qualcuno: gli chiedo scusa.

Questo lavoro è dedicato a Susanna e Gabriele: la mia famiglia.

*“Le cose visibili possono essere invisibili. se qualcuno va a cavallo nel bosco, prima lo si vede, poi no, ma si sa bene che c’è, nella Firma in bianco, la cavallerizza nasconde gli alberi e gli alberi la nascondono a loro volta: tuttavia il nostro pensiero comprende tutte e due, il visibile e l’invisibile. E io utilizzo la pittura per rendere il pensiero visibile.” - René Magritte*



## SOMMARIO

1	Abstract .....	8
2	Articolazione della tesi .....	9
3	La ricerca del triennio .....	11
4	Il semantic web .....	14
4.1	Le origini .....	15
4.2	Cosa significa semantic web? .....	15
4.3	Ontologie e altri strumenti di organizzazione della conoscenza .....	19
4.3.1	Ontologie .....	19
4.3.2	Quali vantaggi derivano dall'impiego di una Ontologia? .....	22
4.3.3	Altri strumenti di organizzazione della conoscenza .....	22
4.3.3.1	Metadati .....	23
4.3.3.2	Tassonomie .....	24
4.3.3.3	Thesauri .....	26
4.3.4	Che differenze ci sono tra tassonomie, thesauri e ontologie? .....	27
4.3.5	SKOS – un ponte verso il semantic web .....	28
4.3.6	Progetti di riferimento per la organizzazione della conoscenza .....	29
4.4	Tecnologie e standard .....	42
4.4.1	RDF, OWL, Ragionatori, SPARQL .....	42
4.4.1.1	RDF .....	42
4.4.1.2	OWL .....	50
4.4.1.3	Ragionatori e SPARQL .....	55
4.5	Il nuovo web - network e progetti di riferimento .....	59
4.5.1	World Wide Web Consortium (W3C) .....	59
4.5.2	Open Geospatial Consortium (OGC) .....	59
4.5.3	Europeana .....	61
4.5.4	Schema.org .....	63
4.5.5	GeoNames .....	65
4.5.6	DBpedia e i Linked Data .....	66
4.5.6.1	DBpedia .....	66
4.5.6.2	Linking Open Data .....	68
5	I sistemi di raccomandazione .....	72
5.1	Difficile scegliere? .....	72

5.2	Sistemi di Raccomandazione: una storia recente .....	74
5.3	A chi serve la “raccomandazione”? .....	77
5.3.1	Le aspettative del Provider .....	77
5.3.2	Le aspettative dello User .....	78
5.4	Quali caratteristiche deve avere un Sistema di Raccomandazione? .....	79
5.5	Modelli e Tecniche di raccomandazione .....	84
5.5.1	Modelli base di raccomandazione .....	84
5.5.1.1	Modello collaborativo .....	84
5.5.1.2	Modello basato sui contenuti .....	85
5.5.1.3	Modello basato sulla conoscenza .....	87
5.5.1.4	Modello ibrido .....	89
5.5.2	Altre tecniche di raccomandazione .....	91
5.5.2.1	Ontology based .....	91
5.5.2.2	Context aware .....	94
5.5.2.3	Tecniche social .....	96
5.5.2.4	Tecniche Conversazionali .....	100
6	Raccomandazione per servizi turistici avanzati .....	104
6.1	Studio per l'impianto di un sistema di raccomandazione .....	104
6.1.1	Definizione degli obiettivi generali .....	105
6.1.2	Principali problematiche di Ricerca e Sviluppo .....	107
6.1.3	Luogo di svolgimento del progetto .....	108
6.1.4	Obiettivi specifici, attività e tempistiche .....	109
6.1.4.1	Obiettivi realizzativi del progetto .....	111
6.1.4.2	Descrizione degli obiettivi realizzativi .....	112
6.1.5	Verifica dell'esito del progetto di ricerca .....	122
6.1.5.1	Risultati disponibili a fine attività .....	122
6.1.5.2	Modalità con cui sarà verificabile l'esito dell'intera ricerca .....	123
6.1.6	Interesse tecnico-scientifico .....	125
6.1.6.1	Novità e originalità delle conoscenze acquisibili .....	125
6.1.6.2	Utilità delle conoscenze acquisibili .....	127
6.1.7	Validità industriale del progetto .....	129
6.1.7.1	Coerenza strategica e gestione del progetto .....	129
6.1.7.2	Competitività tecnologica .....	130
6.2	Il modello ontologico evocativo – sperimentazione .....	131
6.2.1	Costruzione del Modello Ontologico Evocativo “EvOntology” .....	132
6.2.2	Istanze e caratterizzazione degli Item e generazione del DataSet .....	134

6.2.3	Sviluppo del percorso evocativo .....	139
6.2.4	Realizzazione dell'applicazione “EvOntologyReader” .....	141
6.2.5	Realizzazione della applicazione “EvMobile” .....	144
6.2.6	Le tecnologie impiegate .....	145
6.2.6.1	Protégé – un editor di ontologie, open source .....	145
6.2.6.2	Apache Jena – un framework per il semantic web e i linked data .....	146
6.2.6.3	Appcelerator Titanium – un ambiente per sviluppare app mobile .....	147
7	Conclusioni e sviluppi futuri .....	149
8	Bibliografia .....	151
9	Appendice – la ricerca del triennio .....	156
9.1	Il progetto: Sistema Informativo Semantico delle risorse Culturali (SISC) .....	157
9.2	Il progetto: AdriaMuse .....	183
9.3	Il progetto: Monitoraggio e controllo partecipato della mobilità .....	211
9.4	Il progetto: Recommending to People .....	224
9.5	Il progetto: OmniTurist .....	228

## 1 ABSTRACT

La ricerca affronta il tema dello sviluppo di sistemi di raccomandazione, fondati su modelli ontologici e sulla interoperabilità semantica, per l'erogazione di servizi avanzati nel campo del turismo. La tesi compendia una serie di esperienze e di progetti sviluppati nel corso di questi ultimi anni sulle tematiche del *semantic web* e dei sistemi di raccomandazione, per arrivare a studiare e a definire l'impianto di un sistema completo applicabile al campo del turismo e della cultura, per la fornitura di un servizio innovativo di guida di un turista alla visita di una città.

L'assunto di base è che “*everything is on the Web*” e noi dobbiamo semplicemente adottare le tecniche giuste per scovare l'informazione di cui abbiamo bisogno; e questo è ancor più vero nel mercato del turismo, considerata la grande eterogeneità dell'offerta e la varietà e ampiezza delle delle fonti di informazione.

In questo lavoro dunque, si vogliono indagare metodi e tecniche innovativi che possano consentire ad un turista di utilizzare l'informazione del web per ricavare una maggior soddisfazione durante il suo percorso di visita.

## 2 ARTICOLAZIONE DELLA TESI

La tesi è articolata in tre parti.

La prima parte è dedicata alla illustrazione delle tecniche e delle tecnologie del *semantic web*, attraverso le quali l'informazione può essere razionalizzata, normalizzata, strutturata, resa in forma non ridondante e coerente dal punto di vista semantico. Tali tecniche migliorano grandemente l'accessibilità dell'informazione, anche attraverso l'impiego di linguaggi simili a quello naturale.

La seconda parte è dedicata alle tecniche e alle tecnologie di raccomandazione, attraverso le quali l'informazione può essere selezionata e proposta all'utente in una forma e in una sostanza che tenga conto delle sue personali aspettative.

La terza parte, infine, illustra l'attività di ricerca più strettamente legata alla redazione di questa tesi, anche con riferimento ad alcune applicazioni sperimentali realizzate nell'ambito di progetti condotti nell'ultimo periodo.

Questa terza parte si suddivide a sua volta in due capitoli: il primo, riguarda la redazione di un modello generale di sistema per l'erogazione di servizi di raccomandazione *location based*, in un'ottica imprenditoriale; il secondo, illustra una attività di ricerca più basilare, che ha lo scopo di compendiare e valorizzare l'insieme delle attività di studio condotte nel corso di questi ultimi anni, approfondendo in particolare il tema dei sistemi di raccomandazione contestuali, realizzati su base ontologica e che attingono da basi di dati interoperabili (i.e. *linked data*), per l'erogazione di servizi territoriali nel campo del turismo e della cultura.

L'idea che viene proposta sviluppa una ipotesi di modello nel quale, ad un generico *User* che può essere oggetto di raccomandazione (il turista nel nostro caso), possono essere suggeriti degli *Item* (offerta) aggregati in *cluster* in base a diversi parametri di compatibilità (che possono riguardare, per esempio, l'interesse del turista verso un certo tipo di visita e di vacanza: arte, divertimento, gastronomia etc.).

L'informazione suggerita allo *User* è ricavata dal web in base al contesto in cui avviene la visita ed utilizzando un modello ontologico realizzato *ad hoc* attraverso cui è possibile sviluppare un percorso conoscitivo non predeterminato. Questo modello è stato chiamato “evocativo”.

In linea di massima il modello evocativo realizzato funziona così:

- lo *User* si trova in un dato luogo in un certo momento;
- lo *User* è caratterizzato da un profilo che ne descrive i principali interessi;
- alcuni *Item* si trovano in prossimità dello *User*: una chiesa, un luogo, una persona;
- ciascun *Item* evoca una storia che può essere raccontata;
- ciascuna storia contiene dei riferimenti a nuovi *Item* che possono essere più o meno compatibili con il profilo dello *User*;
- un nuovo *Item* viene suggerito allo *User*, in funzione della sua compatibilità col profilo di quest'ultimo;
- il nuovo *Item* evoca una storia che può essere raccontata.

Il procedimento può essere reiterato indefinitamente, attingendo dalle risorse web tutte le nuove storie e realizzando una struttura su più livelli: il primo livello è quello degli oggetti che lo *User* “vede” o con cui può entrare in contatto diretto; i livelli successivi sono quelli che non vede ma che sono legati alle storie degli oggetti di primo livello. L'evoluzione del percorso conoscitivo diviene indecidibile e può portare a dei risultati sorprendenti e imprevedibili, sviluppando uno dei concetti più importanti che stanno interessando la realizzazione dei sistemi di raccomandazione di ultima generazione: il concetto di “serendipità”.

### 3 LA RICERCA DEL TRIENNIO

Come anticipato nel *Abstract*, il lavoro di tesi compendia una serie di esperienze e di progetti di ricerca sviluppati a partire dal 2007, sulle tematiche del *semantic web* e dei sistemi di raccomandazione, per applicazioni nell'ambito del turismo e della cultura:

Sistema Informativo Semantico delle risorse Culturali (SISC), progetto di ricerca realizzato tra il 2007 e il 2009 in convenzione tra Università Iuav e Regione Veneto, nel quale ho svolto una attività di “progettazione e coordinamento tecnico” (resp. Scientifico prof. Domenico Patassini, resp. Progetto prof. Antonino Porrello).

Adriamuse, progetto di cooperazione internazionale del programma IPA *Adriatic* 2007-2013, cofinanziato dalla Unione Europea, nel quale ho svolto attività di *management* di progetto (resp. Scientifico prof. Luigi Di Prinzio). Il progetto terminerà a fine aprile 2014.

Monitoraggio e controllo partecipato della mobilità, per l'erogazione di location based services, svolto tra il 2011 e il 2012 in collaborazione tra Iuav, UniSky e la società Movendo, cofinanziato dalla Regione Friuli Venezia Giulia, nel quale ho svolto un ruolo di impianto progetto e collaborazione alle attività tecniche (resp. Scientifico prof. Luigi Di Prinzio, resp. Progetto dott. Niccolò Iandelli).

Recommending to People, proposta di progetto di ricerca industriale e sviluppo precompetitivo realizzato nel 2012 in risposta ad un bando di finanziamento MIUR in seguito soppresso. Di questo progetto sono stato soggetto proponente, assieme al prof. Andrea Prati, il dott. Mauro Da Dalt e il dott. Ivano Boscolo Nale. Sui risultati di questo progetto si basa il primo capitolo dell'attività di ricerca descritta nella terza parte della tesi.

OmniTurist, progetto cofinanziato dal Ministero per lo sviluppo economico, attualmente in corso di realizzazione per conto di Andreani Tributi SpA, con la collaborazione di UniSky srl e OmniGis srl e nel quale rivesto il ruolo di responsabile di progetto per la parte UniSky. Sui dati di questo progetto si basa il secondo capitolo dell'attività di ricerca descritta nella terza parte della tesi.

Con lo scopo di non appesantire la trattazione, tali progetti vengono descritti nella

parte di Appendice, essendo frequente il loro richiamo nei successivi capitoli. I primi tre sono descritti col discreto dettaglio, riportando in parte alcuni prodotti dei progetti stessi; gli ultimi due sono accennati brevemente in quanto i contenuti di questi stessi progetti vengono ripresi nella terza parte della tesi.

# PRIMA PARTE

## IL SEMANTIC WEB

## 4 IL SEMANTIC WEB

Non è facile dire che cosa sia oggi il *semantic web*. Normalmente viene inteso come il prodotto dell'impiego di una serie di tecniche e di tecnologie in grado di rendere i contenuti del web, oltre che interpretabili dall'uomo, come sono sempre stati, anche comprensibili ai computer, così che possano essere elaborati automaticamente. La sua concezione e il suo significato originari però, sono cambiati nel tempo e si sono via via evoluti per passare da quella che fu la “visione romantica” di Sir Tim Berners Lee, che ne coniò il termine, alla attuale del W3C<sup>1</sup> sul *web of data* e sui *linked data*.

Questo diceva Berners Lee nel 2001:

*“I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web: the content, links, and transactions between people and computers. A Semantic Web, which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The “intelligent agents” people have touted for ages will finally materialize.”*

e questo dice oggi il W3C (di cui, ricordiamo, Berners Lee è fondatore e presidente):

*“In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.”*

I motivi di questa evoluzione (ovvero involuzione, per certi versi), sono dovuti al fatto che il web così come è strutturato attualmente non può diventare “semantico” in senso classico se non riscrivendo buona parte di tutto ciò che è stato fatto fino ad ora,

---

<sup>1</sup> World Wide Web Consortium – organizzazione internazionale che ha lo scopo di definire standard e tecnologie per lo sviluppo del web. Comprende alcune centinaia di membri (380 nel 2012) tra aziende informatiche, università e centri di ricerca, aziende telefoniche, altre aziende e organizzazioni.

nei termini che il *semantic web* richiederebbe per poter essere considerato tale.

La visione attuale è diversa: meno profonda e pervasiva rispetto alla visione originaria, ma senz'altro più coinvolgente e facilmente realizzabile da un punto di vista applicativo e soprattutto basata sui due concetti che oggi sono da considerare imprescindibili: il *collaborative networking*<sup>2</sup> e la *collective intelligence*<sup>3</sup>.

#### **4.1 LE ORIGINI**

*“The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation” (Tim Berners Lee - 2001)*

Questa famosa definizione, da cui prese il via la ricerca sul *semantic web*, rappresenta i fondamenti su cui poggia l'idea stessa.

In questa definizione si immagina che computer e utenti possano lavorare in cooperazione.

Per meglio spiegare quest'ultima importante affermazione, si dovrebbe pensare al *semantic web* come ad un insieme di tecniche e di tecnologie che permettono ai computer di “capire” il significato delle risorse informative con cui hanno a che fare nel web. Infatti, il termine “Venezia”, al quale ciascuno di noi attribuisce una serie di significati ben precisi (è una città, è una città italiana, è una città di mare, e così via), non ha alcun significato per un computer se non si trova il modo di spiegarglielo.

Il W3C fondato da Lee nel 1994 al MIT<sup>4</sup> si occupa, tra le altre cose, proprio di definire gli standard e le tecnologie che possono consentire di realizzare questo passaggio.

#### **4.2 COSA SIGNIFICA SEMANTIC WEB?**

Fino a non molto tempo fa, la contestualizzazione e la descrizione delle risorse del

---

2 Si tratta di una rete di soggetti eterogenei, per lo più anonimi e geograficamente distribuiti, che collaborano per realizzare un comune obiettivo, avvalendosi delle tecnologie del *computer networking*.

3 Definita come l'intelligenza che emerge dalla collaborazione, dalla competizione e dallo scambio di informazioni di individui all'interno di un gruppo, per esempio all'interno di un *social network*

4 Massachusetts Institute of Technology

web avveniva utilizzando delle parole chiave che ne richiamavano sommariamente il contenuto. Se avessimo fatto una qualsiasi ricerca di informazioni, per esempio utilizzando *Google*, il risultato che avremmo ottenuto sarebbe stato strettamente legato a queste parole chiave e sarebbe stato, normalmente, una lista di risorse contenenti tra queste parole, quella che noi stavamo cercando.

Ora è chiaro che quando facciamo una ricerca approfondita, normalmente cerchiamo dei contenuti e non un elenco di documenti che richiamano tali contenuti. Oltretutto, se i contenuti che cerchiamo sono distribuiti (come è nella maggior parte dei casi) tra più risorse, è altrettanto chiaro che una volta effettuata una ricerca di questo tipo, siamo poi costretti ad effettuare una serie di ulteriori operazioni di selezione della informazione, a partire dai diversi documenti che abbiamo raccolto, e infine una loro ricomposizione in un nuovo documento: attività che può essere molto dispendiosa e talvolta addirittura impossibile da portare a termine.

Il motivo di questa inefficienza non è tanto riconducibile ad eventuali carenze dei motori di ricerca, quanto piuttosto al fatto che le risorse web sono memorizzate in un formato comprensibile per gli esseri umani (*human-accessible*) ma affatto interpretabile da parte dei computer (*machine-accessible*), e questo implica che noi utilizzatori siamo costretti a lavorare “a mano” anziché avvalerci delle potenzialità computazionali di questi ultimi.

Quando Berners Lee auspica che computer e utenti lavorino in cooperazione, vuole dire proprio questo: fare in modo che i computer siano messi in condizione di elaborare i contenuti del web.

In base a questa idea, possiamo quindi dare una nuova definizione di *semantic web*:

“il *semantic web* è una innovazione in cui, utilizzando opportune tecniche e tecnologie, siamo in grado di attribuire significato alle risorse presenti nel web (documenti, immagini etc..) affinché tale significato possa essere elaborato dalle macchine, oltre che essere compreso dall'uomo.”

Ora, per realizzare questa nuova idea di web, che cosa bisogna fare?

Col *semantic web* le risorse vengono messe in rete sotto forma di concetti e di

relazioni tra concetti in un formato *machine-accessible*. Facendo una similitudine forse un po' azzardata e considerando il *semantic web* come un nuovo linguaggio, anziché mettere in rete una poesia bell'e fatta, mettiamo in rete il vocabolario, la grammatica e le relazioni tra le parole che servono a comporre la nostra poesia. Possiamo effettivamente utilizzare le parole per ricomporre la poesia ma le possiamo utilizzare anche per creare un nuovo componimento.

Quando parliamo di *semantic web*, la nostra grammatica e le regole di composizione, si chiamano “ontologia”. Con l'ontologia definiamo, nell'ambito di un tema di interesse (dominio), i concetti utilizzabili e le regole che legano tali concetti. Questa definizione viene fatta proprio per “istruire” un computer riguardo al contesto in cui i dati e i concetti vengono utilizzati.

Per noi umani infatti è chiaro che se siamo ad un incontro di *blogger* e parliamo di “reti” ci riferiamo ad una interconnessione di persone che utilizzano una infrastruttura di comunicazione e non ad un attrezzo per pescare i pesci; per un computer questa cosa non è altrettanto chiara e gli deve essere spiegata.

A questo punto, una volta introdotti nel web le nostre risorse in formato *machine-readable* e le nostre ontologie di riferimento, per potervi accedere manca ancora un tassello: uno strumento che consenta di effettuare delle interrogazioni su tali dati. Questo strumento è una interfaccia (un *semantic browser*) attraverso cui poter utilizzare un apposito linguaggio (SPARQL<sup>5</sup>) per l'interrogazione di dati rappresentati in RDF<sup>6</sup> (che è la tecnologia utilizzata per descrivere le risorse).

Il *Semantic Web Layer Cake* (si veda la figura a pag. seguente) illustra una gerarchia di tecnologie che vengono utilizzate per realizzare quanto si diceva: il web comprensibile alle macchine. Ciascun livello del *layer cake* è compatibile e utilizza le tecnologie dei livelli su cui poggia. Questa architettura (ideata ancora una volta da Berners Lee) è tuttora in divenire.

---

5 SPARQL Protocol and RDF Query Language

6 Resource Description Framework

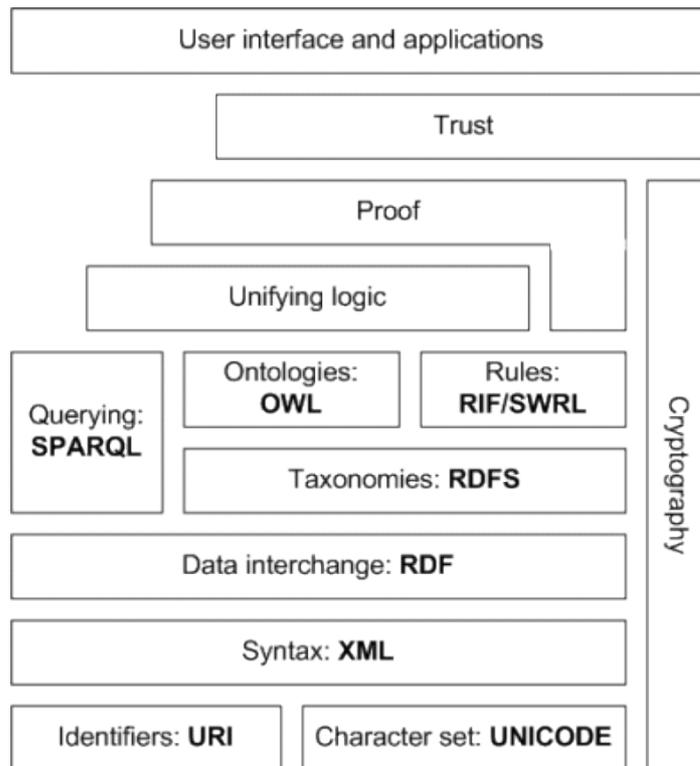


Figura 1: Il Semantic Web Layer Cake illustra l'architettura del semantic web e le tecnologie che la supportano

Il livello di base definisce le tecnologie e gli standard di identificazione e di manipolazione dei testi; il livello XML serve a definire le regole sintattiche e le loro relazioni; RDF definisce le regole per lo scambio di dati tra applicazioni nel web (interoperabilità); il livello delle ontologie, delle tassonomie e delle regole definisce i metodi di organizzazione dei concetti su cui opera SPARQL come linguaggio di interrogazione.

Gli ultimi livelli indicano l'impiego di tecnologie che non sono ancora state standardizzate, ma che possono essere state sviluppate in progetti specifici. Per esempio, all'interno di un *semantic browser* possono essere implementati i successivi livelli del *layer cake* attraverso la verifica dei contenuti e la loro eventuale protezione e crittografia.

### **4.3 ONTOLOGIE E ALTRI STRUMENTI DI ORGANIZZAZIONE DELLA CONOSCENZA**

#### **4.3.1 ONTOLOGIE**

Ontologia è un termine utilizzato in filosofia per riferirsi allo studio della natura dell'esistenza e delle cose. Nell'informatica il significato di ontologia è molto meno profondo e affascinante e quando si parla di ontologie, ci si riferisce piuttosto al tentativo di formulare uno schema concettuale esaustivo e rigoroso nell'ambito di un dato dominio; si tratta generalmente di un repository che contiene tutte le entità rilevanti, le relazioni esistenti fra di esse, le regole determinate da tali relazioni. Una ontologia, definendo questo sistema di relazioni tra entità, fa sì che allorché sia definito un oggetto (nel nostro caso una risorsa web) come Istanza della ontologia stessa, questo assuma tutte le relazioni e le proprietà caratteristiche della classe in cui è collocato, portandosi dietro, per così dire, tutta la conoscenza implicita in tale sistema di relazioni.

Nella *computer science* dunque, una ontologia è riferita ad un dominio e serve a rappresentare e a descrivere una specifica area di conoscenza. Quello dei beni culturali è un dominio, come all'interno di questo lo sono quelli della archeologia e della numismatica, per cui risulta abbastanza chiaro che la ontologia di un dato dominio può anche essere definita attraverso la collezione di più ontologie di sottodomini afferenti al dominio principale.

Da un punto di vista formale l'ontologia può essere rappresentata attraverso un grafo che contiene i concetti relativi al dominio di interesse e le relazioni tra tali concetti.

In prima battuta possiamo semplificare il grafo con un albero in cui i rapporti tra i concetti sono espressi attraverso una struttura gerarchica: classi che rappresentano concetti di livello superiore che via via si suddividono in sottoclassi che rappresentano concetti più fini che godono della proprietà di ereditare le caratteristiche dei concetti di livello superiore, che li hanno generati.

C'è però, oltre a questo rapporto gerarchico tra le classi, un ulteriore livello di

relazione che viene espressa utilizzando il termine di “proprietà”. Le proprietà sono utilizzate per descrivere le caratteristiche (o gli attributi) dei concetti, e possono anche essere utilizzate per collegare tra loro diverse classi, superando il modello strettamente gerarchico di cui in prima battuta ci siamo avvalsi per descrivere l'ontologia. Pertanto le relazioni tra classi non sono soltanto del tipo classe-sottoclasse, ma anche le relazioni tra classi diverse espresse in termini di proprietà.

A questo punto, avendo definito i concetti e le relazioni tra questi, siamo in grado di codificare la conoscenza del dominio di interesse in modo tale che questa possa essere interpretata da un computer.

Le immagini che seguono, derivate dalla ontologia SISC, vengono presentate a titolo di esempio:

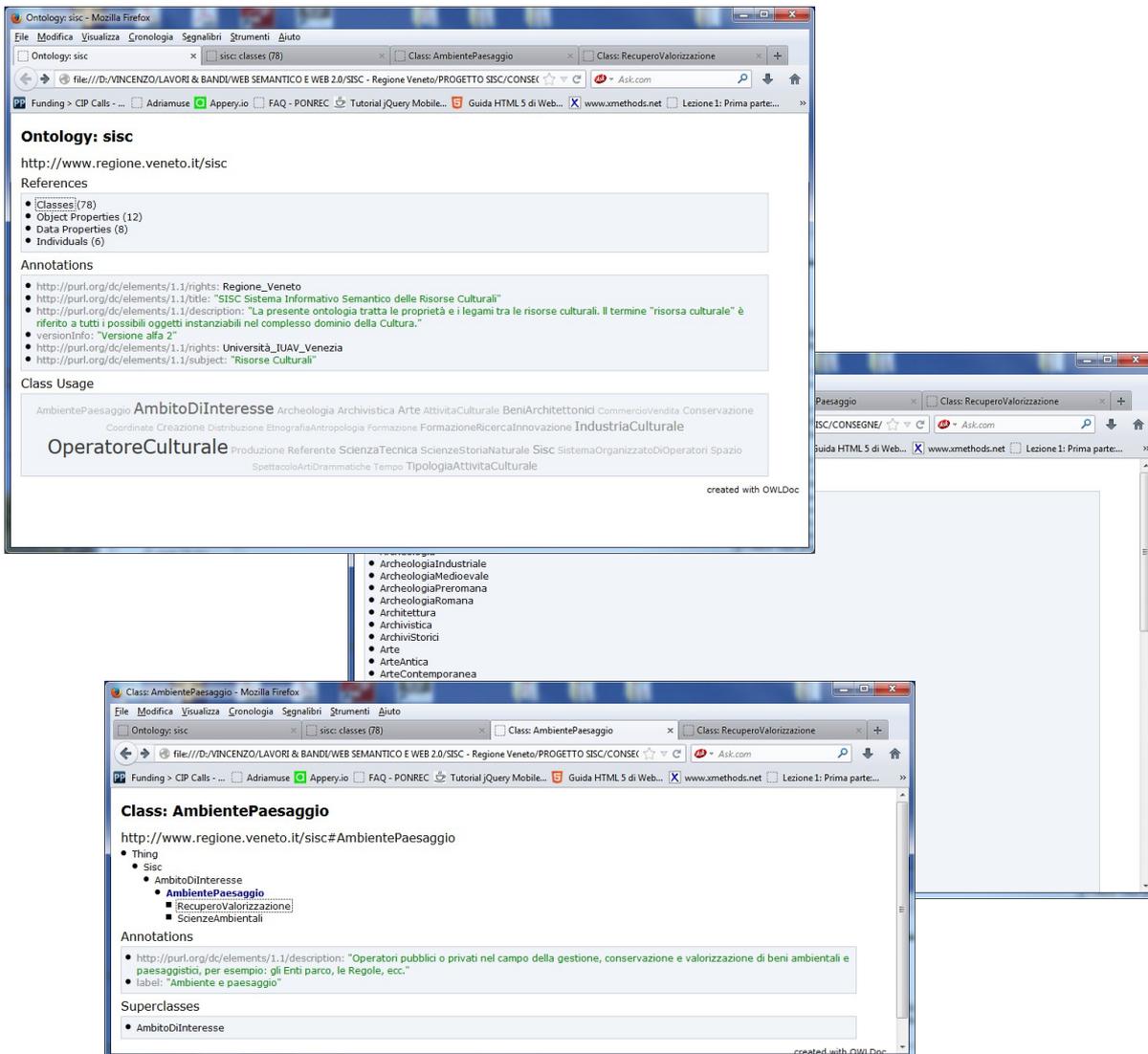


Figura 2: estratto dell'ontologia SISC rappresentata in HTML col plugin "OWLdoc" di Protégé

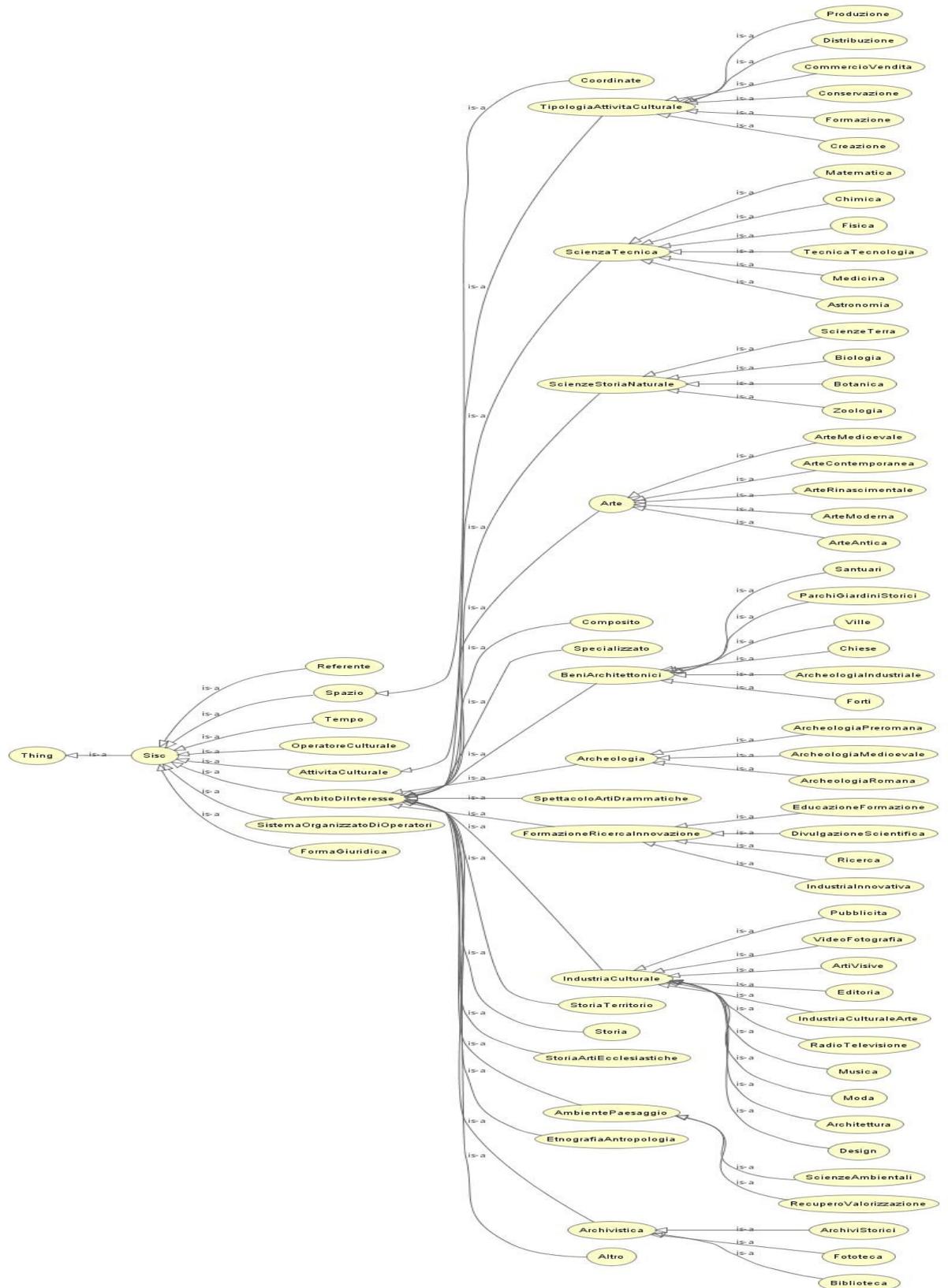


Figura 3: un estratto della ontologia SISC – che riporta solo relazioni gerarchiche – realizzata con l'editor Protégé

### 4.3.2 QUALI VANTAGGI DERIVANO DALL'IMPIEGO DI UNA ONTOLOGIA?

I vantaggi che sono dati dall'uso di una Ontologia sono veramente molti: in primo luogo essa consente di fissare una serie di concetti chiave e di definizioni (relativi ad un dato dominio) che possono essere condivisi, mettendo a disposizione le corrette terminologie per chi deve costruire i propri documenti (condivisione della conoscenza); in secondo luogo consente un riutilizzo completo della conoscenza in essa codificata, anche all'interno di altre ontologie o per il loro completamento (non ridondanza dell'informazione); infine, essendo interpretabile dai computer, abilita il trattamento automatico della conoscenza, con incommensurabili vantaggi di cui solo ora cominciamo ad avere una sia pur vaga comprensione (*semantic web*).

### 4.3.3 ALTRI STRUMENTI DI ORGANIZZAZIONE DELLA CONOSCENZA

Quando parliamo di strumenti e di sistemi di organizzazione della conoscenza (*Knowledge Management Systems* – KOS in inglese), ci riferiamo in particolar modo ad alcuni ben noti schemi come le Tassonomie, i Thesauri o altri tipi di vocabolari. Anche questi, come le ontologie, sono strumenti che consentono di modellare la conoscenza in concetti e in relazioni tra concetti, e infatti, normalmente, sono facilmente riconducibili ad ontologie.

Questi strumenti – spesso *multilanguage* e organizzati in banche dati relazionali o direttamente in RDF - possono essere utilizzati all'interno e in cooperazione con le ontologie, per descriverne compiutamente il significato, per disambiguare termini, per contestualizzare le ricerche effettuate oppure per rendere più efficaci le ricerche stesse associandovi dei concetti affini al concetto immesso dall'utente (si confronti con l'utilizzazione di WordNet descritto più avanti nel cap. 4.3.5).

Secondo *Marcia Lei Zeng* della *Kent State University*, i Sistemi di Organizzazione della Conoscenza possono essere organizzati in quattro classi<sup>7</sup>

1. Liste di termini, come per esempio i dizionari o i glossari;

---

<sup>7</sup> Overview of types of Knowledge Organization Systems - Marcia Lei Zeng

2. Metadati: per esempio gli *authority files* su cui si basa un sistema archivistico, in cui vengono raccolte diverse chiavi di accesso alle risorse dell'archivio;
3. Schemi di categorizzazione e di classificazione, come per esempio le tassonomie;
4. Modelli relazionali: per esempio i thesauri e le ontologie.

Il grafico seguente, denominato “*A Taxonomy of KOS*” e tratto anch'esso da un lavoro della Lei Zeng, illustra in una forma più compiuta, quanto appena affermato:

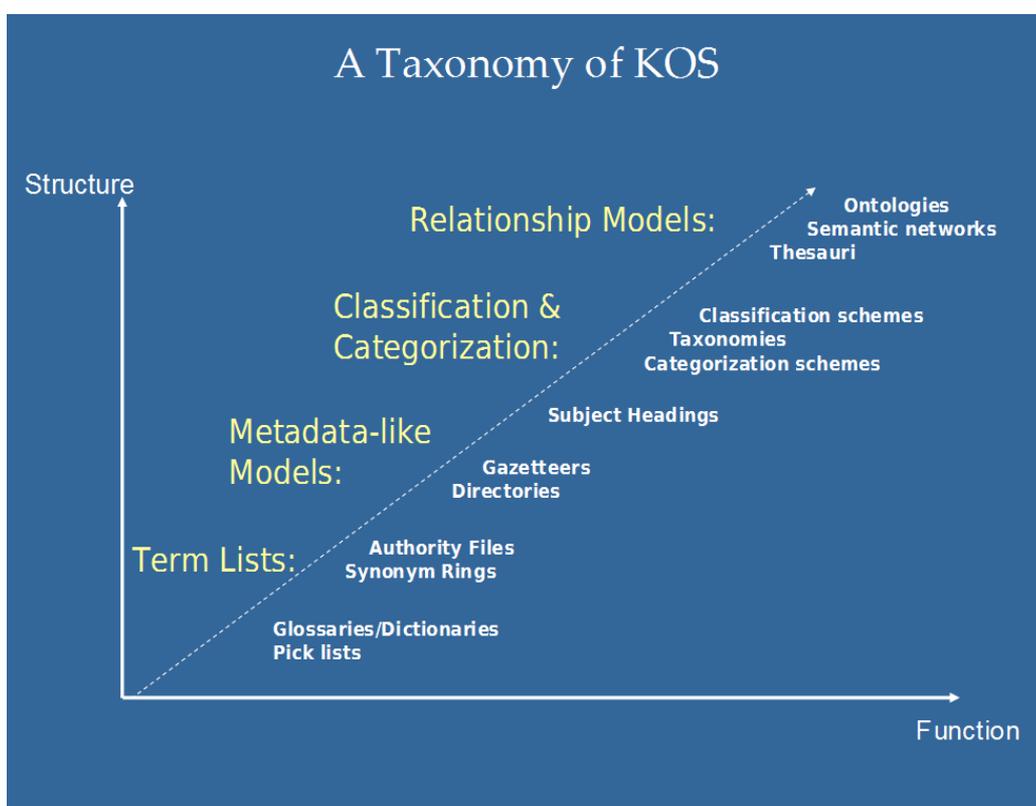


Figura 4: tassonomia di un sistema di organizzazione della conoscenza (Lei Zeng)

Vediamo allora le caratteristiche di alcuni di questi modelli.

#### 4.3.3.1 METADATI

Per poter accedere agevolmente alle risorse di un archivio è auspicabile, se non indispensabile quando l'archivio è molto grande, avere a disposizione uno strumento che ci permetta di descrivere tutte le caratteristiche più importanti di un dato documento. Non solo il titolo e l'autore, ma lo stato di conservazione, le attività di

manutenzione o di restauro effettuate sul documento, il luogo e la data di pubblicazione una descrizione sommaria dei contenuti e così via. Queste informazioni ci consentono di realizzare diverse operazioni: ricerche sui contenuti e sugli autori, statistiche sullo stato di conservazione, selezione e raccolta di dati, pianificazione di attività di manutenzione, gestione della distribuzione e così via.

Le informazioni di cui stiamo parlando sono dette “Metadati” i quali possono essere definiti come “informazioni strutturate relative a risorse su cui debbono essere effettuate delle operazioni”.

Se queste informazioni sono interpretabili da un computer e se questo è in grado esso stesso di effettuare delle operazioni (per esempio di ricerca e di raccolta), stiamo evidentemente parlando di *semantic web*.

*“W3C's Metadata Activity is concerned with ways to model and encode metadata. A particular priority of W3C is to use the Web to document the meaning of the metadata. Our strong interest in metadata has prompted development of the Resource Description Framework (RDF) and its relative Platform for Internet Content Selection (PICS).*

*PICS consists of a suite of specifications which enable people to distribute metadata about the content of digital material in the form of "labels". These contain information about the content in simple, computer-readable form”<sup>8</sup>.*

#### 4.3.3.2 TASSONOMIE

L'etimologia della parola Tassonomia è data dalla composizione delle parole dal greco τάξις (taxis) cioè "ordinamento, disposizione" e da νόμος (nomos) ossia "norma, regola"; significa quindi "ordinamento dei nomi". Tassonomia è dunque: “scienza della classificazione”.

Originariamente con la parola “tassonomia” ci si riferiva alla classificazione degli organismi viventi in biologia. Ora viene normalmente utilizzata per riferirsi alla classificazione di cose e di concetti.

Tale classificazione avviene in forma strettamente gerarchica e la usuale forma di

<sup>8</sup> Tratto dal W3C web site: <http://www.w3.org/Metadata/Activity>

rappresentazione di una tassonomia è quella ad “albero”:



*Figura 5: l'immagine di un albero (vero) è più rilassante*

### 4.3.3.3 THESAURI

Il nome “Thesaurus” venne utilizzato, dal Medioevo in poi, come titolo di repertori scientifici, enciclopedie e dizionari. Esso indica una collezione di termini che non sono definiti ma che hanno tra loro delle relazioni semantiche. Il thesaurus può essere considerato come una estensione della tassonomia in cui i concetti sono ugualmente organizzati in forma gerarchica (*broader terms* e *narrower terms*) ma con la possibilità di definire anche delle relazioni di tipo diverso (*related terms*), basate, appunto, sui significati (per esempio sinonimie o antinomie).

Esempi di thesauri sono il database lessicale WordNet o il GEMET (*General Multilingual Environmental Thesaurus*) di cui si parlerà più oltre nel capitolo dedicato ai “progetti di riferimento per la organizzazione della conoscenza”.

Un esempio interessante da consultare è dato dal “*Visual Thesaurus*” sviluppato dalla società newyorkese *Thinkmap Inc.*

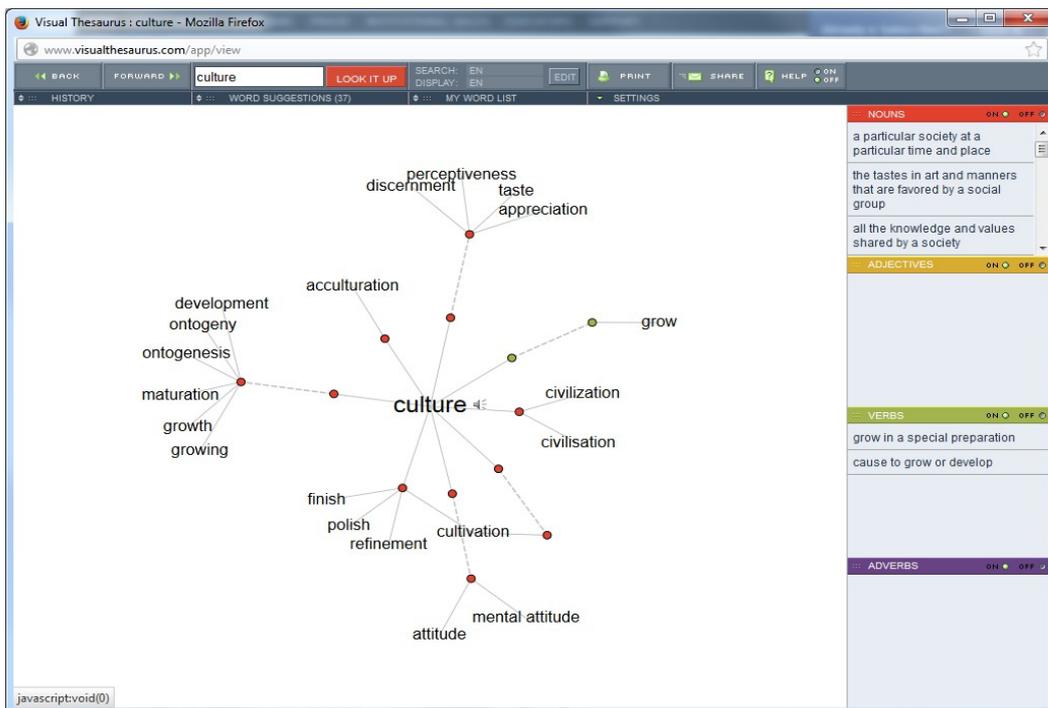


Figura 6: *Visual Thesaurus* è un dizionario interattivo che crea una mappa dei termini sviluppandone le correlazioni

#### 4.3.4 CHE DIFFERENZE CI SONO TRA TASSONOMIE, THESAURI E ONTOLOGIE?

Abbiamo già avuto modo di vedere come la tassonomia possa essere considerata un caso particolare di Thesaurus in cui le relazioni tra i concetti sono esclusivamente di tipo gerarchico.

Se vogliamo dunque rispondere alla domanda del titolo è sufficiente effettuare una comparazione tra Thesaurus e Ontologia.

Innanzitutto tutti i sistemi di organizzazione della conoscenza, inclusi i Thesauri, fanno esattamente questo: organizzano la conoscenza. Non sono però adatti a “rappresentare” la conoscenza e questo semplicemente perché hanno una capacità descrittiva troppo debole. Sebbene il Thesaurus possa arricchire le tipiche relazioni tassonomiche (basate sulla gerarchia) con relazioni ulteriori di tipo semantico-lessicale, tuttavia la sua capacità descrittiva rimane troppo limitata per poter consentire di realizzare le inferenze di tipo logico, possibili con l'ontologia:

*“KOSs are semantically much less rigorous than ontologies, and no formal reasoning can be conducted by just having KOSs. .... ontologies are based upon description logic, therefore logical inferencing can be conducted. However, in KOSs, relationships between concepts are semantically weak. For example, ontologies can specify a is-a relationship, while in thesauri, the hierarchical relation can represent anything from is-a to part-of, depending on the interpretations rooted from the domain and application”<sup>9</sup>.*

Nel senso appena detto, gli usuali sistemi di organizzazione della conoscenza non possono competere con le Ontologie in quanto a capacità di fornire una rappresentazione completa nei termini che sono richiesti dal *semantic web*. Tuttavia sarebbe estremamente utile e opportuno far sì che tutto ciò che si è prodotto e che si produce con sistemi di organizzazione della conoscenza potesse essere in qualche modo ricondotto ad una utilizzazione all'interno del *semantic web*, sia per recuperare un lavoro immenso già fatto, sia per sviluppare nuove circostanze di miglioramento di tali sistemi, anche in termini di condivisione e di accessibilità.

---

<sup>9</sup> Tratto da Liyang Yu - “A developers guide to semantic web” - Springer 2011

E allora la domanda è: come possiamo far sì che ciò accada?

La risposta è data da SKOS (*Simple Knowledge Organization System*): un sistema di linguaggi sviluppati appositamente con lo scopo di rappresentare qualsiasi tipo di vocabolario strutturato, per il *semantic web*.

#### 4.3.5 SKOS – UN PONTE VERSO IL SEMANTIC WEB

Molti sistemi di organizzazione della conoscenza possiedono strutture simili che vengono utilizzate in applicazioni simili. Attraverso SKOS è possibile individuare tali similarità e renderle esplicite ed utilizzabili in forma condivisa tra diverse applicazioni. Infatti SKOS definisce una metodologia standard per la rappresentazione dei sistemi di organizzazione della conoscenza attraverso l'impiego di un vocabolario RDF, garantendo in tal modo sia la *machine readability* dei contenuti, sia l'interoperabilità delle applicazioni che ne fanno uso.

SKOS viene sviluppato da un apposito Gruppo di Lavoro del W3C (*the W3C Semantic Web Development Working Group* (SWD WG)) ed è divenuto una *recommendation* W3C nell'agosto del 2009, da cui l'estratto che segue:

*“The Simple Knowledge Organization System ... aims to provide a bridge between different communities of practice within the library and information sciences involved in the design and application of knowledge organization systems. In addition, SKOS aims to provide a bridge between these communities and the Semantic Web, by transferring existing models of knowledge organization to the Semantic Web technology context, and by providing a low-cost migration path for porting existing knowledge organization systems to RDF”<sup>10</sup>.*

È probabilmente inutile, in questo contesto di tesi, addentrarci nei tecnicismi di SKOS, ma vale forse la pena riportare un ulteriore estratto delle W3C *recommendation* in cui ne vengono citati i costrutti di base:

---

<sup>10</sup> Tratto dalla W3C Recommendation - “SKOS Simple Knowledge Organization System, Reference”

- 18 agosto 2009

*“Using SKOS, concepts can be identified using URIs, labeled with lexical strings in one or more natural languages, assigned notations (lexical codes), documented with various types of note, linked to other concepts and organized into informal hierarchies and association networks, aggregated into concept schemes, grouped into labeled and/or ordered collections, and mapped to concepts in other schemes.”*<sup>11</sup>

Da questo breve estratto risulta abbastanza intuitivo che SKOS offre delle caratteristiche orientate a consentire la rappresentazione della maggior parte dei sistemi di organizzazione della conoscenza, nell'ambito del *semantic web*, infatti, riprendendo altri punti citati nelle *recommendation* e riportandoli alla architettura generale del *semantic web* descritta attraverso il *layer cake* (si veda il cap. 4.2), possiamo affermare che:

- *SKOS concepts can be labeled with any number of lexical (UNICODE) strings, such as "romantic love" or "れんあい", in any given natural language, such as English or Japanese* (riconducibile al livello 1 del *layer cake*)
- *SKOS concepts can be identified using URI* (riconducibile al livello 2 del *layer cake*)
- *SKOS data are expressed as RDF triples, and may be encoded using any concrete RDF syntax (such as RDF/XML or Turtle)* (riconducibile ai livelli 3 e 4 del *layer cake*)
- *The SKOS data model is formally defined in this specification as an OWL full ontology* (riconducibile al livello 5 del *layer cake*)

#### **4.3.6 PROGETTI DI RIFERIMENTO PER LA ORGANIZZAZIONE DELLA CONOSCENZA**

Nell'ambito della organizzazione della conoscenza, sono state sviluppate diverse iniziative che hanno prodotto ontologie e sistemi di metadati e thesauri, in alcuni specifici domini di riferimento. Di seguito vengono riportati alcuni esempi di progetti che riteniamo degni di nota, o perché sono stati in qualche modo utilizzati come

---

<sup>11</sup> Tratto dalla W3C Recommendation - “SKOS Simple Knowledge Organization System, Reference”  
- 18 agosto 2009

fonte di ispirazione per la realizzazione di progetti in ambito Iuav (Sisc, Mace, Adriamuse) o perché sono citati in qualche capitolo di questa tesi o semplicemente perché si ritiene utile che ne sia conosciuta l'esistenza.

### **ISO 21127** “*A reference ontology for the interchange of cultural heritage information*”

Si tratta di uno standard introdotto nel 2006 che definisce un modello concettuale per l'interscambio di informazioni sui beni culturali: uno standard che le istituzioni operanti in questo campo sono incoraggiate ad utilizzare per descrivere le loro collezioni e aumentarne l'accessibilità e la condivisione.

Lo standard ISO 21127 deriva dall'adozione del CIDOC *Conceptual Reference Model* realizzato a partire dal 1996 da un apposito gruppo di lavoro istituito dal *International Council of Museums*<sup>12</sup>.

*“The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.*

*The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.*

*The CIDOC CRM is the culmination of over 10 years work by the CIDOC Documentation Standards Working Group and CIDOC CRM SIG which are working groups of CIDOC. Since 9/12/2006 it is official standard ISO 21127:2006*<sup>13</sup>

---

12 L'International Council of Museums (ICOM), istituito nel 1946, è una organizzazione non governativa, relazionata a Unesco, alla quale partecipano istituzioni museali e professionalità con lo scopo di promuovere e proteggere a livello globale il patrimonio naturale e culturale.

13 Tratta dal sito ICOM <http://www.cidoc-crm.org>

CIDOC CRM è a tutti gli effetti una ontologia di dominio le cui classi principali sono quelle di spazio-tempo, eventi, beni materiali e beni immateriali.

Una parte di questa ontologia è stata utilizzata come riferimento per la compilazione della ontologia del progetto SISC.

**MACE** (*Metadata for Architectural Contents in Europe*) <http://www.mace-project.eu/index.php> è una iniziativa pan-europea di interconnessione e diffusione dell'informazione digitale nel dominio dell'architettura.

*“We connect various repositories of architectural knowledge and enrich their contents with metadata. The result are unique services for searching and browsing architectural contents, for instance, by conceptual connection, geography, language”.*

The screenshot displays four distinct browsing modalities for architectural content:

- Browse by classification:** A hierarchical tree diagram with 'Systems and Equipments' highlighted. Below it, the text reads: "Find contents organized by architectural knowledge."
- MACE Community:** A grid of profile pictures of community members. Below it, the text reads: "Join the MACE Community to share and annotate resources together."
- Browse by competence:** A table showing the number of contents associated with various skills. Below it, the text reads: "Find contents by architectural competence and skill level."
- Browse by location:** A map of Europe with numbered markers (2, 4, 6, 7) indicating project locations. Below it, the text reads: "Use our map to browse architectural projects all over the world."

Competence	Count 1	Count 2	Count 3	Count 4
Knowledge of fine arts	2	2	2	2
Knowledge of History	198	205	205	205
Ability to Create Architectural Design	550	556	565	564
Ability to match requirements and cost factors in	4	4	4	4
Knowledge of Urban Design	156	163	166	163
Underst. people/building relations	537	546	554	557
Underst. of architectural profession	309	313	315	317
Underst. project preparation methods	142	142	144	143

Figura 7: alcune modalità di browsing dei contenuti possibili con MACE

MACE ha visto la partecipazione della Università Iuav di Venezia in qualità di coordinatore lo sviluppo della tassonomia, su cui si basa il sistema di attribuzione dei metadati, realizzata in collaborazione col *Getty Research Institute* di Los Angeles che ha

messo a disposizione il “*repertoire of "Art & Architecture Thesaurus (AAT)"*”.

Il progetto è stato finanziato dal 2006 al 2009 all'interno del programma “eContent-plus Programme, Content Enrichment Project”, coordinato dal *Fraunhofer FIT* di Bonn. Il gruppo di ricerca Iuav è stato diretto dal prof. Vittorio Spigai.

The screenshot shows the MACE website interface. At the top, there is a navigation bar with links for Home, Search & Browse, Community, Feedback & Help, and Extras. Below this, there are tabs for Filtered Search, Browse by Classification, Browse by Competence, Browse by Location (which is selected), and Social Search. The main content area features a map search interface with a Google Maps view of Venice, Italy. The map shows various locations and search results, with two specific locations highlighted with numbered markers (2 and 3). Below the map, there is a search input field with the placeholder text "Enter an address, a place, or a geo-position" and a "Find" button. The results section shows 34 results for the search around the coordinates 45.4335°, 12.3359° (zoom:15). The results are displayed in a grid format, with three visible results: "Die Punta della DoganaThe Punta della...", "Giudecca competition" (with a thumbnail image of a building), and "San Giorgio Maggiore - Venice (Italy)" (with a thumbnail image of a church). Each result includes a "more info" link.

Figura 8: MACE - esempio di ricerca "by location"

**GEMET (General Multilingual Environmental Thesaurus)** è un Thesaurus realizzato dalla *European Environmental Agency* (EEA) utilizzando il *network* denominato EIONET (*European Environment Information and Observation Network*). GEMET nasce come una raccolta di diversi vocabolari multilingue (per esempio comprende *in toto* anche il Thesaurus italiano per l'ambiente), ed è stato progettato come un thesaurus generale, con l'obiettivo di definire una terminologia di base generale per l'ambiente.

Il *merging* dei diversi thesauri è stato realizzato sia dal punto di vista concettuale, sia dal punto di vista formale; concetti coincidenti in più thesauri vengono identificati e marcati; una notazione alfanumerica di tipo neutrale permette di identificare un dato concetto, indipendentemente dalla lingua utilizzata dall'utente. La versione corrente è disponibile in 33 lingue, e contiene oltre 6.000 termini organizzati in uno schema di classificazione composto da 3 super-gruppi, 30 gruppi e 5 gruppi accessori, strumentali alla utilizzazione del thesaurus. GEMET è disponibile in diversi formati: si possono eseguire ricerche *on-line*; è accessibile attraverso *web services*; è scaricabile in diversi formati. Nelle immagini che seguono viene riportato un esempio tratto dal sito del thesaurus, accessibile all'indirizzo <http://www.eionet.europa.eu/gemet>, dove si approfondisce il concetto di “turismo” e all'interno di questo quello di “alpinismo”.

The screenshot shows the EIONET GEMET Thesaurus interface. At the top, there is a navigation bar with links for SERVICES, REPORTNET, TOOLS, and TOPICS (ETCS). Below this, a breadcrumb trail indicates the user's location: "You are here: Eionet > GEMET".

On the left side, there is a "Local navigation" menu with links to User directory, Roles, Organisations, NFP/Eionet IG, Mails to NFPs, SERIS, Workplan/planner, Meetings & events, and Priority dataflows. Below this is a "Find a person" search box and "Account services" including "I have" and "lost my password".

The main content area features a "Thematic Listings" section with links to INSPIRE Spatial Data Themes, Alphabetic Listings, Hierarchical Listings, and Search Thesaurus. Below this is a language selection bar with buttons for various languages, including ar, bg, ca, cs, da, de, el, en, en-US, es, et, eu, fi, fr, ga, hr, hu, it, it, tv, mt, nl, no, pl, pt, ro, ru, sk, sl, sv, tr, uk, and zh-CN.

The "Themes list" section displays a grid of terms in Italian. The term "turismo" is highlighted with a yellow background. Other terms include: acqua, agricoltura, alimenti, acqua potabile, allevamento, ambiente urbano, stress urbano, amministrazione, aree naturali, paesaggio, ecosistemi, aria, aspetti militari, aspetti sociali, popolazione, biologia, chimica, clima, commercio, servizi, dinamica naturale, disastri, incidenti, rischi, economia, edilizia, energia, fisica, forestazione, generale, geografia, industria, informazione, inquinamento, legislazione, materiali, pesca, politica ambientale, radiazioni, ricerca, rifiuti, risorse, rumore, vibrazioni, salute umana, spazio, suolo, trasporti.

At the bottom of the page, there is a footer with links for Download, Administration, Alphabets, About GEMET, Web services, Definition sources, and History of changes. Below these links, it states "GEMET - Themes, version 2.4, 2010-01-13".

A yellow banner at the very bottom of the page contains the text: "Document last modified 2012/07/20 13:27:18.859659 GMT+2. Content in this portal is modified daily by a community of providers. Disclaimer Copyright (c) [redacted]"

The screenshot displays the EIONET GEMET Thesaurus interface. At the top, there is a header with the EIONET logo and 'GEMET Thesaurus'. Below this is a navigation bar with tabs for SERVICES, REPORTNET, TOOLS, and TOPICS (ETCS). The main content area is divided into several sections:

- Local navigation:** A sidebar menu with options like User directory, Roles, Organisations, NFP/Eionet IG, Mails to NFPs, SERIS, Workplan/planner, Meetings & events, and Priority dataflows.
- Concepts list for turismo:** A list of related concepts such as accampamento, acqua dolce idonea a balneazione, and consumo del paesaggio.
- GEMET Thesaurus (alpinismo):** The main entry for 'alpinismo', which includes:
  - Definition:** No definition needed.
  - Broader terms:** attività ricreative.
  - Scope note:** scope note is not available.
  - Groups:** RICREAZIONE, TURISMO.
  - Themes:** turismo.
  - Other relations:** Wikipedia article: [Mountaineering](#).
  - Has close match:** [UMTHES: bergsteigen](#).
  - Multilingual translations:** A list of translations for 'alpinismo' in various languages, including Arabic (تسلق), Basque (mendizaleatasun), Bulgarian (Алпинизъм), Catalan (muntanyisme (UP: alpinisme)), Chinese (登山), Croatian (planinarjenje), Czech (horolezectví), Danish (bjergbestigning), Dutch (bergsport), English (mountaineering), English (US) (mountaineering), Estonian (mägironimine), Finnish (vuorikipelyvarusteet), French (escalade), German (Bergsteigen), Greek (ορειβασία), Hungarian (hegymászás), Irish (síebhteoireacht), Latvian (alpinisms), Lithuanian (alpinizmas), Maltese (alpinizmu), Norwegian (fjellsport), Polish (turystyka górska), Portuguese (alpinismo), Romanian (alpinism), Russian (альпинизм), Slovak (horolezectvo), Slovenian (gornštvo, planinstvo), Spanish (montañismo (UP: alpinismo)), Turkish (dağcılık), and Ukrainian (альпінізм).
- Account services:** A sidebar menu with options like I have lost my password.
- Footer:** A yellow box containing the text 'Document last modified 2012/07/20 13:27:18.859859 GMT+2. Con Disclaimer: Copyright [©] [ ]' and a 'Find a person' search box.

At the bottom of the page, a yellow box displays the Concept URL: <http://www.eionet.europa.eu/gemet/concept/5404>.

Figura 9: il Thesaurus GEMET

## FOAF

*“The Friend of a Friend (FOAF) project is creating a Web of machine-readable pages describing people, the links between them and the things they create and do; it is a contribution to the linked information system known as the Web. FOAF defines an open, decentralized technology for connecting social Web sites, and the people they describe.”*<sup>14</sup>

FOAF è un progetto avviato nel 2000 da *Dan Brickley* e *Libby Miller* che ha prodotto come risultato una delle ontologie più utilizzate in campo *social*.

FOAF nacque come un esperimento rivolto a verificare e dimostrare i vantaggi che si sarebbero ottenuti dalla realizzazione di un *semantic web*. Per questo esperimento fu scelto un dominio di applicazione enormemente vasto e relativamente semplice da trattare: quello delle persone. Esistevano già a quel tempo un grandissimo numero di pagine web personali, dalle quali si sarebbero potute trarre numerose informazioni personali e tra queste anche informazioni su passioni e interessi: sarebbe stato sufficiente coinvolgere i possessori di tali pagine in una comunità che avesse lo scopo di fornire di metadati tali pagine utilizzando una ontologia comune (FOAF appunto) rendendole in tal modo, *machine readable*.

Oggi FOAF può essere considerato come uno standard di fatto per quanti desiderino marcare una pagina web descrivendo informazioni di carattere personale. Le specifiche si possono trovare all'indirizzo <http://xmlns.com/foaf/spec/20100809.html>.

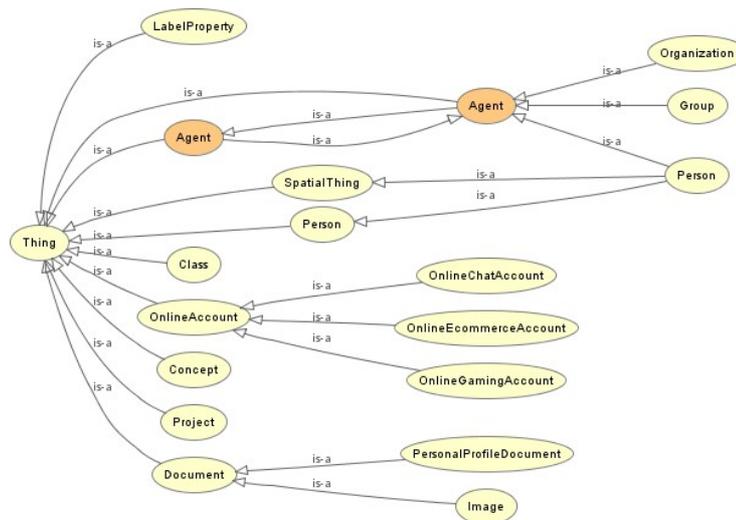


Figura 10: l'ontologia FOAF

<sup>14</sup> Tratto dal sito <http://www.foaf-project.org/>

Il **Dublin Core Metadata Initiative** (DCMI) – si tratta di un progetto avviato nel 1995 durante un *workshop* tenutosi nella città di Dublin in Ohio, che aveva lo scopo di definire un set di elementi di base con cui descrivere risorse digitali. In seguito all'avvio di questa iniziativa, nel 1996 fu definito un primo set di 15 elementi di base – il *Dublin Core Metadata Element Set* – che include, tra gli altri: il Titolo l'Autore, il Soggetto, la Descrizione, un Identificatore, la Fonte e così via..

A partire dal 2000, con l'avvio dei primi studi sul *semantic web* e dello studio di RDF come generico *data model* per la codifica di metadati, Dublin Core divenne uno dei più popolari e utilizzati vocabolari per descrivere documenti RDF:

*“Starting in 2000, the Dublin Core community focused on “application profiles” -- the idea that metadata records would use Dublin Core together with other specialized vocabularies to meet particular implementation requirements. During that time, the World Wide Web Consortium’s work on a generic data model for metadata, the Resource Description Framework (RDF), was maturing. As part of an extended set of DCMI Metadata Terms, Dublin Core became one of most popular vocabularies for use with RDF, more recently in the context of the Linked Data movement.”*

Il sito del DCMI si trova all'indirizzo <http://dublincore.org> e in questo si può trovare qualsiasi ulteriore approfondimento.

The screenshot shows the 'DCMI Specifications' page. At the top, there is a navigation menu with links for Home, Metadata Basics, DCMI Specifications (active), Community and Events, Join/Support, and About Us. Below the menu is a search bar with the text 'Enter keyword' and a 'Search' button. The main content area is divided into several sections:

- DCMI Specifications:** A list of links including Recommendations, Proposed, Working Drafts, Superseded, Recommended Resources, Approval Processes, and Translations.
- Semantic Recommendations:** A section titled 'DCMI Metadata Terms [DCMI Recommendation]' with a sub-section 'DCMI Metadata Terms [DCMI Recommendation]' and a list of bullet points:
  - DCMI Namespace Policy [DCMI Recommendation]
  - DCMI term declarations represented in RDF schema language
  - The Dublin Core Metadata Registry [DCMI Recommended Resource]
  - Dublin Core Metadata Element Set [DCMI Recommendation]
  - ISO Standard 15836:2009
  - ANSI/NISO Standard Z39.85-2012
- User guidelines:** A section titled 'Interoperability Levels for Dublin Core Metadata [DCMI Recommended Resource]' and 'Guidelines for Dublin Core Application Profiles [DCMI Recommended Resource]'. It also includes a section for 'Singapore Framework for Dublin Core Application Profiles [DCMI Recommended Resource]' and 'Using Dublin Core [DCMI Recommended Resource]'.

Figura 11: Dublin Core metadata initiative

**WordNet** è un database semantico-lessicale per la lingua inglese elaborato dal linguista George Armitage Miller presso l'Università di Princeton.

WordNet raggruppa le parole in un set di sinonimi chiamati *synset* dei quali vengono fornite una descrizione generale e le relazioni semantiche tra questi stessi *synset*. L'obiettivo è quello di fornire in combinazione un dizionario e un thesaurus che possano supportare l'automazione dell'analisi di un testo. Date queste sue caratteristiche, WordNet è molto utilizzato nella realizzazione di sistemi di disambiguazione di termini.

WordNet è in grado di distinguere sostantivi, verbi, aggettivi e avverbi e attualmente comprende oltre 150 mila termini organizzati in oltre 115 mila *synset*. La maggior parte dei *synset* sono tra loro interconnessi da relazioni semantiche che variano in base al tipo di parola: per esempio possiamo trovare la relazione di “olonomia” tra due sostantivi (A è olonimo di B se B è parte di A – la testa è olonimo del naso) oppure la relazione di “implicazione” tra due verbi (B è implicazione di A se per fare A bisogna per forza fare anche B – svegliarsi implica dormire).

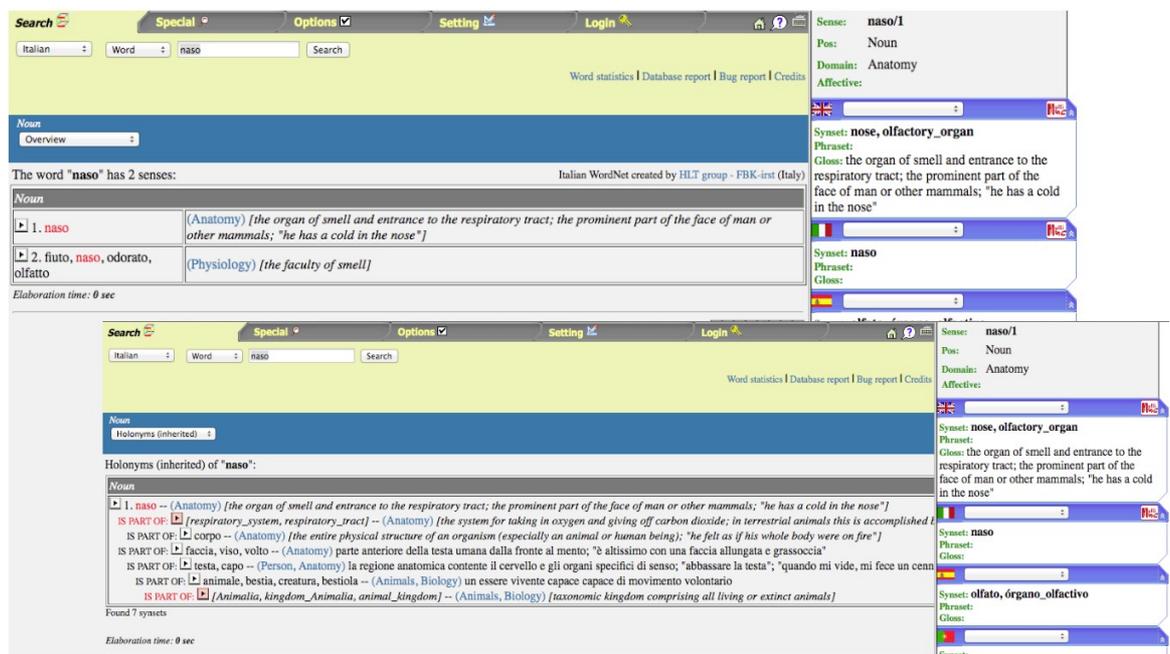


Figura 12: esempio tratto da "MultiWordNet" il WordNet italiano realizzato dal centro ICT della Fondazione Bruno Kessler di Trento

## **Euromuse.net - l'ontologia Harmonize e i servizi Harmosearch**

Euromuse.net è una piattaforma che viene utilizzata come *broker* di informazioni inerenti il dominio dei beni culturali. Attraverso Euromuse.net vengono raccolte informazioni sui Musei e sugli eventi ed esibizioni da essi organizzati, aggregandole da diverse sorgenti e rendendole pubblicamente accessibili attraverso il web. Inoltre vengono rese disponibili all'interno dei circuiti che si occupano di beni culturali e di turismo, affinché possano essere integrate in altre piattaforme come per esempio: portali che si occupano di turismo, agenzie viaggi *online* etc..

Questa opportunità è stata sfruttata anche nel progetto Adriamuse, come viene spiegato nel capitolo in appendice, ad esso dedicato.

Purtroppo il dominio dei beni culturali e del turismo è molto frammentato e ad esso fa riferimento una enorme quantità di informazioni su piattaforme web e su repository ancora molto disomogenei tra loro e ciò costringerebbe ad un grande lavoro di recupero del materiale informativo. Per tentare di superare questo problema, Euromuse.net ha utilizzato la soluzione “Harmonize”.

*“Harmonise enables seamless data exchange by following a semantic mapping approach. Instead of a fixed standard, Harmonise provides a reference model, the Harmonise ontology, which local data formats can be mapped against and in this way mapped to other local data formats or standards. In the sense of such a reference model, the Harmonise ontology defines all relevant concepts and relationships between concepts of the problem domain. A specific mapping mechanism and reconciliation engine then enable the translation of data from one local format to another and enable seamless data exchange and global interoperability.”*<sup>15</sup>

All'interno del progetto Euromuse.net, la *Harmonize Ontology* viene utilizzata per consentire uno scambio di dati continuativo tra differenti *data provider*. Essa è stata adottata per coprire i più importanti sotto domini del turismo (*accomodation, events, attraction, gastronomy*) ed è stata estesa all'interno del progetto Euromuse.net per poter essere utilizzata nel dominio specifico dei musei e delle esibizioni.

---

<sup>15</sup> Tratto dal documento “Harmonization, Interoperability and Standard” del progetto Euromuse.net



Figura 13: homepage di Euromuse.net

La soluzione tecnologica fornita da HarmoSearch - un progetto realizzato in parallelo al progetto Adriamuse - riguarda la fornitura di servizi web per l'interscambio di dati relativi ad eventi, tra diversi utilizzatori registrati.

I servizi web di HarmoSearch consentono tre tipi di operazioni:

*pushData operation*: consente di inviare dati da un utilizzatore ad un altro;

*simpleSearch operation*: consente di richiedere dati specifici ad un utente, effettuando delle operazioni di ricerca semplificate:

*advancedSearch operation*: consente di richiedere dati specifici ad un utente, effettuando delle operazioni di ricerca avanzate.

Tutte queste operazioni richiedono in input almeno tre parametri:

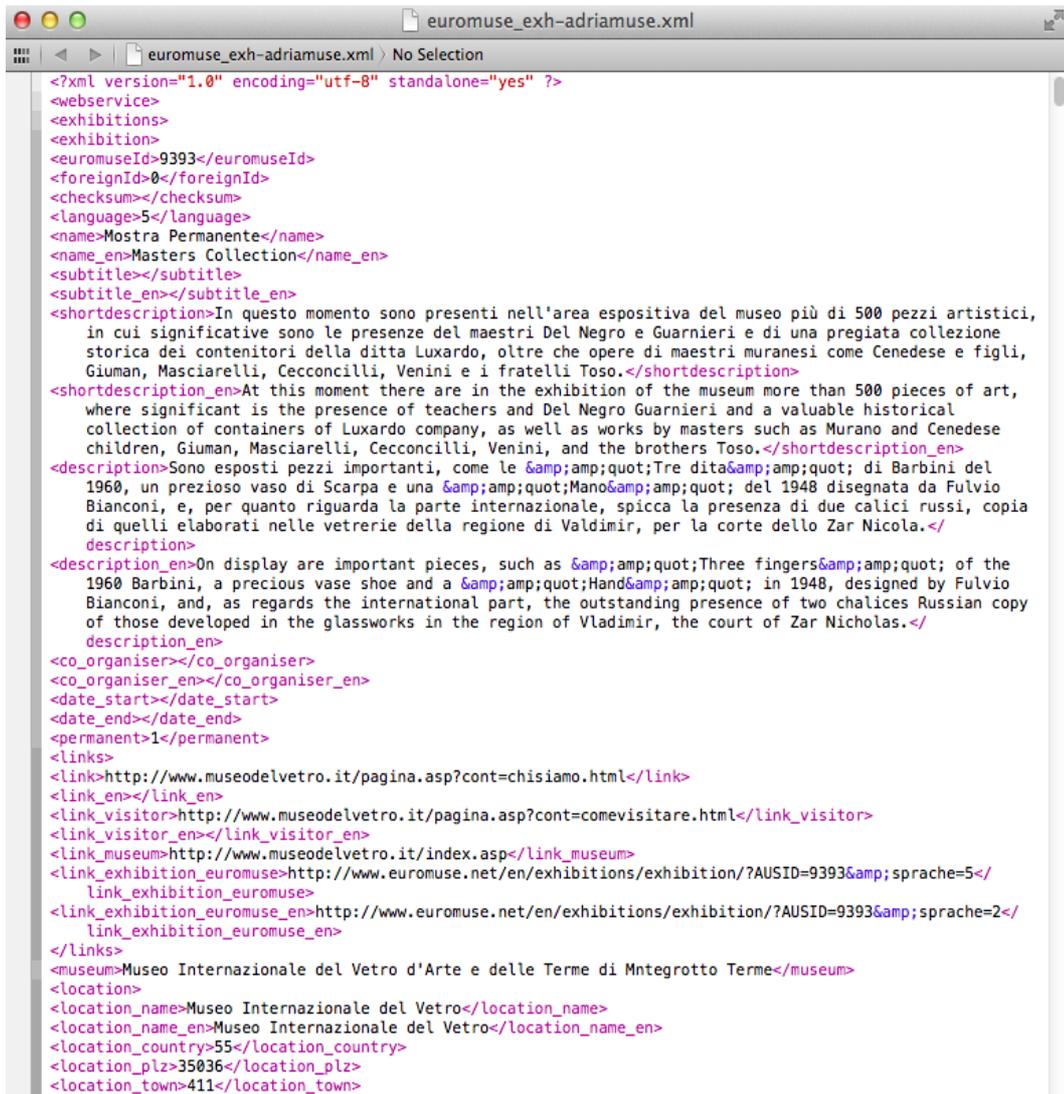
*sender*: il nome di chi invia la richiesta (per esempio un museo);

*receivers*: una lista contenente il nome dei soggetti a cui viene inviata la richiesta;

*data to be sent*: i dati da inviare nel caso di operazione *push*, ovvero la il testo della *query* nel caso di una operazione *search*.

Le informazioni riguardanti ciascun singolo evento, basate sulla ontologia *Harmonize*, includono:

1. *id*: name and identifier of the data provider
2. *eventTitle*: title and sub title of the event
  - *mainTitle* (en + original)
  - *subTitle* (en + original)
3. *description*: description of the event
  - *shortDescription*
  - *longDescription*
4. *category*: list of categories the event belongs to
5. *location*: event location
  - *name*
  - *address*
  - *country*
  - *city*
6. *timeline*: event calendar and timetable
7. *organizer*: who organizes the event
  1. *name*
8. *price*: the price of tickets
9. *documentation*: links to online resources such as photos, web pages ...



```

<?xml version="1.0" encoding="utf-8" standalone="yes" ?>
<webservice>
<exhibitions>
<exhibition>
<euromuseId>9393</euromuseId>
<foreignId>0</foreignId>
<checksum></checksum>
<language>5</language>
<name>Mostra Permanente</name>
<name_en>Masters Collection</name_en>
<subtitle></subtitle>
<subtitle_en></subtitle_en>
<shortdescription>In questo momento sono presenti nell'area espositiva del museo più di 500 pezzi artistici,
in cui significative sono le presenze dei maestri Del Negro e Guarnieri e di una pregiata collezione
storica dei contenitori della ditta Luxardo, oltre che opere di maestri muranesi come Cenedese e figli,
Giuman, Masciarelli, Ceconcelli, Venini e i fratelli Toso.</shortdescription>
<shortdescription_en>At this moment there are in the exhibition of the museum more than 500 pieces of art,
where significant is the presence of teachers and Del Negro Guarnieri and a valuable historical
collection of containers of Luxardo company, as well as works by masters such as Murano and Cenedese
children, Giuman, Masciarelli, Ceconcelli, Venini, and the brothers Toso.</shortdescription_en>
<description>Sono esposti pezzi importanti, come le &quot;Tre dita&quot; di Barbini del
1960, un prezioso vaso di Scarpa e una &quot;Mano&quot; del 1948 disegnata da Fulvio
Bianconi, e, per quanto riguarda la parte internazionale, spicca la presenza di due calici russi, copia
di quelli elaborati nelle vetrerie della regione di Valdimir, per la corte dello Zar Nicola.</
description>
<description_en>On display are important pieces, such as &quot;Three fingers&quot; of the
1960 Barbini, a precious vase shoe and a &quot;Hand&quot; in 1948, designed by Fulvio
Bianconi, and, as regards the international part, the outstanding presence of two chalices Russian copy
of those developed in the glassworks in the region of Vladimir, the court of Zar Nicholas.</
description_en>
<co_organiser></co_organiser>
<co_organiser_en></co_organiser_en>
<date_start></date_start>
<date_end></date_end>
<permanent>1</permanent>
<links>
<link>http://www.museodelvetro.it/pagina.asp?cont=chiamo.html</link>
<link_en></link_en>
<link_visitor>http://www.museodelvetro.it/pagina.asp?cont=comevisitare.html</link_visitor>
<link_visitor_en></link_visitor_en>
<link_museum>http://www.museodelvetro.it/index.asp</link_museum>
<link_exhibition_euromuse>http://www.euromuse.net/en/exhibitions/exhibition/?AUSID=9393&sprache=5</
link_exhibition_euromuse>
<link_exhibition_euromuse_en>http://www.euromuse.net/en/exhibitions/exhibition/?AUSID=9393&sprache=2</
link_exhibition_euromuse_en>
</links>
<museum>Museo Internazionale del Vetro d'Arte e delle Terme di Mntegrotto Terme</museum>
<location>
<location_name>Museo Internazionale del Vetro</location_name>
<location_name_en>Museo Internazionale del Vetro</location_name_en>
<location_country>55</location_country>
<location_plz>35036</location_plz>
<location_town>411</location_town>

```

Figura 14: questa schermata mostra un estratto XML delle informazioni relative ad un evento organizzato da un Museo della rete Adriamuse, compilato secondo lo schema HarmoSearch sulla base della ontologia Harmonize.

## 4.4 TECNOLOGIE E STANDARD

### 4.4.1 RDF, OWL, RAGIONATORI, SPARQL

Il W3C ha definito una serie di raccomandazioni per la realizzazione del *semantic web*, tra cui un modello a grafo orientato ed etichettato per la rappresentazione delle risorse web (RDF), un linguaggio specifico per la creazione di ontologie (OWL), strumenti in grado di elaborare processi di inferenza<sup>16</sup> a partire da un modello ontologico (Ragionatori), uno strumento di query per il recupero delle informazioni (SPARQL).

Questi strumenti vengono brevemente descritti nei successivi capitoli.

#### 4.4.1.1 RDF

RDF (*Resource Description Framework*) fu pensato inizialmente come strumento per la codifica di metadati. La sua prima versione, elaborata dal W3C, risale al 1999: prima che fosse proposta da Berners Lee l'idea di base del *Semantic web* nel 2001.

Se lo scopo iniziale di RDF era dunque quello di descrivere i dati contenuti nel web, dopo il 2001 questo fu notevolmente arricchito, con la prospettiva di essere impiegato non solo per descrivere dati, ma per descrivere qualsiasi risorsa del mondo reale, incluse le relazioni intercorrenti tra tali risorse. Questa nuova idea venne formalizzata nel 2004 attraverso una nuova raccomandazione del W3C, tuttora valida.

Le principali proprietà di RDF sono due:

1. consente di descrivere in maniera flessibile qualsiasi tipo di informazione
2. consente di interconnettere informazioni distribuite nel web

da cui possiamo citare la seguente definizione:

*“RDF is a standard published by W3C, and it can be used to represent distributed*

---

<sup>16</sup> Ragionamento con cui si dimostra il logico conseguire di una verità da un'altra (dal vocabolario della lingua Italiana - Treccani)

*information - knowledge in a way that computer applications can use and process in a scalable manner*<sup>17</sup>

Vediamo dunque come vengono realizzate queste proprietà.

### **Gli statement RDF**

La prima cosa da vedere è come RDF consente di rappresentare in maniera flessibile qualsiasi tipo di informazione, realizzando in tal modo la proprietà n. 1.

L'unità di base che viene utilizzata per ottenere questo risultato è data dal cosiddetto *statement*, ossia una tripla del tipo

Soggetto – Predicato – Oggetto

dove il soggetto è una risorsa, il predicato è una proprietà e l'oggetto è un'altra risorsa o un valore.

Una risorsa può essere qualsiasi cosa: per esempio, questo stesso documento, ovvero la persona che lo sta leggendo, o il suo codice fiscale o la sua foto pubblicata su *facebook*.

Per esempio, se volessi presentarmi come ingegnere e volessi affermare di avere scritto io stesso questa tesi e supponendo che questa sia reperibile ad un dato indirizzo web, potrei scrivere gli *statement*:

“Vincenzo Giannotti è autore di <http://www.iuav.dottorato.tesi/tesivincenzo>”

“Vincenzo Giannotti è Ingegnere”

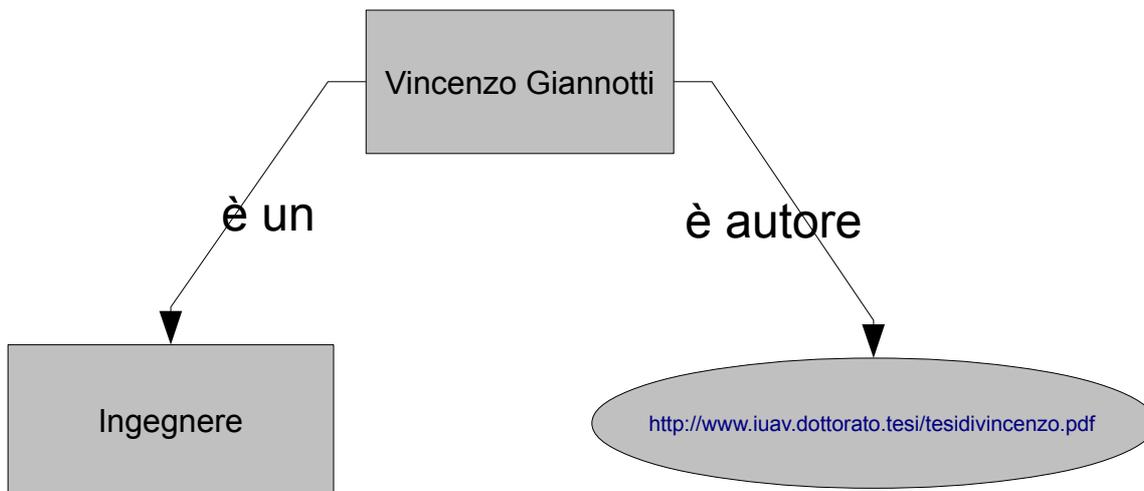
dove:

SOGGETTO	PREDICATO	OGGETTO
Vincenzo Giannotti	è autore di	<a href="http://www.iuav.dottorato.tesi/tesivincenzo">http://www.iuav.dottorato.tesi/tesivincenzo</a>
Vincenzo Giannotti	è	Ingegnere

Ognuno di questi *statement* (o tripla) rappresenta un elemento di informazione mentre un insieme di *statement* (grafo RDF) rappresenta un dato ambito di

<sup>17</sup> Tratto da Liyang Yu - “A developers guide to semantic web” - Springer 2011

conoscenza.



### Gli Uniform Description Identifier in RDF

Tutte le risorse di cui stiamo parlando, possono essere identificate in maniera univoca attraverso degli URI<sup>18</sup> (*Uniform Resource Identifier*).

L'importanza di utilizzare un identificativo univoco per descrivere una risorsa è presto detto: non si può permettere che risorse uguali siano descritte con nomi diversi e viceversa; infatti "Vincenzo Giannotti" o "Vincenzo Bruno Giannotti" che è il mio nome completo, rappresentano comunque la stessa risorsa, la quale a sua volta è diversa da un qualsiasi altro "Vincenzo Giannotti" mio omonimo.

L'importanza di utilizzare un URI come identificativo univoco è ugualmente presto detta: il URI (a differenza del URL) consente di riferirsi a risorse generiche, eventualmente di natura astratta, che possono essere identificate nel web ma che non è detto siano anche recuperabili dal web. Per esempio, nello *statement*: Venezia - è - bella, "Venezia" è una città (di cui esistono varie descrizioni nel web ma che non è direttamente recuperabile dal web in quanto città) e "bella" è un termine astratto che

<sup>18</sup> URI è una stringa che identifica univocamente una generica risorsa che può essere un indirizzo web, come pure un file, un documento, una persona etc.. URI è un concetto più generale rispetto a quello di URL (*Uniform Resource Locator*) in quanto quest'ultimo fornisce un mezzo di accesso a quella data risorsa (p.e. il suo indirizzo http)

non ha alcun corrispondente fisico.

Tralasciamo di approfondire questioni prettamente tecniche legate alla sintassi specifica con cui gli URI vengono definiti<sup>19</sup>; basti sapere che se una risorsa è già definita attraverso un URI, è buona norma che questa venga riutilizzata, quando possibile, evitando di appesantire il web con definizioni ridondanti.

Per quanto riguarda invece la definizione dei predicati, che come abbiamo visto servono a mettere in relazione un soggetto e un oggetto, anche questi vengono identificati univocamente tramite URI e anche in questo caso l'impiego dell'identificativo unico consente di evitare ambiguità nell'utilizzazione di uno stesso predicato in domini di applicazione diversi. Inoltre, poiché nella sostanza il predicato viene definito alla stessa stregua di una risorsa, esso può essere utilizzato come un oggetto, anziché come un predicato, potendolo così caratterizzare con informazioni in grado di descriverne più compiutamente la natura.

Una tale possibilità si dimostra estremamente utile nel momento in cui desideriamo precisare, per esempio, che a un dato predicato può essere attribuita una proprietà di simmetria (se Marco è fratello di Luca, allora Luca è fratello di Marco) consentendo di realizzare i processi di inferenza di cui parleremo poi quando accenneremo ai Ragionatori.

Per concludere questo capitolo dedicato all'utilizzo del URI in RDF, è opportuno far notare un fatto che risulta tuttavia abbastanza intuitivo: URI è anche lo strumento che consente di interconnettere le informazioni distribuite nel web e che realizza, in tal modo, la proprietà n. 2 del RDF. Infatti esso, da un lato consente l'identificazione univoca di una risorsa, da un altro lato favorisce il riuso di risorse esistenti, la qual cosa ha una fondamentale implicazione: una volta che un concetto sia stato definito, non c'è alcun motivo di ridefinirlo e tutto ciò che si potrà dire di aggiuntivo su di esso, sarà semplicemente considerato “ulteriore informazione”, a sua volta riutilizzabile.

Su questo concetto si basa uno dei più importanti progetti che riguardano lo sviluppo del *semantic web* e di cui parleremo più oltre: il *Linking Open Data Project*.

---

<sup>19</sup> hash URI e slash URI

**Le risorse anonime in RDF**

Le risorse anonime (*blank nodes*) sono risorse che non sono individuate da un URI. Tuttavia, pur non essendo identificate da un URI, possono essere utilizzate nelle triple RDF, come soggetto o come oggetto. L'interesse di questo particolare tipo di risorsa risiede nel fatto che essa può essere utilizzata per superare il rapporto di tipo binario che sta alla base della relazione soggetto – predicato – oggetto. Con la risorsa anonima è infatti possibile definire una struttura articolata di questo tipo:

Mario – è – Professore (dove Professore è una risorsa anonima)

Professore – ha nome - Mario

Professore – ha cognome – Rossi

Professore – lavora a - Venezia

In questo caso viene creata una URI composta da più elementi, consentendo di definire una cosiddetta “relazione a più vie”. Si tratta di uno strumento molto potente che tuttavia, come è facilmente comprensibile, consente di utilizzare le nuove URI solo a un livello locale, potendosi queste riferire al solo caso particolare per cui sono state definite.

**RDFS (*Resource Description Framework Schema*)**

Come abbiamo visto, RDF fornisce un meccanismo per identificare le risorse e relazionarle tra di loro ma non consente di dichiarare quali risorse siano lecite, quali proprietà siano applicabili o quali caratteristiche contraddistinguano una proprietà che lega due risorse. Per poter fare ciò è necessario definire classi e proprietà utilizzando “RDF Schema” (RDFS).

RDFS permette di costruire dei vocabolari in grado di definire il significato, le convenzioni e le caratteristiche di risorse e di proprietà. Gli elementi principali che caratterizzano tali vocabolari sono la strutturazione in “classi” e “sottoclassi” delle risorse e la possibilità di definire il dominio e l'intervallo di validità delle proprietà. Una classe può essere definita tramite il costrutto *rdfs:Class*, mentre una risorsa può

essere definita come appartenente ad una determinata classe tramite il costrutto *rdfs:type*. Il processo di ereditarietà tra classi (*rdfs:subClassOf*) in RDFS è esteso anche alle proprietà (*rdfs:subPropertyOf*).

I concetti di dominio (*rdfs:domain*) e di intervallo (*rdfs:range*) sono utilizzati per limitare l'insieme delle risorse che possono avere una data proprietà e l'insieme di valori validi per tale proprietà.

In sostanza RDFS non è altro che un linguaggio che permette di realizzare delle Ontologie, analogamente a quanto si può fare, con maggiore espressività, utilizzando il linguaggio OWL di cui si parla al capitolo successivo.

### **XML/RDF**

RDF come è stato finora descritto, fornisce solamente un *framework* concettuale che serve a descrivere delle risorse in maniera che possano essere elaborate da un computer, ma affinché ciò possa effettivamente avvenire, serve una sintassi che consenta a noi di formalizzare il modello e al computer di leggerlo ed elaborarlo. Il W3C definisce la sintassi XML come adatta a questo scopo. Questa sintassi, chiamata “XML/RDF”, consente di rappresentare un grafo RDF attraverso un documento scritto in XML.

La sintassi XML/RDF non è l'unica adatta a questo scopo, ne esistono anche altre come la “Notation3” (N3) o la “Turtle” (*Terse RDF Triple Language*) che tralascieremo di approfondire in questo contesto, limitandoci a qualche breve cenno sulla prima.

```

- <rdf:RDF>
- <rdf:Description rdf:about="http://www.regione.veneto.it/sisc#operatore_312">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
  <rdf:type rdf:resource="http://www.regione.veneto.it/sisc#OperatoreCulturale"/>
  - <foaf:name>
    Museo Lapidario Greco Latino/Raccolta di mappe e disegni antichi dal sec. XVX/Villa Contarini- Fonda
  </foaf:name>
  <sisc:indirizzo>Villa Contarini Via L. Camerini 1</sisc:indirizzo>
  <sisc:comune>Piazzola sul Brenta</sisc:comune>
  <sisc:provincia>PD</sisc:provincia>
  <sisc:localita>Piazzola sul Brenta</sisc:localita>
  <sisc:haCoordinate rdf:resource="http://www.regione.veneto.it/sisc#coord_312"/>
  <sisc:telefono>049/5590238</sisc:telefono>
  <sisc:fax>049/9619182</sisc:fax>
  <sisc:mbox rdf:resource="contar@tin.it"/>
  <sisc:web rdf:resource="www.villacontarini.com"/>
  <sisc:haFormaGiuridica rdf:resource="http://www.regione.veneto.it/sisc#EntePubblico"/>
  - <sisc:ambito>
    - <rdf:Description>
      <rdf:type rdf:resource="http://www.regione.veneto.it/sisc#Archeologia"/>
      <rdfs:label>Archeologia</rdfs:label>
    </rdf:Description>
  </sisc:ambito>
  - <sisc:ambito>
    - <rdf:Description>
      <rdf:type rdf:resource="http://www.regione.veneto.it/sisc#StoriaTerritorio"/>
      <rdfs:label>Storia del territorio</rdfs:label>
    </rdf:Description>
  </sisc:ambito>
</rdf:Description>
- <rdf:Description rdf:about="http://www.regione.veneto.it/sisc#operatore_311">
  <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Organization"/>
  <rdf:type rdf:resource="http://www.regione.veneto.it/sisc#OperatoreCulturale"/>
  <foaf:name>Museo Civico</foaf:name>
  <sisc:indirizzo>Via Lorenzo Luzzo 23</sisc:indirizzo>
  <sisc:comune>Feltre</sisc:comune>
  <sisc:provincia>BL</sisc:provincia>
  <sisc:localita>Feltre</sisc:localita>
  <sisc:haCoordinate rdf:resource="http://www.regione.veneto.it/sisc#coord_311"/>
  <sisc:telefono>0439/8851</sisc:telefono>

```

Figura 15: estratto di un documento XML/RDF utilizzato per la descrizione delle risorse SISC: si noti l'impiego delle ontologie FOAF e SISC per descrivere le proprietà degli operatori culturali

La sintassi XML/RDF definisce un set di URI alla stregua di un vocabolario in cui ad alcuni prefissi vengono attribuiti dei ben precisi significati. Questo gruppo di URI predefiniti sono chiamati RDF *vocabulary* e tutti condividono il seguente prefisso:

<http://www.w3.org/1999/02/22-rdf-syntax-ns#>

il quale prefisso viene associato convenzionalmente al termine (*namespace*) **rdf:**

Altri esempi di *namespace* da conoscere sono quelli che si riferiscono a RDF Schema e a OWL:

[rdfs: http://www.w3.org/2000/01/rdf-schema#](http://www.w3.org/2000/01/rdf-schema#)

[owl: http://www.w3.org/2002/07/owl#](http://www.w3.org/2002/07/owl#)

Analizzando il listato della “Figura 15” possiamo facilmente riconoscere alcuni termini appartenenti al RDF *vocabulary*, per esempio: “rdf:RDF” e “rdf:Description”, che fanno parte dei *syntax names*, oppure “rdf:type”, che fa parte dei *property names*.

Quindi, ricapitolando brevemente, la notazione “rdf:description” sarà utilizzata per indicare un ben preciso termine contenuto nel RDF *vocabulary* il cui URI sarà dato dalla concatenazione del *namespace* “rdf:” col termine “*description*”

<http://www.w3.org/1999/02/22-rdf-syntax-ns#description>

#### 4.4.1.2 OWL

“OWL is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. This representation of terms and their interrelationships is called an ontology.”<sup>20</sup>

OWL (*Ontology Web Language*) è il linguaggio più avanzato e ricco utilizzato per descrivere ontologie e dal 2004 è una raccomandazione del W3C.

Nel 2007 fu costituito un nuovo W3C OWL *Working Group* per revisionare il linguaggio raccomandato nel 2004 con una nuova versione denominata OWL2.

OWL2 costituisce, dal ottobre 2009, una nuova raccomandazione del W3C. Nel seguito ci si riferirà a OWL2 col termine OWL.

Questo linguaggio viene fornito in tre diverse forme di espressività e complessità crescenti:

- OWL *Lite* è il più semplice dei tre linguaggi e può essere vantaggiosamente utilizzato quando si abbia necessità di definire una gerarchia di classificazione e semplici restrizioni. È facilmente utilizzabile quando si debba realizzare un percorso di migrazione rapido di un thesaurus verso altri sistemi di organizzazione della conoscenza.
- OWL DL (*Description Logic*) è un linguaggio molto più espressivo del *Lite* e comunque garantisce la completezza computazionale (tutte le conclusioni hanno la garanzia di essere calcolabili) e la decidibilità (tutte le computazioni finiscono in un tempo definito). Queste proprietà rendono possibile la computazione automatica della classificazione e la verifica della consistenza della ontologia (*reasoning*).
- OWL *Full* consente la massima espressività ma non è in grado di garantire la

---

<sup>20</sup> Tratto da “OWL Web Ontology Language Overview” - W3C Recommendation 10 February 2004

completezza computazionale. In questo caso non è possibile sviluppare il ragionamento automatico.

OWL può essere considerato come una collezione di termini che possono essere utilizzati per definire classi e proprietà all'interno di uno specifico dominio di conoscenza.

Tutti i termini che vengono utilizzati per questo scopo hanno un unico URI di riferimento:

<http://www.w3.org/2002/07/owl#>

L'elemento basilare di una ontologia compilata in OWL è il “assioma” che è una asserzione del tipo:

B è sottoclasse di A

ed è data per certamente vera.

Ciascun assioma è una asserzione che coinvolge Classi, Proprietà e talvolta Individui; se ritorniamo velocemente a quanto visto a proposito di RDF in cui uno *statement* era dato da una tripla del tipo:

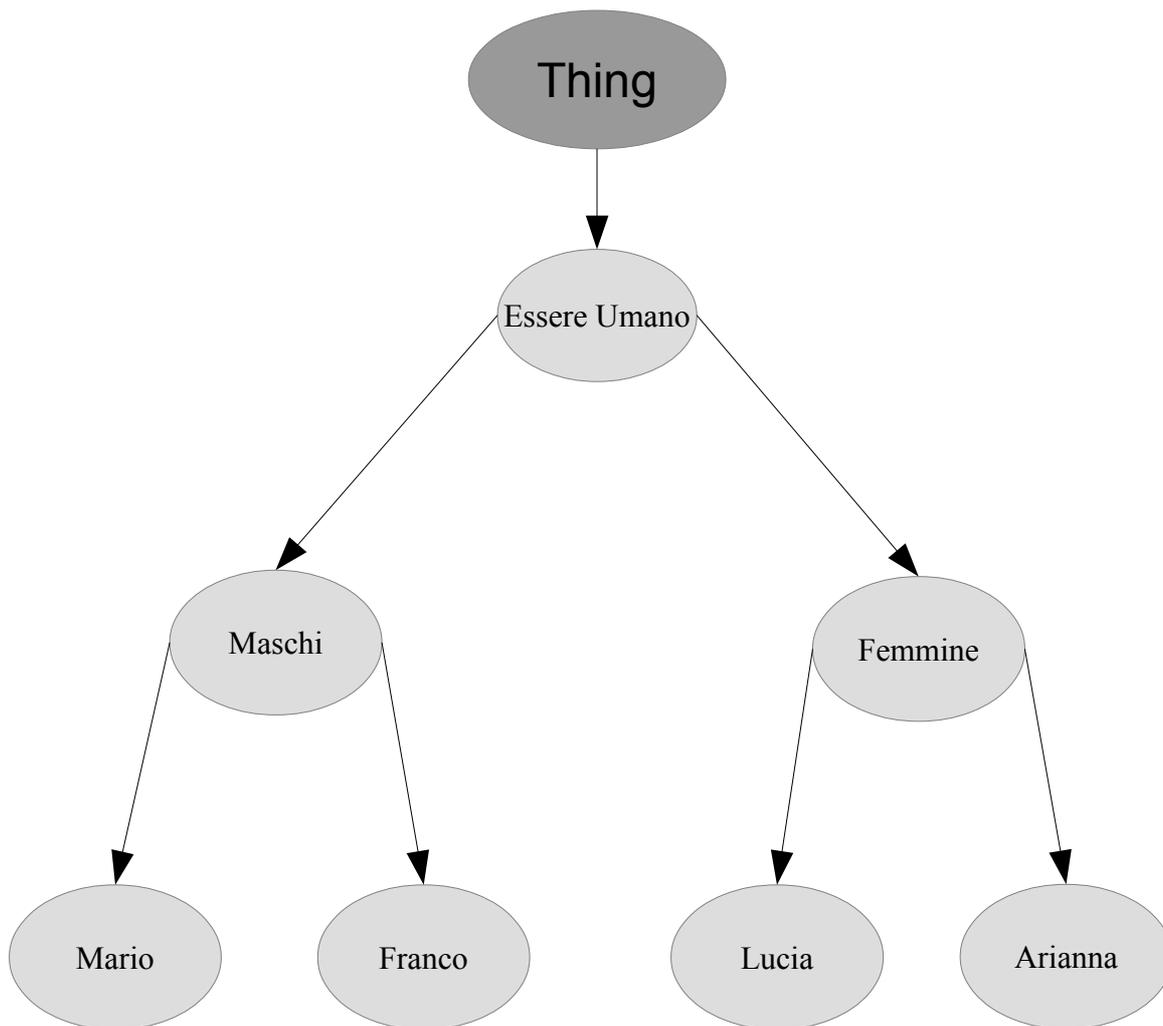
Mario – ha degli – Amici

possiamo facilmente vedere la corrispondenza.

### **Classi**

In OWL le Classi possono essere considerate come dei raggruppamenti di Individui che condividono le medesime proprietà. Le Classi sono organizzate gerarchicamente in una forma del tipo Classe superiore – Classe inferiore, a partire da una Classe genitrice denominata *Thing* che rappresenta il contenitore di tutti gli Individui e la Superclasse di tutte le Classi OWL.

Per esempio, la Classe “Essere Umano”, che deriva dalla Classe *Thing*, può riferirsi alle sottoclassi “maschi” e “femmine” e queste a loro volta contenere degli individui:



L'organizzazione gerarchica in classi e sottoclassi non è nient'altro che una Tassonomia, come abbiamo visto in precedenza, in cui ciascuna Sottoclasse acquisisce le proprietà della Classe da cui proviene:

se Mario appartiene alla classe dei “maschi” allora è un “essere umano”.

Esiste anche una classe specifica chiamata *Nothing* che è una classe priva di Individui ed è una Sottoclasse di tutte le Classi OWL.

Con OWL DL è possibile computare automaticamente le relazioni intercorrenti nella gerarchia Classi-Sottoclassi, utilizzando un “Ragionatore”.

## **Proprietà**

Le Proprietà, in OWL, sono intese come relazioni binarie tra Individui o tra Individui e Valori. Per esempio possiamo dire:

Vincenzo – ha Fratello – Giovanni

oppure

Betty – ha età – 24

Le proprietà (che in altri formalismi possono essere chiamate “relazioni” o “attributi”) possono essere ulteriormente caratterizzate dalle medesime proprietà (si perdoni l'uso ambiguo del termine) che in logica e in matematica si riferiscono alle relazioni tra insiemi: per esempio possono essere simmetriche ovvero riflessive o transitive etc...

Nel precedente esempio si vede subito che se Vincenzo ha come fratello Giovanni, allora è vero che Giovanni ha come fratello Vincenzo (proprietà simmetrica).

Anche le proprietà possono essere organizzate in una gerarchia tale per cui da una proprietà possono derivare più sottoproprietà.

Le proprietà possono avere anche altre caratteristiche: l'appartenenza ad un dominio di riferimento e il *Range*.

L'appartenenza ad un dominio di riferimento implica il fatto che, per esempio, se Vincenzo ha fratello, il fratello faccia parte del dominio degli essere umani e non di quello del mondo vegetale.

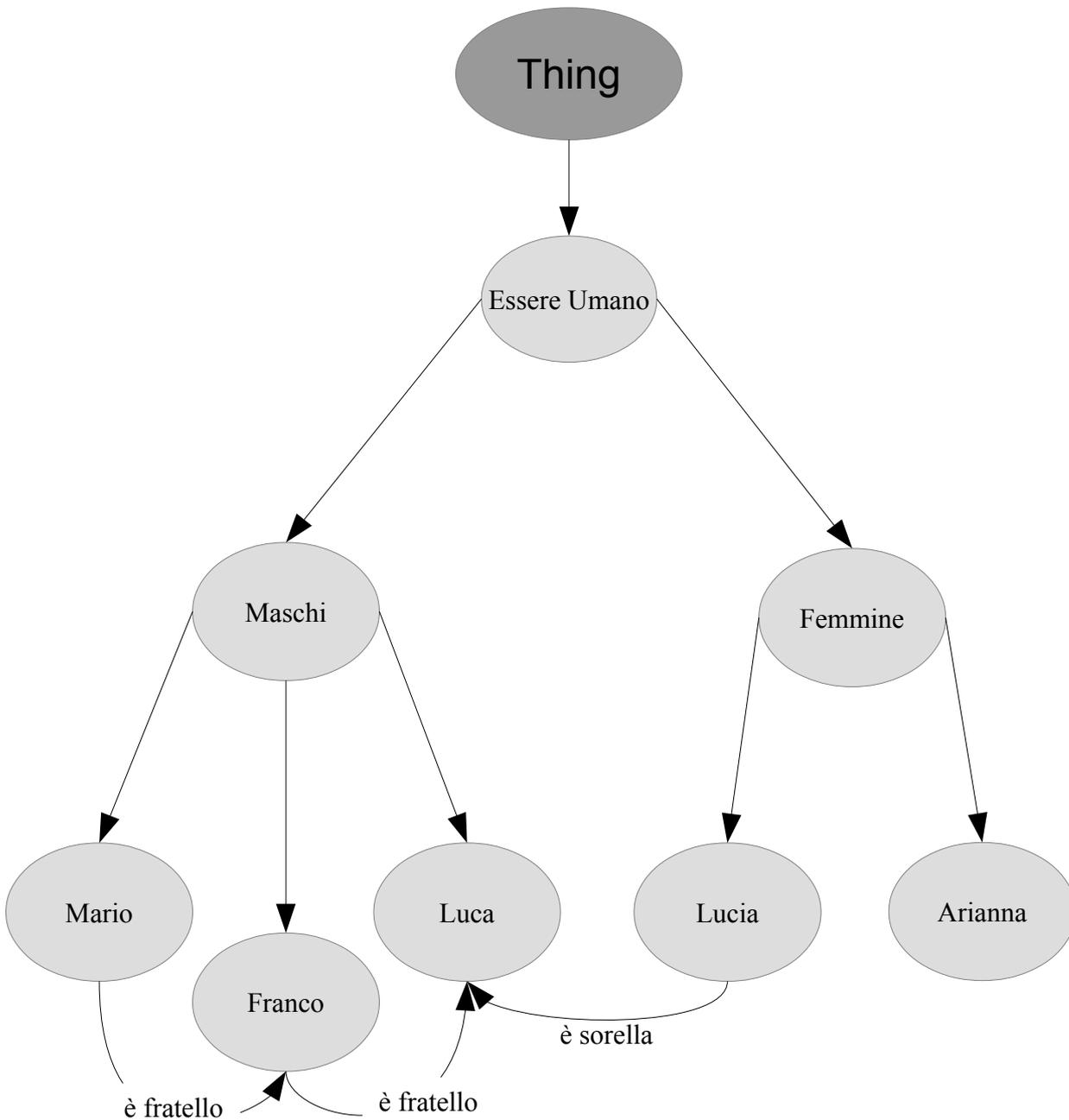
Il *Range* di una proprietà consente di limitare l'insieme dei valori attribuibili a un dato individuo per cui Betty non potrà avere -2745,7 anni.

## **Individui**

Gli Individui, in OWL, sono Istanze (come vengono chiamati in altri formalismi) di Classi che rappresentano oggetti in un dato Dominio di interesse, collegati tra loro tramite delle Proprietà.

Per esempio, a partire dalla classe dei “maschi” possono essere definiti gli Individui

“Mario”, “Franco”, “Luca”, ... e questi possono essere a loro volta relazionati attraverso dei vincoli di parentela tra loro e con gli individui della classe delle “femmine”:



Un'ultima cosa che vale la pena di menzionare quando si parla di OWL (e in particolare di OWL2), riguarda il concetto di IRI (*Internationalized Resource Identifiers*) che non è altro che un URI arricchito dell'intero range dei caratteri UNICODE. A differenza di URI, che utilizza il set di dei 127 caratteri ASCII, IRI

può utilizzare un set di caratteri molto più ricco, in grado di soddisfare le esigenze di Internazionalizzazione (da cui il nome IRI) del *semantic web*.

#### 4.4.1.3 RAGIONATORI E SPARQL

Uno strumento importante e di grande utilità sia per la manutenzione, sia per l'arricchimento di una base di conoscenza, è il ragionatore semantico (*semantic reasoner*).

Il ragionatore semantico, o semplicemente “Ragionatore”, è un motore in grado di validare la coerenza di una base di conoscenza e di ricavare nuove affermazioni RDF da un insieme di regole definite in OWL o attraverso RDF Schema.

Per esempio, riprendendo l'esempio riportato nel grafo della pagina precedente, a partire dalle affermazioni “Franco è fratello di Luca” e “Luca è fratello di Mario”, un ragionatore può facilmente dedurre (inferire) che “Franco è fratello di Mario”:

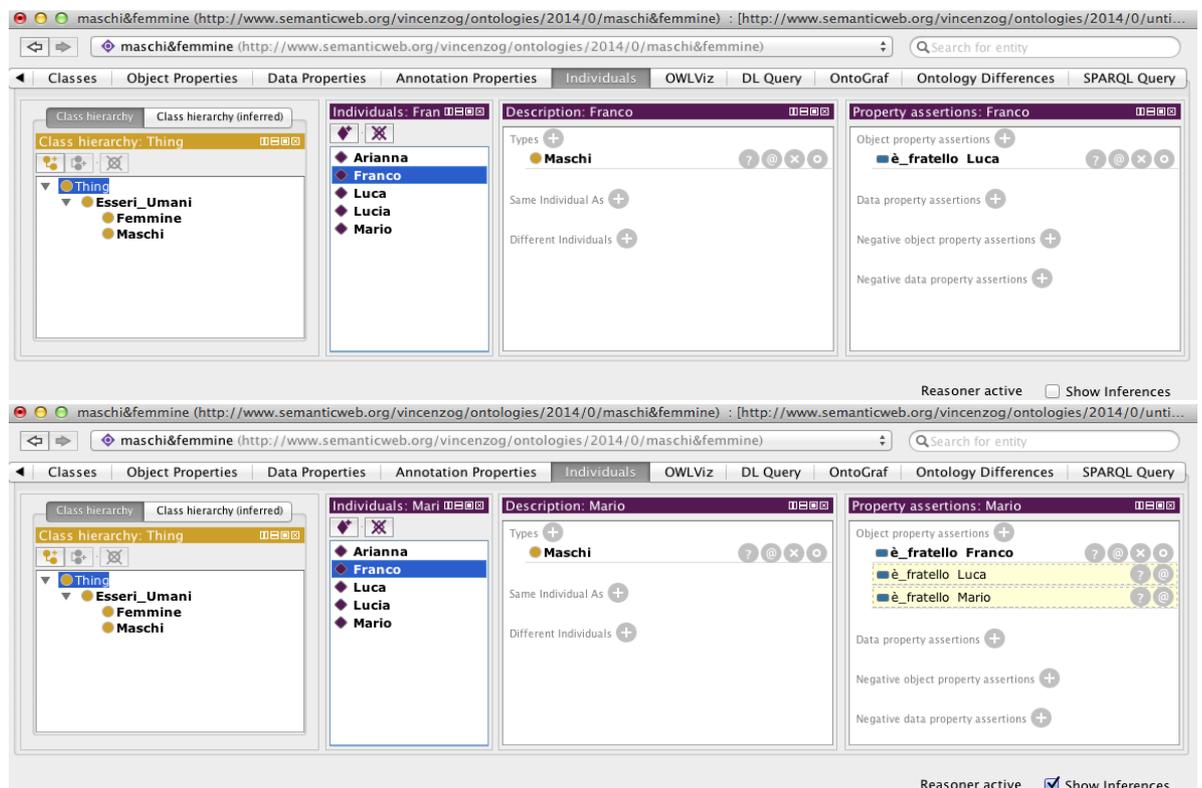


Figura 16: applicazione del ragionatore di Protégé alla semplice ontologia dell'esempio maschi e femmine; nella seconda schermata vengono visualizzate le inferenze

Il ragionatore può essere attivato in due momenti diversi del processo di creazione e di consultazione di una base di conoscenza:

- subito prima di una interrogazione (*query*) e in questo caso generalmente le nuove affermazioni RDF inferite sono memorizzate in modo permanente;
- nel momento in cui vengono aggiunte nuove risorse o apportate modifiche all'ontologia; in questo caso il ragionatore memorizza le nuove affermazioni RDF inferite in modo persistente.

Infine, per poter accedere all'insieme degli *statement* RDF disponibili è necessario disporre di uno strumento che consenta di interrogare tale insieme: questo strumento è il linguaggio di interrogazione **SPARQL** (*SPARQL Protocol and Rdf Query Language*).

SPARQL è il linguaggio di interrogazione standard per il *semantic web* ed è una raccomandazione del W3C. Le *query* vengono richiamate sui dati attualmente disponibili e per questo motivo sarebbe opportuno che il ragionatore fosse attivato subito prima dell'inoltro della *query*: in tal modo le istanze inferite sono memorizzate in modo permanente.

Analogamente a come vengono espresse le affermazioni RDF, le interrogazioni SPARQL si basano su un costrutto, cosiddetto, di *triple pattern matching*.

Un esempio di interrogazione SPARQL realizzata su DBpedia è la seguente, dove chiediamo che siano elencati i film realizzati da Vittorio De Sica e l'anno di produzione di ciascuno:

SPARQL Explorer for <http://dbpedia.org/sparql>

```

SPARQL:
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX do: <http://purl.org/dc/elements/1.1/>
PREFIX <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

prefix dbo: <http://dbpedia.org/ontology/>

SELECT ?subject ?label ?released ?abstract WHERE {
?subject rdf:type <http://dbpedia.org/ontology/Film>.
?subject dbpedia2:starring <http://dbpedia.org/resource/Vittorio_De_Sica>.
?subject rdfs:comment ?abstract.
?subject dbo:releaseDate ?released
FILTER(lang(?abstract) = "it" && lang(?label) = "it").
}
    
```

Results:

SPARQL results:

subject	label	released	abstract
<a href="#">:Il_signor_Max</a>	-	*1939-10-05**xsd:date	"Il signor Max è un film del 1937 diretto da Mario Camerini, aiuto regista Mario Soldati, con musiche di Renzo Rossellini, dirette da Ugo Giacomozzi, direttore fotografia Anchise Brizzi, fonico Vittorio Trentino. Il costumista Gino Carlo Sensani oltre che dei costumi, si è occupato anche dell'arredamento, collaborando logicamente co lo scenografo Gastone Medin. Il film è stato girato negli stabilimenti di Cinecittà."@it
<a href="#">:Cameriera_bella_presenza_offresi...</a>	-	*1951-11-03**xsd:date	"Cameriera bella presenza offresi... è una commedia a episodi prodotta da Carlo Civaliero per la Cines spa, diretta nel 1951 da Giorgio Pàstina, che riuni un cast all-stars intorno ad Elsa Merlini, che faceva il suo rientro sul grande schermo dopo nove anni di assenza."@it
<a href="#">:The_Earrings_of_Madame_de...</a>	-	*1953-09-16**xsd:date	"I gioielli Madame de... è un film del 1953, diretto da Max Ophüls."@it
<a href="#">:Bread,_Love_and_Dreams</a>	-	*1953-12-22**xsd:date	"Pane, amore e fantasia è un film del 1953 diretto da Luigi Comencini. Primo episodio della tetralogia "Pane, amore e...", il film offre un affresco dell'Italia di provincia del dopoguerra. Gli altri sono Pane, amore e gelosia, Pane, amore e... di Dino Risi e Pane, amore e Andalusia di Javier Setó."@it
<a href="#">:The_Earrings_of_Madame_de...</a>	-	*1954-07-19**xsd:date	"I gioielli Madame de... è un film del 1953, diretto da Max Ophüls."@it
<a href="#">:The_Miller's_Beautiful_Wife</a>	-	*1955-10-27**xsd:date	"La bella mugnaia è un film del 1955 diretto da Mario Camerini e girato a Cerreto Sannita, San Lorenzello, Amorosi, Guardia Sanframondi e a Pontassieve."@it
<a href="#">:Pane,_amore_e...</a>	-	*1955-12-22**xsd:date	"Pane, amore e... è un film del 1955 diretto da Dino Risi. È il terzo film della tetralogia Pane, amore e..., preceduto da Pane, amore e fantasia e Pane, amore e gelosia e seguito da Pane, amore e Andalusia."@it
<a href="#">:Fathers_and_Sons_(1957_film)</a>	-	*1957-02-21**xsd:date	"Padri e figli è un film del 1957 diretto da Mario Monicelli."@it
<a href="#">:Pane,_amore_e...</a>	-	*1957-06-19**xsd:date	"Pane, amore e... è un film del 1955 diretto da Dino Risi. È il terzo film della tetralogia Pane, amore e..., preceduto da Pane, amore e fantasia e Pane, amore e gelosia e seguito da Pane, amore e Andalusia."@it
<a href="#">:General_della_Flovere</a>	-	*1959-10-07**xsd:date	"Il generale Della Flovere è un film del 1959 diretto da Roberto Rossellini, realizzato su un soggetto di Indro Montanelli, dalla rielaborazione del quale prese forma l'omonimo romanzo."@it
<a href="#">:It_Started_in_Naples</a>	-	*1960-08-07**xsd:date	"La baia di Napoli (It Started in Naples) è un film diretto da Melville Shavelson."@it
<a href="#">:The_Angel_Wore_Red</a>	-	*1960-09-28**xsd:date	"La sposa bella è un film del 1960 diretto da Nunnally Johnson, tratto da un romanzo di Bruce Marshall, con protagonista Ava Gardner. È stato girato in parte ad Acireale."@it
<a href="#">:The_Last_Judgement_(1961_film)</a>	-	*1961-10-26**xsd:date	"Il giudizio universale è un film del 1961 diretto da Vittorio De Sica. Da un soggetto e una sceneggiatura di Cesare Zavattini un film "corale" caratterizzato dalla presenza di numerosi attori "di grido", voluti dal produttore Dino De Laurentiis, che voleva cercare di replicare il successo de La ciociara con un film assolutamente diverso ma con la stessa accoppiata vincente autore-regista."@it
<a href="#">:The_Biggest_Bundle_of_Them_All</a>	-	*1968-01-17**xsd:date	"Colpo grosso alla napoletana (The Biggest Bundle of Them All) è un film del 1968 diretto da Ken Annakin."@it

Figura 17: esempio di interrogazione SPARQL

questa interrogazione la possiamo leggere così:

Nella prima parte la lista dei *PREFIX* specifica tutti i *namespace* che saranno utilizzati nella interrogazione SPARQL.

L'istruzione *SELECT* definisce quali saranno le variabili di ricerca che saranno prese in considerazione nel risultato (le colonne della tabella SPARQL *results*).

L'istruzione *WHERE* definisce i criteri di selezione specificando, tra parentesi graffe, uno o più *triple patterns* separati dal “.”. Per esempio, il pattern:

`?subject rdf:type <http://dbpedia.org/ontology/Film>.`

Richiede che siano selezionati i soggetti che nella ontologia dbpedia hanno la proprietà di essere dei Film.

Il *pattern* successivo:

`?subject dbpedia2:starring <http://dbpedia.org/resource/Vittorio\_De\_Sica>`

richiede che tra i soggetti precedentemente individuati, siano selezionati quelli in cui compare la risorsa “Vittorio De Sica”.

Il *pattern* successivo:

`?subject rdfs:comment ?abstract`

richiede che a partire dai soggetti precedentemente individuati, siano restituiti gli “abstract” che si riferiscono alla classe “comment” di RDF Schema.

Infine, le istruzioni FILTER servono chiaramente per filtrare le selezioni ottenute rispetto a dei parametri d'ingresso: nel nostro caso, la *label* in lingua italiana e la data del soggetto.

## **4.5 IL NUOVO WEB - NETWORK E PROGETTI DI RIFERIMENTO**

Di seguito vengono illustrati alcuni network e progetti che possono contribuire a definire lo stato dell'arte del *semantic web* anche con riferimento agli ambiti dei sistemi geografici e dei beni culturali.

### **4.5.1 WORLD WIDE WEB CONSORTIUM (W3C)**

Il W3C è una comunità internazionale, guidata da Tim Berners Lee, in cui le Organizzazioni Membro (Imprese, Università, Centri di ricerca) e gli utenti lavorano insieme per sviluppare standard Web.

*“The W3C mission is to lead the World Wide Web to its full potential by developing protocols and guidelines that ensure the long-term growth of the Web.”*

Tra gli standard che il W3C si occupa di sviluppare vi sono anche quelli che si riferiscono al *semantic web* e alla tecnologia XML:

*“In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases.....Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.”*

### **4.5.2 OPEN GEOSPATIAL CONSORTIUM (OGC)**

OGC è un'organizzazione internazionale che si occupa di definire specifiche tecniche per i servizi geospaziali e di localizzazione (*location based*). L'obiettivo dell'ente è quello di sviluppare ed implementare standard per il contenuto, i servizi e l'interscambio di dati geografici che siano aperti ed estensibili.

*“The Open Geospatial Consortium (OGC) is an international industry consortium of 474 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards. OGC® Standards support interoperable solutions that “geo-enable” the Web, wireless and location-based services and mainstream IT. The standards empower technology developers to make complex spatial information*

*and services accessible and useful with all kinds of applications.*”

Le specifiche definite da OGC sono pubbliche e disponibili gratuitamente.

In particolare GML (*Geography Markup Language*) è una grammatica XML utilizzata per descrivere oggetti (*feature*) con caratteristiche geografiche e serve per modellare sistemi geografici e come formato aperto di interscambio di informazioni geografiche via Internet. La specifica del GML è stata definita dal OGC ed è anche standard ISO 19136.

GeoSPARQL (*Geographic Query Language for RDF Data*) è invece uno standard che definisce un set di funzioni SPARQL, un set di RIF *rules*<sup>21</sup>, e un vocabolario RDF/OWL per le informazioni geografiche, basato sul *General Feature Model*.

**GeoSPARQL**

**Vocabulary**

- gs:nearby(lat lon limit)
- gs:nearby(?location limit)
- gs:within(lat1 lon1 lat2 lon2)
- gn:name
- gn:population
- gn:alternateName
- gn:inCountry
- gn:featureCode (just cities or airports)

```

PREFIX co: <http://www.geonames.org/countries/#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX geo: <http://www.w3.org/2003/01/geo/wgs84_pos#>

SELECT ?link ?name ?lat ?lon
WHERE {
  ?link gs:within(51.139725 -0.895386 51.833232 0.645447) .
  ?link gn:name ?name .
  ?link gn:featureCode gn:S.AIRP .
  ?link geo:lat ?lat .
  ?link geo:long ?lon
}

```

[About SPARQL](#) | [geospatial index for jena](#) | presented by [KONA](#)

For further information on the new GeoSPARQL standard see [OGC GeoSPARQL - A Geographic Query Language for RDF Data](#)

**recent queries**

```

PREFIX co:
<http://www.geonames.org/countries/#
>

PREFIX xsd:
<http://www.w3.org/2001/XMLSchema#>
PREFIX geo:
<http://www.w3.org/2003/01/geo/wgs84_pos#>

SELECT ?link ?name ?pop ?lat ?lon
WHERE {
  ?link gs:within(40.157623
-74.855347 41.077281 -73.586426) .
  ?link gn:name ?name .
  ?link gn:population ?pop .
  ?link geo:lat ?lat .
  ?link geo:long ?lon
}

```

Figura 18: esempio di interrogazione GeoSPARQL

Infine sono da citare gli standard che si riferiscono al dominio dei servizi di localizzazione (OGC *Mobile Internet and Location Services*): “OGC standards are implemented in applications, often browser-based, that run on mobile devices such as smartphones, tablet computers or embedded computers (in cars, for example) that have wireless access to the Internet. Most such devices can determine their position (through

21 Il *Rule Interchange Format* (RIF) è una *W3C Recommendation* che ha lo scopo di consentire l'interscambio di regole tra i diversi linguaggi di regole esistenti.

*GPS or other means) and report their position to applications that deliver location services, such as wayfinding. Driven by members' requirements, the OGC works, in concert with other standards organizations, to make indoor and outdoor location information more usable and useful.*<sup>22</sup>

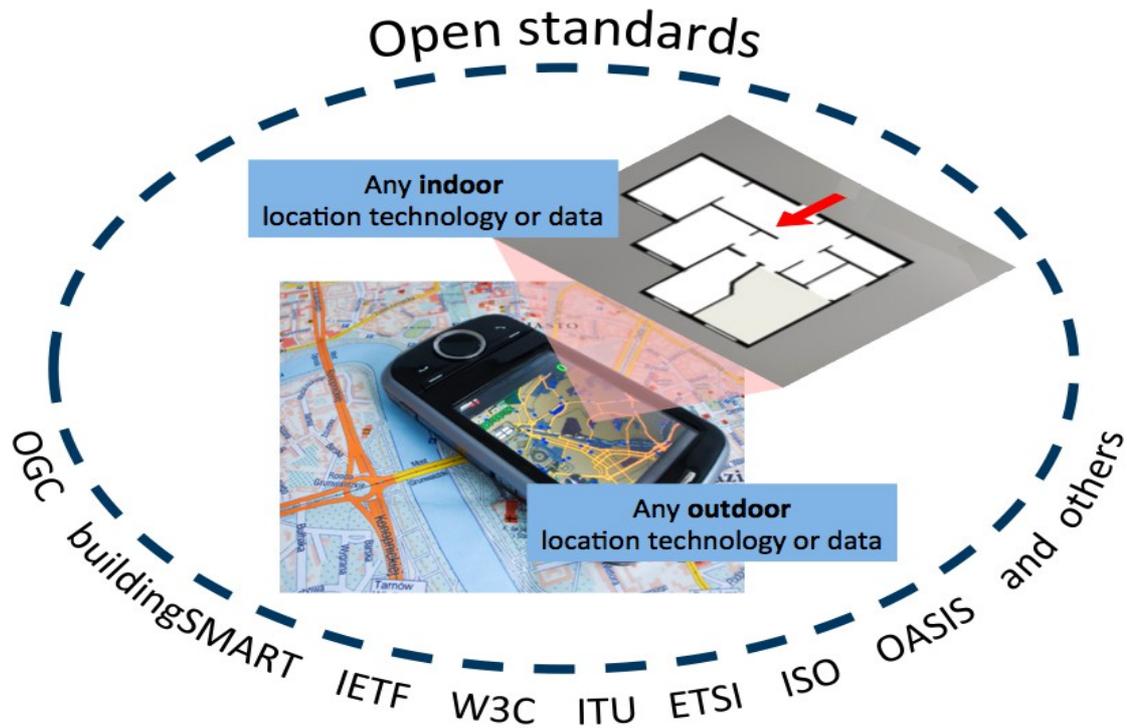


Figura 19: gli standar per la localizzazione indoor e outdoor sono definiti da OGC insieme con altro Organizzazioni analoghe

#### 4.5.3 EUROPEANA

Europeana è una biblioteca digitale europea che allo stato attuale raccoglie, nell'ambito dei beni culturali, oltre 30 milioni di oggetti digitali, usufruendo del contributo di diverse istituzioni europee. La sua dotazione include libri, film, dipinti, giornali, archivi sonori, mappe, manoscritti ed archivi.

Il soggetto che ha in carico il servizio Europeana è la *Europeana Foundation* che si occupa anche della gestione del portale Europeana.eu e della realizzazione dei servizi ad esso correlati.

<sup>22</sup> <http://www.opengeospatial.org/domain/mobileinternet>

Ogni oggetto digitale reso accessibile da Europeana, è ospitato direttamente dalla istituzione che l'ha in carico per cui, una volta che l'oggetto che si sta cercando viene individuato tramite Europeana, per accedere ai contenuti completi ci si può connettere al sito originale.

Per fare sì che le diverse organizzazioni culturali - biblioteche, musei, archivi e collezioni audiovisive - rendano l'informazione ricercabile, è stato studiato lo standard *Europeana Semantic Elements*. Questo standard di metadati, basato su *Dublin Core*, definisce un minimo comune denominatore per l'integrazione di diversi tipi di contenuti digitali.

*“The Europeana Semantic Elements (ESE) format provides a basic set of elements for describing objects in the cultural heritage domain in a way that is usable for Europeana. The current version of the ESE is an updated version of the one that has been used from the start of the Europeana prototype in November 2008. It is a Dublin Core based application profile providing a generic set of terms that can be applied to heterogeneous materials thereby providing a baseline to allow contributors to take advantage of their existing rich descriptions.”*<sup>23</sup>

Tra i vari progetti avviati all'interno di Europeana, di particolare interesse è stato il *Europeana Connect*, un progetto della Commissione Europea che ha sviluppato uno studio approfondito per rendere la libreria di Europeana veramente interoperabile, multilingua e *user oriented*.

Un altro progetto attualmente in corso e avviato nel 2012 è il *Linked Open Data pilot* per la trasformazione di un ampio *subset* di dati di Europeana in *linked data*, rendendoli accessibili attraverso <http://pro.europeana.eu/linked-open-data>.

I dati sono rappresentati secondo il *Europeana Data Model* (EDM) e accessibili attraverso lo SPARQL *endpoint* <http://europeana.ontotext.com>.

---

23 Tratto da “Europeana Semantic Elements Specification and Guidelines” 14/07/2013



### Welcome on the SPARQL end-point of [data.europeana.eu](http://data.europeana.eu)!

[data.europeana.eu](http://data.europeana.eu) currently contains open metadata on 20 million texts, images, videos and sounds gathered by Europeana. The data follows the terms of the [Creative Commons CC0 public domain dedication](#). The data is described using the [Resource Description Framework \(RDF\)](#) format, and structured using the [Europeana Data Model \(EDM\)](#). We provide more detail about the published EDM data at the [technical details page](#).

Please take the time to check out the [list of collections currently included in the pilot](#).

The terms of use and external data sources appearing at [data.europeana.eu](http://data.europeana.eu) are provided on the [Europeana Data sources](#) page.

Sample queries are available on the [sparql](#) page.

## Repository overview

<b>Engine:</b> OWLIM SE	<b>Version:</b> 5.3
<b>Cultural Heritage Objects:</b> 20,163,672	<b>Data version:</b> 2.0
<b>From:</b> <a href="http://data.europeana.eu/download/2.0/">http://data.europeana.eu/download/2.0/</a>	<b>Last update:</b> 14-Sep-2012
<b>Inference ruleset:</b> owl-horst-optimized	<b>Number of expl. statements:</b> 998,471,854
<b>Number of entities:</b> 265,799,020	<b>Number of statements:</b> 3,798,446,742

Figura 20: lo SPARQL Endpoint di Europeana

#### 4.5.4 SCHEMA.ORG

Schema.org è un'iniziativa lanciata il 2 giugno 2011 da Bing, Google e Yahoo! per creare e sostenere un insieme comune di schemi (ontologie intermedie) per la marcatura di dati strutturati sulle pagine, utilizzando i *microdata*<sup>24</sup> di html5.

Schema.org è stato avviato partendo dalla considerazione che

*“Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data.”*

---

<sup>24</sup> *Microdata* è un set di tag, introdotti con HTML5, che permette di fornire una semantica alle pagine web, consentendo la descrizione di contenuti specifici

Una via abbastanza semplice per abilitare questo accesso diretto da parte dei motori di ricerca, consiste nell'effettuare un markup delle pagine sulla base di un vocabolario condiviso:

*“A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.”*

## Thing > Event

An event happening at a certain time and location, such as a concert, lecture, or festival. Ticketing information may be added via the 'offers' property. Repeated events may be structured as separate Event objects.

Property	Expected Type	Description
<b>Properties from Thing</b>		
<a href="#">additionalType</a>	URL	An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax – the 'typeof' attribute – for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.
<a href="#">alternateName</a>	Text	An alias for the item.
<a href="#">description</a>	Text	A short description of the item.
<a href="#">image</a>	URL	URL of an image of the item.
<a href="#">name</a>	Text	The name of the item.
<a href="#">sameAs</a>	URL	URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Freebase page, or official website.
<a href="#">url</a>	URL	URL of the item.
<b>Properties from Event</b>		
<a href="#">attendee</a>	<a href="#">Organization</a> or <a href="#">Person</a>	A person or organization attending the event.
<a href="#">attendees</a>	<a href="#">Organization</a> or <a href="#">Person</a>	A person attending the event (legacy spelling; see singular form, attendee).
<a href="#">doorTime</a>	DateTime	The time admission will commence.
<a href="#">duration</a>	Duration	The duration of the item (movie, audio recording, event, etc.) in <a href="#">ISO 8601 date format</a> .
<a href="#">endDate</a>	Date	The end date and time of the event or item (in <a href="#">ISO 8601 date format</a> ).
<a href="#">eventStatus</a>	<a href="#">EventStatusType</a>	An eventStatus of an event represents its status; particularly useful when an event is cancelled or rescheduled.
<a href="#">location</a>	<a href="#">Place</a> or <a href="#">PostalAddress</a>	The location of the event, organization or action.
<a href="#">offers</a>	<a href="#">Offer</a>	An offer to sell this item—for example, an offer to sell a product, the DVD of a movie, or tickets to an event.
<a href="#">performer</a>	<a href="#">Organization</a> or <a href="#">Person</a>	A performer at the event—for example, a presenter, musician, musical group or actor.
<a href="#">performers</a>	<a href="#">Organization</a> or <a href="#">Person</a>	The main performer or performers of the event—for example, a presenter, musician, or actor (legacy spelling; see singular form, performer).
<a href="#">previousStartDate</a>	Date	Used in conjunction with eventStatus for rescheduled or cancelled events. This property contains the previously scheduled start date. For rescheduled events, the startDate property should be used for the newly scheduled start date. In the (rare) case of an event that has been postponed and rescheduled multiple times, this field may be repeated.
<a href="#">startDate</a>	Date	The start date and time of the event or item (in <a href="#">ISO 8601 date format</a> ).
<a href="#">subEvent</a>	<a href="#">Event</a>	An Event that is part of this event. For example, a conference event includes many presentations, each are a subEvent of the conference.
<a href="#">subEvents</a>	<a href="#">Event</a>	Events that are a part of this event. For example, a conference event includes many presentations, each are subEvents of the conference (legacy spelling; see singular form, subEvent).
<a href="#">superEvent</a>	<a href="#">Event</a>	An event that this event is a part of. For example, a collection of individual music performances might each have a <a href="#">music festival</a> as their <a href="#">superEvent</a> .

Figura 21: questa immagine tratta da <http://schema.org/Event> descrive il concetto di evento.

Questa descrizione è stata utilizzata nel progetto *AdriaMuse* per rimappare il medesimo concetto (definito nella ontologia *Harmonize*) sulla base di *Schema.org*

#### 4.5.5 GEONAMES

GeoNames è un progetto che integra dati di natura geografica e che allo stato attuale rende disponibili oltre 10 milioni di oggetti geografici. L'accessibilità ai dati (che è gratuita con licenza *Creative Commons*) è garantita da *web services* e attraverso una esportazione giornaliera del database.

*“GeoNames is integrating geographical data such as names of places in various languages, elevation, population and others from various sources. All lat/long coordinates are in WGS84 (World Geodetic System 1984). Users may manually edit, correct and add new names using a user friendly wiki interface. GeoNames has Ambassadors in many countries who assist with their help and expertise.”*

Vale la pena di sottolineare il fatto che nell'esempio relativo a GeoSPARQL, citato in precedenza nel capitolo sul OGC, uno dei *dataset* cui viene fatto riferimento è proprio quello di GeoNames.

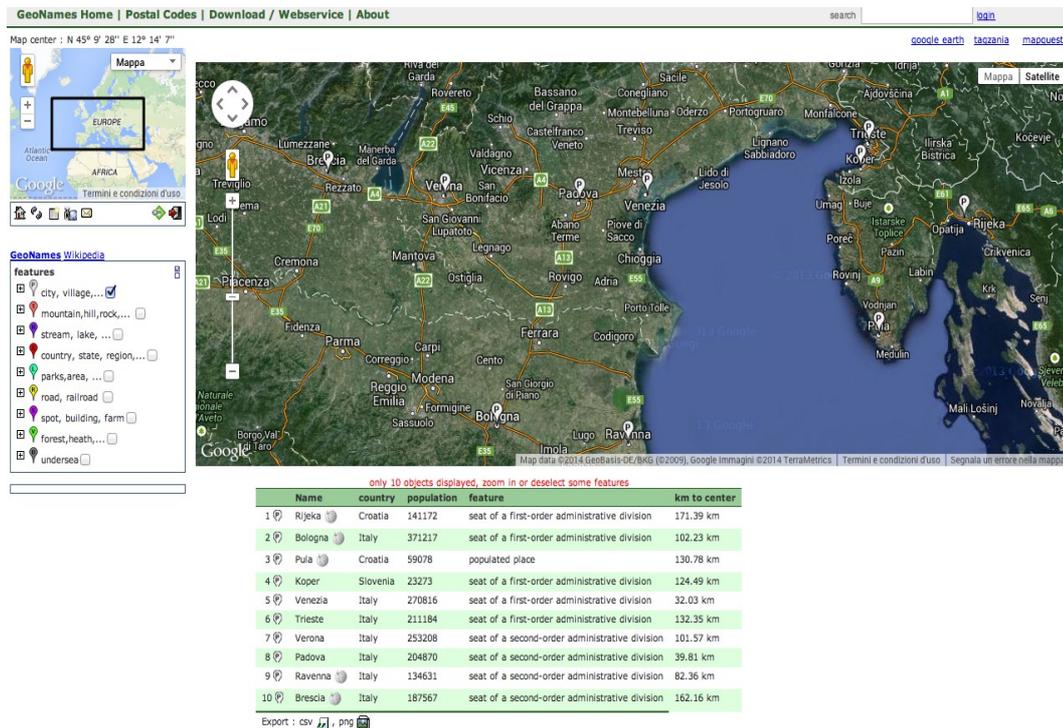


Figura 22: esempio di selezione in GeoNames da <http://www.geonames.org/maps/cities.html>

#### 4.5.6 DBPEDIA E I LINKED DATA

Dbpedia e *Linking Open Data* sono probabilmente i due progetti attualmente più importanti quando si parla di *Semantic web* poiché rappresentano il cuore del già citato *web of data* ovvero di quella che può essere considerata la nuova visione del *semantic web*:

*“The Semantic Web is a Web of data - of dates and titles and part numbers and chemical properties and any other data one might conceive of. RDF provides the foundation for publishing and linking your data. Various technologies allow you to embed data in documents (RDFa, GRDDL<sup>25</sup>) or expose what you have in SQL databases, or make it available as RDF files.”*<sup>26</sup>

##### 4.5.6.1 DBPEDIA

Dbpedia è un progetto avviato nel 2007 dalla Università Libera di Berlino e dalla Università di Leipzig in collaborazione con la società americana OpenLink Software. Il progetto è tuttora attivo ed ha lo scopo di estrarre contenuti strutturati a partire dalle informazioni generate nell'ambito del progetto Wikipedia, rendendo a loro volta disponibili nel web tali informazioni strutturate.

*“DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data.”*<sup>27</sup>

Nella sostanza, le informazioni estratte da Wikipedia sono rappresentate nella forma di grafo RDF e fino ad oggi hanno consentito di realizzare un enorme *dataset* di triple RDF che condividono lo stesso set di ontologie, rendendo di fatto Wikipedia completamente *machine-readable*.

<sup>25</sup> Gleaning Resource Descriptions from Dialects of Languages è una specifica W3C per l'ottenimento di triple RDF da documenti XML

<sup>26</sup> Dal sito del W3C <http://www.w3.org/standards/semanticweb/>

<sup>27</sup> Dal sito Dbpedia <http://wiki.dbpedia.org/About>

Il procedimento di estrazione delle informazioni, realizzato attraverso delle procedure automatizzate, si basa sul fatto che Wikipedia viene realizzata utilizzando dei *template* standard detti *Infoboxes* dai quali è relativamente semplice estrarre l'informazione desiderata visto che possiedono tutti la medesima struttura qualsiasi sia la pagina che analizziamo.

Per garantire la correttezza semantica delle informazioni, ogniquale volta venga estratto un attributo contenuto in un *infobox*, questo viene mappato sulle classi e sulle proprietà definite nella corrispondente ontologia: la *Dbpedia Ontology*.

The figure shows a side-by-side comparison of the Wikipedia article for Tiziano and its corresponding Dbpedia entry. On the left is the Wikipedia page, which includes a navigation sidebar, the article title 'Tiziano', and the main text. On the right is the Dbpedia page, which displays the same article content but with a focus on the structured data (RDF) associated with it. The Dbpedia page includes a table for 'Property' and a large block of RDF code. The RDF code is an XML-based format that maps the information from the Wikipedia article to a structured ontology. It includes various namespaces (e.g., http://www.w3.org/1999/02/22-rdf-syntax-ns#, http://www.w3.org/2000/01/rdf-schema#, http://www.w3.org/2002/07/owl#) and uses properties like rdfs:label, rdfs:comment, and owl:sameAs to describe the entity 'Tiziano' and its relationships to other entities like 'agent', 'Italian painter', and 'Venetian school'.

Figura 23: la pagina di Wikipedia su Tiziano, la corrispondente pagina su Dbpedia e la pagina dei dati Dbpedia

La versione inglese della base di dati di Dbpedia, attualmente descrive circa 4 milioni di oggetti dei quali circa 3,22 milioni sono mappati sulla *Dbpedia Ontology* (versione 3.9). Tra gli oggetti istanziati tramite *Dbpedia Ontology* possiamo trovare 639.000

luoghi, 832.000 persone, 209.000 organizzazioni e molto altro. Inoltre esistono versioni locali di Dbpedia, in 199 lingue, che complessivamente descrivono quasi 25 milioni di oggetti attraverso quasi 2 miliardi di Triple RDF.

Per sviluppare servizi su Dbpedia (in particolare servizi di *browsing* e di ricerca) sono stati sviluppati numerosi progetti satellite. Senza dilungarci oltre, se si desidera farsi un'idea in merito, conviene consultare il sito [www.dbpedia.org](http://www.dbpedia.org) alla voce *Applications*. In questa pagina si potranno trovare numerosi esempi di applicazioni, tra cui alcune interessanti SPARQL *query interfaces*.

#### 4.5.6.2 LINKING OPEN DATA

Si tratta di un progetto di grandissima importanza, avviato nel 2007 e supportato dal W3C. Anch'esso è scaturito, come molte iniziative in questo campo, da una proposta di Tim Berners Lee che di questa nuova idea scrisse in un articolo del 2006, del quale viene riportato un breve estratto:

*“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.*

*Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF. The URIs identify any kind of object or concept. But for HTML or RDF, the same expectations apply to make the web grow:*

- 1. Use URIs as names for things*
- 2. Use HTTP URIs so that people can look up those names.*
- 3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)*
- 4. Include links to other URIs. so that they can discover more things.*

*Simple. In fact, though, a surprising amount of data isn't linked in 2006, because of problems with one or more of the steps. This article discusses solutions to these problems,*

*details of implementation, and factors affecting choices about how you publish your data.”*

Attualmente il *Linking Open Data Community Project* ha lo scopo di pubblicare in RDF il maggior numero possibile di *dataset open* e di definire una serie di *link* RDF tra questi stessi *dataset* così da potersi spostare da un *Item* contenuto in un *dataset*, ad un altro contenuto in un diverso *dataset*.

L'immagine che segue, illustra lo stato dei *dataset* pubblicati nel formato *linked data*, sia dalla comunità *Linking Open Data*, sia da da altri contributori. L'ultimo aggiornamento risale al settembre 2011 e tuttavia questo diagramma è di grande interesse poiché, accedendo al sito che lo propone (<http://lod-cloud.net/>), è possibile cliccare su questa stessa immagine per accedere ad una nuova immagine cliccabile da cui è possibile raggiungere i metadati di un qualsiasi *dataset*.

Da notare il fatto che Dbpedia risulta il *dataset* più corposo e più “collegato” tra quelli proposti nel diagramma.



# SECONDA PARTE

## *I SISTEMI DI RACCOMANDAZIONE*

## 5 I SISTEMI DI RACCOMANDAZIONE

### 5.1 DIFFICILE SCEGLIERE?

I sistemi di raccomandazione sono un insieme di strumenti e tecniche rivolti a realizzare dei servizi per la “guida” di un utente nel *mare magnum* della informazione. In particolare, in questo contesto di tesi, ci si riferisce alla informazione nel web.

Tutti noi siamo consapevoli di quale sia stata l'evoluzione del web negli ultimi anni e a quale impressionante ritmo sia cresciuta la varietà e la quantità dell'informazione che attraverso il web viene resa potenzialmente accessibile. Le espressioni *Internet of things* e *web 2.0*, ormai entrate nel lessico comune, rendono conto del fatto che ormai la tendenza è quella di “mettere in rete” qualsiasi cosa e di essere “costantemente connessi”.

Basti pensare che 15 anni fa non esisteva Google, il motore di ricerca più utilizzato (la società fu registrata nel 1998), Amazon cominciò a svilupparsi nel colosso che conosciamo solo nel 2003, Facebook, il più noto dei *social network*, risale al 2004 e al suo debutto in borsa nel 2012 capitalizzava oltre 100 miliardi di dollari!

Le cifre sono impressionanti:

Year	IP Traffic (PB/month)	Fixed Internet Traffic (PB/month)	Mobile Internet Traffic (PB/month)
1990	0.001	0.001	
1991	0.002	0.002	n/a
1992	0.005	0.004	n/a
1993	0.01	0.01	n/a
1994	0.02	0.02	n/a
1995	0.18	0.17	n/a
1996	1.9	1.8	n/a
1997	5.4	5.0	n/a
1998	12	11	n/a
1999	28	26	n/a
2000	84	75	n/a
2001	197	175	n/a
2002	405	356	n/a
2003	784	681	n/a
2004	1,477	1,267	n/a
2005	2,426	2,055	0.9
2006	3,992	3,339	4
2007	6,430	5,219	15
2008	9,927	7,639	38
2009	14,414	10,676	92
2010	20,197	14,929	256
2011	27,483	20,634	597
2012	-	31,338	885

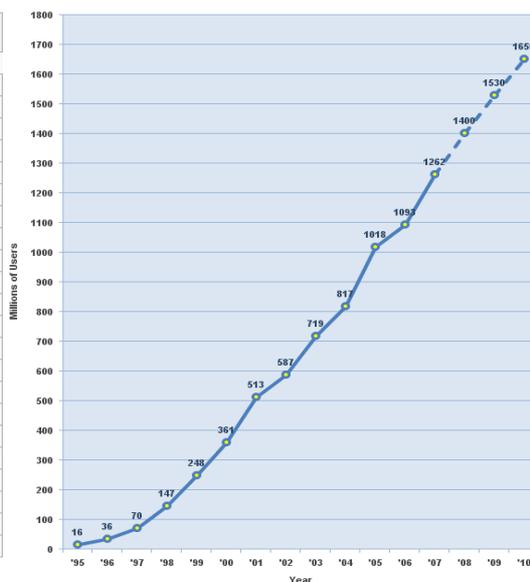


Figura 24: la crescita del traffico Internet per anno (a sx. fonte Cisco Systems) e la crescita degli utenti Internet per anno (fonte Internetworldstats)

Rank	Paese	N. di cellulari	Popolazione	% di cellulari sulla popolazione
-	<a href="#">World</a>	Over 6.8 billion	7,012,000,000	87
01	<a href="#">China</a>	1,150,000,000	1,349,585,838	85.21
02	<a href="#">India</a>	867,800,000	1,220,800,359	70.72
03	<a href="#">United States</a>	327,577,529	310,866,000	103.9
04	<a href="#">Brazil</a>	268,440,423	192,379,287	135.4
05	<a href="#">Russia</a>	256,116,000	142,905,200	155.5
06	<a href="#">Indonesia</a>	236,800,000	237,556,363	109.28
07	<a href="#">Pakistan</a>	125,012,860	178,854,781	69.90
08	<a href="#">Japan</a>	121,246,700	127,628,095	95.1
09	<a href="#">Germany</a>	107,000,000	81,882,342	130.1
10	<a href="#">Philippines</a>	106,987,098	94,013,200	113.8
11	<a href="#">Nigeria</a>	114,000,000	165,200,000	69
12	<a href="#">Bangladesh</a>	106,000,000	148,090,000	68
13	<a href="#">Iran</a>	96,165,000	73,973,000	130
14	<a href="#">Mexico</a>	92,900,000	112,322,757	82.7
15	<a href="#">Italy</a>	88,580,000	60,090,400	147.4
16	<a href="#">United Kingdom</a>	75,750,000	61,612,300	122.9
17	<a href="#">Vietnam</a>	72,300,000	90,549,390	79
18	<a href="#">France</a>	72,180,000	63,573,842	114.2
19	<a href="#">Egypt</a>	92,640,000	82,120,000	112.81
20	<a href="#">Thailand</a>	69,000,000	65,001,021	105

Tabella 1: lista dei paesi nel mondo, ordinata per numero di telefoni cellulari in uso

Questa esplosione incontrollata di connessioni globali e costantemente attive, in cui l'informazione è anche del tipo “cosa sto facendo, dove, ora” può evidentemente far sorgere un problema di capacità di decidere all'interno di una offerta troppo vasta e varia, col rischio di portare a fare delle scelte sbagliate o poco soddisfacenti. Anche se il contesto in cui la scelta viene operata può apparire totalmente libero da qualsivoglia costrizione, paradossalmente libero non è, poiché talvolta induce a delle decisioni disinformate e superficiali, anche a causa dei nostri limiti umani e alla incapacità di valutare troppe alternative.

*“Indeed, choice, with its implications of freedom, autonomy, and self-determination can become excessive, creating a sense that freedom may come to be regarded as a kind of misery-inducing tyranny.”*<sup>28</sup>

Tutto ciò si traduce nella impossibilità di analizzare consapevolmente la quantità e

<sup>28</sup> “The Paradox of Choice” ECCO, New York (2004) – Barry Schwartz, Department of Psychology, Swarthmore College, PA

soprattutto la qualità dell'informazione che ci viene proposta<sup>29</sup>.

I sistemi di raccomandazione sono nati proprio con l'obiettivo di facilitare (talvolta orientare) le scelte dell'utente, cercando di comprenderne i bisogni e di guidarlo lungo dei percorsi in cui scegliere sia più facile e la scelta, alla fine, risulti più soddisfacente.

## 5.2 SISTEMI DI RACCOMANDAZIONE: UNA STORIA RECENTE

*“Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user<sup>30</sup>”*

La nascita dei Sistemi di Raccomandazione, nella loro accezione moderna, si può far risalire ad un esperimento effettuato da Peter Foltz e Susan Dumais dell'Università del Colorado e pubblicata in un loro articolo del 1992. Essi inviarono per alcuni mesi, ad alcuni impiegati della Bellcore, gli *abstract* relativi ai *Technical Memorandum* pubblicati mensilmente dalla Società, cercando di predire quali avrebbero incontrato maggiormente i loro interessi (in base ad interessi dichiarati o dedotti da precedenti letture ritenute rilevanti):

*“This research tested methods for predicting which Technical Memos (TMs) best match people's technical interests. Within Bellcore, approximately 150 new TMs are published each month, yet very few are relevant to any single person's interests. In a six month study, 34 Bellcore employees were sent monthly personalized lists of new TM abstracts that were*

---

29 “I repeat Sturgeon's Revelation, which was wrung out of me after twenty years of wearying defense of science fiction against attacks of people who used the worst examples of the field for ammunition, and whose conclusion was that ninety percent of SF is crud. Using the same standards that categorize 90% of science fiction as trash, crud, or crap, it can be argued that 90% of film, literature, consumer goods, etc. are crap. In other words, the claim (or fact) that 90% of science fiction is crap is ultimately uninformative, because science fiction conforms to the same trends of quality as all other artforms.” (Theodore Sturgeon) - 1958

30 Introduction to Recommender Systems Handbook (Francesco Ricci, Lior Rokach and Bracha Shapira) - 2011

*predicted to best match their interests. These predictions were made using two methods for describing technical interests; one based on sets of keywords that the employees provided, and the other using feedback about previous abstracts they found relevant. Two information retrieval methods were tested to make the predictions, one using standard keyword matching, and the other using Latent Semantic Indexing (LSI). All four methods effectively selected relevant abstracts. The best method for filtering used LSI with feedback about previous relevant abstracts. Feedback using previous relevant abstracts provided an efficient and simple way of modeling people's interests. Overall, the filtering methods show promise for presenting personalized information.*<sup>31</sup>

I sistemi di questo tipo erano chiamati di “*collaborative filtering*” (il primo fu Tapestry, un sistema di filtraggio delle e-mail sviluppato alla Xerox di Palo Alto in California) in quanto adottavano delle tecniche di filtraggio dell'informazione basate sulla somiglianza degli utenti.

L'utilizzo del termine “*Recommender System*”, fu invece proposto da Paul Resnick (AT&T Labs) e Hal Varian (Università della California) in un loro articolo del 1997:

*“The developers of the first recommender system, Tapestry, coined the phrase “collaborative filtering” and several others have adopted it. We prefer the more general term “recommender system” for two reasons. First, recommenders may not explicitly collaborate with recipients, who may be unknown to each other. Second, recommendations may suggest particularly interesting items, in addition to indicating those that should be filtered out.”*

Negli ultimi anni l'interesse per i sistemi di raccomandazione è enormemente cresciuto e oggi giorno il loro studio e il loro impiego viene considerato imprescindibile se si vuole competere nel campo del commercio elettronico e nello sviluppo di servizi web. Giusto per fare un paio di esempi noti a tutti, e che possiamo facilmente verificare, Google e Amazon hanno sviluppato dei sistemi di raccomandazione molto sofisticati e potenti che considerano non solo i nostri interessi resi in forma esplicita, ma una serie molto ampia di strumenti di profilazione

---

31 Personalized Information Delivery: An Analysis of Information Filtering Methods - Peter W. Foltz, Susan T. Dumais *Communications of the ACM*, 35(12), 51-60, 1992.

che possono includere, a puro titolo di esempio, le ricerche che abbiamo effettuato in precedenza, i commenti e le *reviews* che scriviamo e leggiamo, i gusti e gli interessi dei nostri amici, il luogo in cui ci troviamo e così via, con lo scopo di fornire servizi sempre più personalizzati ed efficaci.

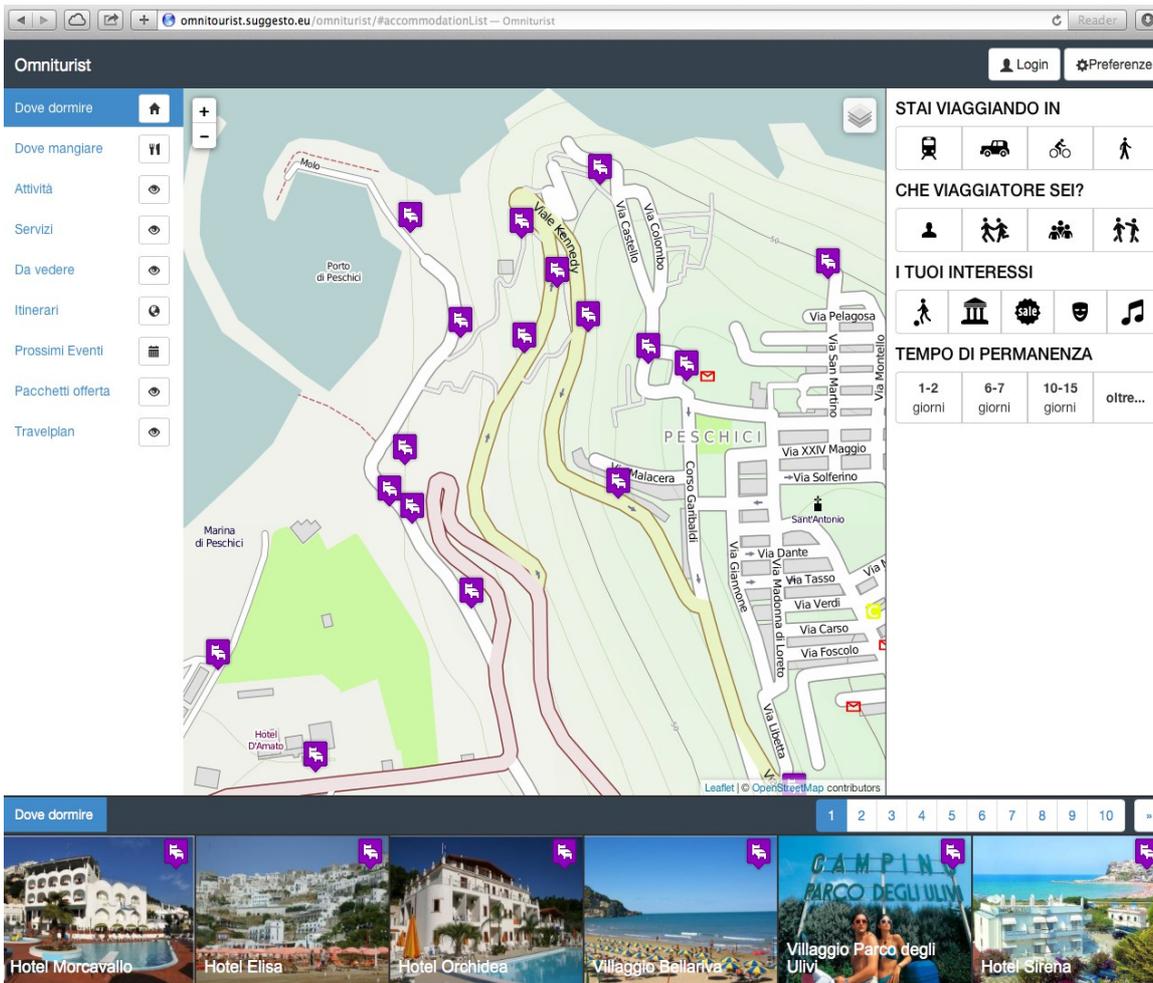


Figura 25: nel sistema OmniTurist l'utente ha a disposizione una apposita sezione per descriversi ed esprimere i propri interessi

### 5.3 A CHI SERVE LA “RACCOMANDAZIONE”?

Prima di passare a trattare gli aspetti più tecnici che contraddistinguono i sistemi di raccomandazione, proviamo ad analizzare brevemente quale sia lo scopo per cui vengono realizzati e l'utilità e i vantaggi che possono portare sia per chi provvede la raccomandazione (*Provider*), sia per chi ne è oggetto (*User*). Una analisi di questo tipo, che qui viene solo accennata in riferimento ad alcuni casi tipo, deve essere senz'altro condotta quando si debba decidere se e come realizzare un sistema di raccomandazione; infatti uno dei requisiti fondamentali di questo tipo di sistemi è che debbono soddisfare sia le aspettative del *Provider*, sia quelle dello *User* e questo implica che debbono essere ben chiari i bisogni di entrambi.

#### 5.3.1 LE ASPETTATIVE DEL *PROVIDER*

L'impiego di tecniche di raccomandazione, dal punto di vista del *Provider*, hanno degli obiettivi facilmente intuibili e possono fornire diversi vantaggi, in ogni caso legati al contesto in cui avviene la raccomandazione. Se si tratta di promuovere la vendita di beni *online* da parte di una multinazionale ovvero di fornire servizi turistici e culturali da parte di una pubblica amministrazione, la percezione di quale sia “vantaggio” che il *Provider* ottiene dal processo di raccomandazione può essere sensibilmente diverso; tuttavia le differenze non è detto siano così marcate da dover trattare i vari casi separatamente, per cui possiamo tentare di dire quanto segue:

in primo luogo un vantaggio può essere che “si vendono più beni o servizi”. Questo può sembrare un beneficio soprattutto per una società commerciale (come in effetti è) che utilizzando le giuste tecniche riesce a vendere di più che se non le utilizzasse; ma lo stesso ragionamento può valere per un museo statale, che attrae più visitatori e riesce a proporre un *merchandising* più appropriato. Nel caso delle vendite *online* l'obiettivo è comunque sempre quello di far sì che il visitatore web permanga il più possibile all'interno sito e che sia raggiunto a una offerta che abbia le maggiori possibilità di successo.

Un secondo importante aspetto riguarda la “varietà dell'offerta” che può essere

vantaggiosamente accresciuta se si utilizza un sistema di raccomandazione; chi propone la vendita del bene può avere interesse a bilanciare le proposte su un'ampia gamma di prodotti, così come l'amministrazione di una città turistica potrebbe avere interesse ad indirizzare i visitatori, non solo lungo i classici percorsi turistici, ma anche in aree più periferiche, con lo scopo di promuoverne la valorizzazione.

Un terzo aspetto può essere rivolto poi ad “aumentare la soddisfazione dell'utente” con lo scopo di fidelizzarlo e di far sì che esso stesso divenga promotore della bontà di quanto offerto. Infatti, l'offerta di prodotti in linea con le aspettative (anche inconsapevoli) dello *User*, darà a quest'ultimo la sensazione che “proprio qui si trova quello che cerco”, facendogli apprezzare, non solo il prodotto che viene raccomandato, ma anche il servizio nel suo complesso.

Infine, il fatto che più *User* siano attratti da una data offerta e che questi stessi permangano per un tempo maggiore all'interno del sito che la propone, dà la possibilità al *Provider* di analizzarne a fondo i comportamenti e di capire meglio quali siano le loro aspettative. Su questo specifico argomento, che riguarda la profilazione dell'utente e la sua fidelizzazione, torneremo in seguito.

### 5.3.2 LE ASPETTATIVE DELLO *USER*

Dal punto di vista dello *User*, dando per scontato che il processo di raccomandazione rispetti i giusti canoni di correttezza (la raccomandazione deve avvenire sempre nell'interesse dello *User*), valgono considerazioni analoghe a quelle fatte per il *Provider*.

Un primo evidente vantaggio che lo *User* può ottenere dal fatto di poter usufruire dei servizi di un sistema di raccomandazione, riguarda il fatto che gli potranno essere suggeriti dei “prodotti e servizi in linea con le sue personali attese”, in una forma ordinata a partire dal più appetibile.

Un'altra interessante possibilità per lo *User* è quella di ottenere anche “suggerimenti su prodotti alternativi o complementari” rispetto a quelli cercati, che possono in qualche modo arricchire l'offerta: non sono prodotti che avrebbe scelto (principalmente per ignoranza), ma il sistema di raccomandazione ritiene che gli

possano essere utili.

Nel caso di applicazioni in ambito turistico l'offerta può anche essere del tipo “a pacchetto” in cui allo *User* vengono offerte delle soluzioni di visita organizzate in base ai suoi interessi, alle sue possibilità economiche, al fatto che sia in vacanza da solo o con la famiglia e così via. In questo caso l'offerta non è un prodotto, ma una serie di prodotti: non tutti presi singolarmente sono i migliori, ma complessivamente rappresentano un'offerta attraente per il turista a cui sono rivolti.

Un'altro importante aspetto da tenere in considerazione è dato dal fatto che nel web si possono trovare un sacco di persone volenterose, che non aspettano altro che di fornire i loro personali giudizi e opinioni su questo e quest'altro prodotto. L'aspetto *social* può diventare in taluni casi un formidabile supporto per un sistema di raccomandazione e un fattore di grande attrattività per lo *User*.

#### **5.4 QUALI CARATTERISTICHE DEVE AVERE UN SISTEMA DI RACCOMANDAZIONE?**

Per capire quali debbano essere le caratteristiche fondamentali che un Sistema di Raccomandazione debba soddisfare per rivelarsi utile ed efficace, riproponiamo per un attimo la seguente definizione:

*“Recommender Systems are software tools and techniques providing suggestions for items to be of use to a user”.*

Il soddisfacimento di questa definizione è dunque lo scopo per cui un Sistema di Raccomandazione viene realizzato ed utilizzato. Lo *User* è l'elemento centrale ed è nei suoi confronti che debbono essere garantite alcune basilari proprietà<sup>32</sup>.

#### **La trasparenza dei metodi utilizzati**

In statistica la validità di una previsione vale solo se è corredata da informazioni

---

<sup>32</sup> Questa parte è stata redatta prendendo spunto dalla tesi di Marco Zanchetta “Progettazione di un Service Recommender System basato su Intelligenza Collettiva”

inerenti il campione utilizzato e il metodo di calcolo impiegato: allo stesso modo, quando parliamo di sistemi di raccomandazione è buona norma fornire informazioni sulle tecniche e sui metodi di calcolo utilizzati. La giusta cura degli aspetti di trasparenza da parte del *Provider* può senz'altro comportare un maggior coinvolgimento dello *User* e ciò può rivelarsi un vantaggio se questi è invogliato ad affinare il proprio profilo con lo scopo di verificare un miglioramento del servizio di raccomandazione, ovvero a segnalare eventuali anomalie o incongruenze riscontrate nel servizio stesso. In tal caso, nei sistemi cosiddetti adattivi, sarà possibile tenere in conto la segnalazione, riconfigurando opportunamente la procedura di raccomandazione.

### **L'accuratezza dei risultati**

Affinché un Sistema di Raccomandazione possa essere veramente efficace, è necessario che lo *User* lo consideri affidabile. Ovviamente l'affidabilità, dal suo punto di vista, coincide con la capacità di fornire risultati soddisfacenti, che rispecchino le sue personali attese. Affidabilità dunque significa: accuratezza dei risultati. Poiché l'accuratezza del risultato è strettamente correlata con la capacità di profilare correttamente lo *User*, va da se che il successo di un sistema di raccomandazione dipenderà in larga parte dalla capacità del *Provider* di ottenere su di esso il maggior numero di informazioni, senza adottare procedure troppo invasive o prolisse (che rischierebbero di allontanarlo) e nel contempo garantendone la *privacy*. Spesso, anche in questo caso come nel precedente, può avere successo l'impiego di tecniche che riescano a coinvolgere lo *User*, facendolo diventare parte attiva del processo di raccomandazione, per esempio chiedendogli di valutare l'efficacia di un dato suggerimento.

### **La discrezione**

Un buon Sistema di Raccomandazione deve essere discreto. Se lo *User* ha anche solo la sensazione di sentirsi controllato o manipolato, abbandonerà il sito e

presumibilmente non vi ritornerà. Insistere troppo nel tentare di carpire informazioni attraverso tecniche esplicite (compilazione di *form* di profilazione) può rivelarsi controproducente. Meglio limitare al massimo le tecniche esplicite ed arricchire la profilazione utilizzando altri metodi basati, per esempio, sulla analisi dei comportamenti durante navigazione e sul completamento del profilo utilizzando le informazioni di altri utenti che abbiano avuto comportamenti simili.

### **L'ampiezza dei *dataset***

Può accadere che un Sistema di Raccomandazione sia realizzato con estrema cura e che riesca ad esser estremamente accurato, senza riuscire tuttavia a realizzare il soddisfacimento delle aspettative dei propri utenti. Spesso questo è dovuto ad una carenza del *dataset* di cui il sistema si avvale. L'ampiezza del *dataset* di riferimento (*Item Space Coverage*) e così pure l'ampiezza della base degli utenti (*User Space Coverage*), sono fattori in grado di discriminare tra un buon Sistema di Raccomandazione ed uno di modesta efficacia. I Sistemi di Amazon e di Google sono eccellenti anche perché possono contare su dei vastissimi *Item and User space coverage*.

L'ampiezza del *dataset* è una proprietà che a sua volta è strettamente connessa con un'altra importante proprietà: la scalabilità del sistema. È infatti abbastanza ovvio che dal momento in cui un sito (e con esso il suo Sistema di Raccomandazione) comincia ad avere successo e ad ampliare i suoi *space coverage*, dovrà essere in grado di sopportare efficacemente l'incremento di traffico che si verrà a realizzare, sia in termini di efficienza computazionale, sia in termini di gestione di elevate moli di dati.

### **La gestione delle novità**

La gestione delle novità è una questione piuttosto delicata e di non facile soluzione. Le novità, infatti, normalmente vengono inserite prive di alcuna valutazione da parte degli utenti e d'altro canto, soprattutto nei sistemi commerciali, sono proprio le novità che questi vorrebbero avere suggerite. In questo caso le tecniche maggiormente efficaci sono quelle che prendono come riferimento *Item* vecchi, ma simili ai nuovi e

questi vengono utilizzati per compilare dei giudizi preliminari sulle novità, in attesa che gli utenti comincino a fornire dei loro propri giudizi. È abbastanza chiaro che queste tecniche hanno tanto più successo quanto maggiore risulta l'ampiezza dei *dataset* a disposizione.

### **La serendipità**

Citando il vocabolario Treccani “serendipità ... è la capacità o fortuna di fare per caso inattese e felici scoperte, ... mentre si sta cercando altro”.

In questo contesto, se un Sistema di Raccomandazione fa sempre e solo la scelta giusta, farà senz'altro bene il proprio lavoro, ma poiché l'uomo non è una macchina e una certa imprevedibilità è spesso gradita, anche questa esigenza dovrebbe essere tenuta in debito conto. D'altro canto la maggior parte delle grandi scoperte, non è forse avvenuta per caso ?

L'imprevedibilità, o meglio, la serendipità, non deve essere quindi trascurata e i Sistemi di Raccomandazione più avanzati, infatti, inseriscono ogni tanto, inaspettatamente, dei suggerimenti che “non centrano nulla”. Lo *User* è libero di tralasciarli o di tenerli in conto.

### **La varietà delle raccomandazioni**

Non è sempre vero che ottenere subito la soluzione migliore sia un vantaggio: molto spesso gli *User* balzellano nel web qua e là per farsi un'idea, magari per organizzare un viaggio o per fare un regalo. In tutti questi casi sarebbe opportuno che il Sistema di Raccomandazione si “tenesse sul vago” organizzando le sue proposte per passi successivi: all'inizio esso fornisce indicazioni generali che vengono via via affinate interagendo con lo *User*. Un buon sistema di Raccomandazione dovrebbe essere in grado di percepire le reali intenzioni dello *User* valutando il suo comportamento durante la visita del sito.

### **La gestione dei *feedback***

La gestione dei *feedback* è una questione alquanto delicata: gli *User*, infatti, tramite i loro *feedback* possono intervenire direttamente nelle valutazioni del Sistema di Raccomandazione. Se alcuni *feedback* sono inseriti con scopi devianti o fraudolenti, tutto il sistema può risentirne in maniera irreparabile. Un Sistema di Raccomandazione deve quindi poter discriminare tali comportamenti, quando possibile, adottando per esempio tecniche di analisi statistica e valutando la provenienza delle diverse segnalazioni.

### **Il rispetto della *privacy***

In un Sistema di Raccomandazione spesso lo *User* è invitato a rivelare (talvolta inconsapevolmente) informazioni attinenti alla propria sfera personale. Il fine esplicito è noto: effettuarne una profilazione quanto più accurata possibile e fornirgli suggerimenti attendibili e utili. Altre finalità non sono concesse e le informazioni private debbono essere trattate come tali.

Il *Provider* di un Sistema di Raccomandazione dovrà quindi attuare una politica chiara e trasparente per quanto attiene agli aspetti di *privacy*, adottando una regolamentazione e un modello organizzativo che gli consentano di vigilare efficacemente sulla corretta utilizzazione di tutti i dati sensibili di cui viene in possesso.

## 5.5 *MODELLI E TECNICHE DI RACCOMANDAZIONE*

### 5.5.1 *MODELLI BASE DI RACCOMANDAZIONE*

In questo capitolo vengono illustrati alcuni modelli di raccomandazione che spesso costituiscono la base per sviluppare ulteriori modelli e tecniche come quelli che saranno presentati nel successivo capitolo

#### 5.5.1.1 *MODELLO COLLABORATIVO*

Quello collaborativo (*collaborative filtering*) è forse il più noto tra i modelli di raccomandazione – si ricordi il già citato Tapestry: sistema di filtraggio delle e-mail sviluppato alla Xerox di Palo Alto in California - e si basa sulla considerazione che “utenti simili hanno gusti simili”. Nel modello collaborativo ad un utente vengono forniti dei suggerimenti che si basano sulle scelte e sulle valutazioni che utenti con caratteristiche simili alle sue, hanno fatto in passato.

Le tecniche di analisi su cui si basa questo modello sono svariate e sono tutte focalizzate sui due metodi principali<sup>33</sup>:

***neighborhood based methods*** - “*In this method the user-item ratings stored in the system are directly used to predict ratings for new items. This can be done in two ways known as user-based or item-based recommendation.*”

Nei sistemi focalizzati sull'utente (*User based*), viene valutato il potenziale interesse di un dato *User* per un certo prodotto, utilizzando le valutazioni espresse da altri utenti ritenuti affini (*neighbors*) allo *User* in questione. Gli utenti *neighbors* dello *User* sono normalmente coloro che hanno espresso pareri simili a quelli dello *User* su prodotti valutati da entrambi.

Nei sistemi focalizzati sul prodotto (*Item based*) viene fatta una previsione della valutazione che lo *User* darebbe ad un dato prodotto, basandosi sulla valutazione data

---

<sup>33</sup> La parte che segue è compendiata dal cap. “A Comprehensive Survey of Neighborhood-based Recommendation Methods” - Christian Desrosiers and George Karypis da “F. Ricci et al. (eds.), Recommender Systems Handbook”

dallo stesso *User* su prodotti simili a quello in questione. In questo caso due prodotti sono ritenuti simili se più utenti ne hanno dato una valutazione, secondo modalità e in condizioni analoghe.

***model-based methods*** - *“In contrast to neighborhood-based systems, which use the stored ratings directly in the prediction, model-based approaches use these ratings to learn a predictive model.”*

L'idea che sta alla base di questo metodo è quella di modellare l'interazione utente-prodotto con fattori che rappresentino caratteristiche nascoste, degli utenti e dei prodotti, all'interno del sistema. Tali fattori possono riguardare, per esempio, la classe delle preferenze a cui appartiene un utente o la classe della categoria di un dato prodotto. Una volta individuati i fattori, il sistema viene istruito utilizzando il *dataset* disponibile e successivamente può essere impiegato per prevedere le valutazioni degli utenti sui nuovi prodotti.

Questo ultimo metodo, oltre a presentare sicuri vantaggi rispetto al primo metodo per quanto attiene la accuratezza della raccomandazione fornita, soprattutto nel caso di nuovi prodotti, presenta un altro aspetto degno di nota: poiché opera su caratteristiche recondite degli utenti e dei prodotti, si presta vantaggiosamente ad essere impiegato anche con approcci orientati alla serendipità della raccomandazione.

#### 5.5.1.2 MODELLO BASATO SUI CONTENUTI

Il modello basato sui contenuti (*content based*) si focalizza sulla raccomandazione di prodotti simili a quelli che lo *User* ha già apprezzato in passato<sup>34</sup>:

*“Indeed, the basic process performed by a content-based recommender consists in matching up the attributes of a user profile in which preferences and interests are stored, with the attributes of a content object (item), in order to recommend to the user new interesting items.”*

I sistemi di raccomandazione che implementano un modello basato sui contenuti,

---

<sup>34</sup> La parte che segue è compendiata dal cap. “Content-based Recommender Systems: State of the Art and Trends” - Pasquale Lops, Marco de Gemmis and Giovanni Semeraro da “F. Ricci et al. (eds.), Recommender Systems Handbook”

analizzano una serie di documenti e descrizioni correlate a prodotti in precedenza valutati da uno *User* e ne costruiscono un profilo, basato sulle caratteristiche dei prodotti valutati.

Il processo di raccomandazione consiste, in pratica, nel far corrispondere gli attributi del profilo delle *User* con gli attributi del prodotto, definendo in tal modo il livello di interesse dello *User* per quel dato prodotto. Si può così decidere se valga o meno la pena che questo gli sia suggerito o che sia introdotto nella lista dei prodotti da suggerire.

Esistono delle tecniche ben precise sia per fornire la rappresentazione dei prodotti, sia per definire il profilo dello *User*, come pure esistono delle consolidate strategie per calcolare la corrispondenza di questi due fattori. Nel seguito vengono accennate brevemente alcune di tali tecniche, rimandando ad una letteratura più specifica eventuali approfondimenti:

**analisi del contenuto** – quando si ha a che fare con una informazione non strutturata (per esempio un testo scritto) è necessario effettuare una attività preliminare per l'estrazione delle informazioni rilevanti e una loro strutturazione in modo da renderla fruibile per successive elaborazioni. Si tratta di rendere il testo *machine accessible* attraverso una sorta di metadatazione dei contenuti o più semplicemente la compilazione di una lista di parole chiave riferite al testo stesso.

Il risultato di questa attività di analisi costituisce l'*input* per una successiva attività di:

**apprendimento del profilo** – in questa fase vengono raccolte informazioni utili alla rappresentazione delle preferenze dell'utente, passando poi a generalizzare tali informazioni nel tentativo di costruire un vero e proprio profilo dell'utente stesso. Per effettuare la generalizzazione delle preferenze vengono adottate tecniche attraverso le quali si è in grado di inferire un modello degli interessi dell'utente a partire da una serie di valutazioni su prodotti che egli ha apprezzato o meno in passato. Anche questa tecnica, come la precedente, ha notevoli correlazioni con le tecniche legate allo sviluppo di ontologie e di sistemi di organizzazione della conoscenza di cui si è parlato nella prima parte di questa tesi (cfr. cap. 4.4.1.3 sui Ragionatori).

Il risultato di questa attività di apprendimento, costituisce l'*input* per una successiva attività di:

**filtraggio dei prodotti** – in questa ultima fase il profilo dell'utente viene utilizzato per suggerire prodotti di potenziale interesse. Attraverso una analisi delle corrispondenze tra la rappresentazione del profilo utente e gli attributi dei prodotti, viene effettuato un filtraggio dei prodotti di scarso interesse e viene compilata una lista ordinata di quelli raccomandabili.

### 5.5.1.3 MODELLO BASATO SULLA CONOSCENZA

Il modello basato sulla conoscenza (*knowledge based*) si focalizza sulla raccomandazione di prodotti che possono corrispondere alle aspettative dell'utente:

*“Knowledge-based systems recommend items based on specific domain knowledge about how certain item features meet users needs and preferences and, ultimately, how the item is useful for the user.”*

I sistemi di raccomandazione basati sulla conoscenza sono normalmente di due tipi: *case-based* o *constraint-based*. Nel primo tipo viene ottimizzata una funzione di similarità valutando quanto le aspettative dell'utente (descrizione del problema) corrispondono alla raccomandazione (soluzione del problema). Il valore della funzione rappresenta in questo caso, l'utilità della raccomandazione per l'utente.

Nel secondo caso i presupposti sono più o meno gli stessi se non per le modalità di calcolo delle soluzioni: nell'approccio *constraint-based* infatti, la base di conoscenza viene arricchita con regole esplicite su come mettere in relazione i requisiti dell'utente con le proprietà dei prodotti.

I modelli basati sulla conoscenza presentano dei vantaggi rispetto ad altri modelli, soprattutto nelle fasi iniziali di utilizzo. Infatti gli approcci basati sui contenuti o sulla collaborazione soffrono di un fondamentale problema: non funzionano bene se il *dataset* non è sufficientemente ampio ovvero se non sono noti i comportamenti di una vasta platea di utenti o se non è stato raccolto un sufficiente numero di valutazioni sui prodotti.

I modelli basati sulla conoscenza superano questo problema e possono essere quindi utilmente impiegati come complementari ai modelli *collaborative* e *content-based* in quanto:

*“its recommendations do not depend on a base of user ratings. It does not have to gather information about a particular user because its judgements are independent of individual tastes.”*<sup>35</sup>

Si preferisce parlare di complementarità rispetto agli altri metodi in quanto i *knowledge-based* dovrebbero essere equipaggiati con degli specifici strumenti di apprendimento per poter essere competitivi anche nelle fasi successive a quelle iniziali, in cui il *dataset* comincia a diventare consistente.

Nel seguito viene accennato un approfondimento della tecnica *constraint-based*<sup>36</sup>, rimandando alla lettura del testo citato in nota per ulteriori dettagli:

*“A recommender knowledge base of a constraint-based recommender system typically is defined by two sets of variables (VC, VPROD) and three different sets of constraints (CR, CF, CPROD). Those variables and constraints are the major ingredients of a constraint satisfaction problem. A solution for a constraint satisfaction problem consists of concrete instantiations of the variables such that all the specified constraints are fulfilled.*

**Customer Properties** *VC describe possible requirements of customers, i.e., requirements are instantiations of customer properties. In the domain of financial services willingness to take risks is an example for a customer property and willingness to take risks = low represents a concrete customer requirement.*

**Product Properties** *VPROD describe the properties of a given product assortment. Examples for product properties are recommended investment period, product type,*

---

35 “Knowledge-based recommender systems” - Robin Burke - Department of Information and Computer Science University of California, Irvine – scritto per la “Encyclopedia of Library and Information Science”

36 Tratto da “Developing Constraint-based Recommenders” - Alexander Felfernig, Gerhard Friedrich, Dietmar Jannach and Markus Zanker – da “F. Ricci et al. (eds.), Recommender Systems Handbook”

*product name, or expected return on investment.*

**Constraints** *CR are systematically restricting the possible instantiations of customer properties, for example, short investment periods are incompatible with high risk investments.*

**Filter Conditions** *CF define the relationship between potential customer requirements and the given product assortment. An example for a filter condition is the following: customers without experiences in the financial services domain should not receive recommendations which include high-risk products.*

**Products** *Finally, allowed instantiations of product properties are represented by CPROD. CPROD represents one constraint in disjunctive normal form that defines elementary restrictions on the possible instantiations of variables in VPROD.“*

#### 5.5.1.4 MODELLO IBRIDO

Il modello ibrido è chiaramente un modello nel quale coesistono almeno due dei modelli accennati in precedenza e si basa dunque su una combinazione di tecniche. Per esempio applicando un modello *knowledge-based* combinato con un modello *content-based* potremo ottenere delle buone *performance* sia nella fase iniziale di lancio di un prodotto, nella quale avremo a disposizione poche valutazioni, sia nelle fasi successive in cui il *dataset* avrà raggiunto dimensioni più consistenti. Inoltre nei sistemi ibridi possiamo impiegare contemporaneamente diverse tecniche con lo scopo di definire una raccomandazione più ricca soprattutto in contesti nei quali può avere senso tenere in considerazione l'ambiente in cui l'utente si trova.

Per esempio, rifacendosi ad un caso attinente con la ricerca effettuata per questa tesi, ad un turista che si trova in vacanza in una città d'arte, non andremo a suggerire la visita di un museo che richiede almeno mezza giornata a disposizione, se sono già le cinque di sera, come ad una famiglia con figli piccoli non proporremo una serata in discoteca e a una coppia non proporremo una passeggiata sulla spiaggia se sta piovendo.

Non approfondiamo oltre questo capitolo e in particolare i metodi di contestualizza-

zione in quanto di questi discuteremo più oltre, nel successivo capitolo dedicato alle tecniche e ancor più nella terza parte della tesi, dove affronteremo il tema anche da un punto di vista applicativo.

## 5.5.2 ALTRE TECNICHE DI RACCOMANDAZIONE

In questo capitolo si descrivono alcune tecniche che sono state prese in considerazione, con diversi livelli di approfondimento, per la realizzazione delle attività di ricerca cui si riferisce questa tesi.

### 5.5.2.1 ONTOLOGY BASED

I sistemi di raccomandazione *ontology-based* utilizzano e sfruttano le tecniche e le tecnologie del *semantic web* per godere di un grande vantaggio potenziale: avere a disposizione una base di conoscenza enormemente vasta, *machine-accessible*. I sistemi di questo tipo utilizzano il modello ontologico per ottenere una conoscenza approfondita del dominio in cui la raccomandazione viene normalmente erogata. La base di conoscenza ontologica può essere impiegata sia per caratterizzare il dominio dell'utente, sia per descrivere il dominio del prodotto. Per esempio, se parliamo di turismo, questo sarà il dominio in cui l'utente verrà profilato e l'ontologia sarà utilizzata per descriverlo in un contesto di vacanza, tenendo in considerazione: le diverse fasi in cui si sviluppa il suo viaggio (dalla ideazione, alla pianificazione, fino alla fruizione *in loco*), le modalità di fruizione della vacanza (*all inclusive*, viaggio avventura, da solo, in coppia, in famiglia), i suoi interessi (divertimento, cultura, natura, socializzazione, relax), le abitudini e i comportamenti (si alza presto, tira tardi la notte, si muove in bicicletta) e le eventuali attività *social* (come proposto nel progetto FOAFRealm<sup>37</sup>, un sistema di gestione dei profili basato sul *social networking* e definito secondo il vocabolario FOAF).

D'altro canto, il dominio del prodotto sarà inerente all'offerta turistica e potrà riguardare: i servizi (*accommodation*, ristorazione, escursionismo, *divertissement*), gli eventi (concerti, feste, incontri), l'arte e la cultura (musei, monumenti), la natura (parchi, spiagge, paesaggi) e così via.

Un progetto interessante anche se datato, utilizza un approccio misto come quello appena esemplificato: è stato condotto dai coreani Sungrim Kim e Joonhee Kwon

<sup>37</sup> In questo progetto la profilatura si basa sulla definizione di relazioni sociali più o meno strette (ho una relazione più stretta col mio migliore amico che non con un mio amico qualsiasi)

che nel 2007 hanno sviluppato un sistema di raccomandazione basato su un “*ontology model for grocery store. Our ontology is categorized into four distinctive but related themes: (i)ontology about product, (ii)ontology about location, (iii)ontology about the record, and (iv)consumer. The “Product” ontology specifies the products that are recommended by our method, and the “Location” ontology specifies the contextual information, especially the location context. The “Consumer” and “Record” ontology are used to model the consumer records and the shopping records.*”

Con questo metodo la raccomandazione viene progressivamente mirata sviluppando in sequenza tre fasi:

- nella prima si estrae una lista di prodotti preferiti dall'utente, sulla base di acquisti effettuati in precedenza; per fare questa estrazione vengono selezionate le classi e le istanze di prodotti che superano un determinato punteggio di gradimento, all'interno della ontologia n. 1 (*ontology about product*).
- nella seconda fase viene determinato il livello di informazione che deve essere fornito all'utente, in termini di minore o maggiore dettaglio. Questo significa che in prima battuta può essere fornita una informazione più generale che in seguito può venire raffinata. Per ottenere questo risultato viene costruita una gerarchia di concetti, dove quelli di livello superiore (i genitori) hanno dei significati più generalisti di quelli di livello inferiore (i figli). Il passaggio da un livello superiore ad uno inferiore, dipende dal fatto che l'utente individui un concetto di suo interesse, il che gli consente un “cambio di contesto”.
- Nella terza ed ultima fase la raccomandazione derivante dalla seconda fase viene ulteriormente raffinata in funzione del grado di attenzione dell'utente. Questo viene calcolato in base al livello del contesto e al comportamento dell'utente stesso. Per velocizzare le attività computazionali, viene fatta, quando possibile, una predizione di quelle che potrebbero essere le informazioni da raccomandare: “*When attention to the current context is high,*

*the method estimates which other contexts that are contained by the context are likely to be visited in the near future. The prefetch weight is used to prefetch the information in the recommendation. If the prefetch weight for a context is low, there is a low possibility that the recommended information for the context will be accessed in the near future.”*

I modelli ontologici e in generale le tecniche di caratterizzazione semantica possono essere utilmente impiegate anche nella fase di profilazione utente. Uno studio a riguardo, che vale la pena di citare, è quello condotto da Hubert Kadima, Maria Malek del francese Laris/Eisti Laboratory: *“Toward ontology-based personalization of a Recommender System in social network”*.

In questo studio i due ricercatori affermano quanto segue:

*“typically, the recommender systems have to maintain a model of each user profile with respect to his knowledge of domain concepts and user preferences evolution. A variety of classical techniques ranging from simple statistics to machine learning algorithms (k-nearest neighbour, naive Bayes, SVM) may be applied to deliver personalization solutions. Recently, the emergence of the social semantic web provides us opportunity to revisit personalization approaches for the social networks environment. In general, the different ontologies in social networks should help to distinguish models based on the representation of business data and those established for the representation of document structures and social interactions on the web. Our unified user profile context is represented as an instance of a reference domain ontology in which concepts are annotated by interest scores. We use semantic user profile and the RDF-based user model exchange language UserML to maintain consistency between individual user profile contexts.”*

Oltre ai progetti citati in letteratura possiamo trovare numerose altre iniziative di progetti che sviluppano sistemi di raccomandazione a partire da modelli ontologici: molti di questi sistemi adottano tecniche ibride per cui non è semplice attribuirli rigorosamente ad una categoria piuttosto che a un'altra. Senza dilungarci troppo

quindi, rimandiamo alla lettura del capitolo dedicato alle tecniche *social* dove sono presentati ancora un paio di esempi di progetti che sfruttano la semantica per affinare il processo di raccomandazione.

#### 5.5.2.2 CONTEXT AWARE

*"Context is any information that can be used to characterize the situation of an entity.*

*An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."*<sup>38</sup>

Ci sono in realtà due interpretazioni che possono essere date del termine “contesto” e che sono definite nel libro “*Mobile Design and Development*” di Brian Fling:

- Contesto (con la C maiuscola): che riguarda il “come” una persona percepisce, interpreta, vive, un dato momento o una data circostanza. Per esempio: se mi trovo all'interno della Basilica di Santa Maria del Popolo a Roma e sto guardando “La conversione di San Paolo” di Caravaggio, la mia esperienza sarà ben diversa da quella che avrei se fossi a casa mia sul divano ad ammirare lo stesso quadro stampato su un libro;
- contesto (con la c minuscola): riguarda il “modo” o “l'ambiente” in cui una data attività viene realizzata. Riferendosi ancora all'esempio precedente avremo che il contesto, nel primo caso è “la Basilica”, nel secondo caso è “casa mia”.

Nel caso delle applicazioni nel campo del turismo e in particolare del turismo *mobile*, che sono quelle che maggiormente ci interessano, il contesto (con la c minuscola) è quello che ci interessa: perché è più chiaro a che cosa ci si riferisce ed è più facile trattarlo coi normali strumenti di elaborazione di cui disponiamo. Si tratta infatti di suggerire ad un utente dei percorsi, degli oggetti, dei comportamenti, delle storie che possono essere condizionati dal fatto che egli, durante la visita di una città, si trovi in un certo luogo, in un dato momento e in determinate condizioni ambientali.

---

<sup>38</sup> Anind K. Dey - “*Understanding and using context*” - *Personal and Ubiquitous Computing* 2001.

Da un punto di vista tecnico l'elaborazione *context aware* si basa su tre differenti metodi di calcolo per l'integrazione delle informazioni di contesto, nel processo di raccomandazione:

nella prima, la cosiddetta tecnica *reduction based* (o di *pre-filtering*), per la generazione delle raccomandazioni vengono prese in considerazione solamente informazioni che corrispondono al contesto di uso corrente oppure sono considerate valide come riferimento valutazioni effettuate in un contesto analogo all'attuale.

Nella seconda tecnica detta di *contextual post filtering*, le informazioni di contesto vengono dapprima ignorate. Successivamente l'insieme di tutte le raccomandazioni generate viene filtrato ritenendo valide solo raccomandazioni considerate rilevanti nel contesto attuale.

Nella terza e ultima tecnica, di *contextual modeling*, i dati di contesto vengono esplicitamente utilizzati nel modello predittivo.

Un interessante progetto basato sul *context aware* è il **Progetto meSch** (*Material EncounterS with digital Cultural Heritage*) finanziato dalla Comunità Europea (ICT Call 9:FP7-ICT-2011-9) e coordinato dalla *Sheffield Hallam University* avviato da pochissimi mesi e che terminerà nel 2017:

*“meSch will deliver a platform to enable curators, artists, designers and cultural heritage professionals in general to create smart objects and intelligent spaces. Smart objects (like a magnifying glass or a replica) are enriched with digital technology while intelligent spaces embed sensors: both react to people, spaces and smart objects. A bespoke application will adapt the content and the behaviour of the object or space to visitors, their social context and the environment.”*

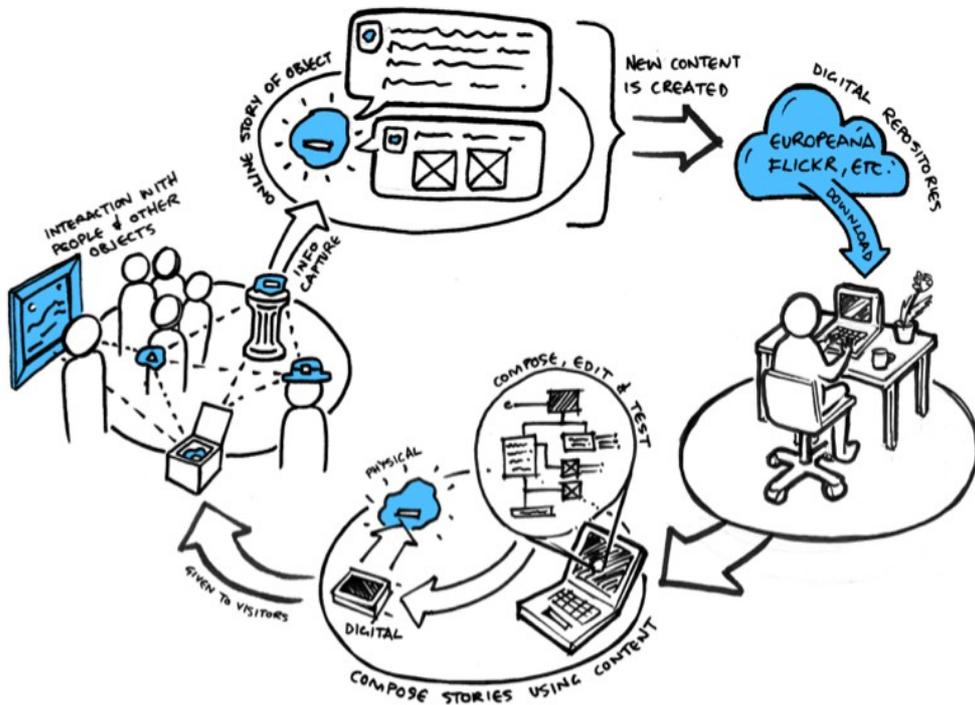


Figura 26: schema del meSch

### 5.5.2.3 TECNICHE SOCIAL

Quelle *social (community based)* sono tra le tecniche più recenti e traggono spunto dalla grande diffusione dei *social network* in cui un utente è in contatto con “amici” coi quali condivide molte delle sue attività quotidiane. Queste tecniche raccomandano allo *User* dei prodotti in base alle preferenze espresse dai suoi amici, considerando che è più facile che un utente tenga in considerazione i suggerimenti di questi ultimi piuttosto che quelli di sconosciuti, benché con gusti simili ai suoi. La ricerca in questo specifico campo è piuttosto recente e tuttavia ha delle implicazioni di grande interesse e molto promettenti.

Il web e in particolare il web sociale e collaborativo (web 2.0) è fonte inesauribile di classificazioni, valutazioni, tassonomie e informazioni personali generate dalla grande *community* degli utenti del web e dei *social network*. Elaborando questa enorme mole di informazioni non è difficile ricavare informazioni anche personali e ricostruire la sua rete di relazioni sociali.

In letteratura si possono trovare diversi approcci e soluzioni per la realizzazione di sistemi di raccomandazione basati su tecniche *social*: ne vediamo un paio.

### **A Multi View Recommendation Engine**

Questo lavoro è stato presentato nel 2009 da Houda Oufaida, Omar Nouali del centro di ricerca algerino CERIST.

*“...we introduce a multi view recommender system that includes collaborative, social and semantic views of the user’s profile. Each view recommends a set of items. Hence, three hybridization strategies are proposed for recommendations re-ranking.”*

Oufaida e Nouali ipotizzano che la raccomandazione sia calcolata a partire da un profilo utente rappresentato su tre dimensioni o viste:

- una vista “collaborativa” che si basa sui ratings espressi dall'utente in forma esplicita o implicita;
- una vista “socio-demografica” che contiene l'anagrafica dell'utente (età, sesso, professione, contatti etc.);
- una vista “semantica” che rappresenta i diversi interessi dell'utente, ordinati secondo un peso che gli è stato attribuito.

Ognuna di queste tre viste viene utilizzata per inserire l'utente in uno specifico ambito (*neighborhood*):

- il **Collaborative Neighborhood**: la vista collaborativa contiene i *rating* espliciti e impliciti espressi dall'utente; in questo caso viene impiegato un algoritmo di correlazione per calcolare la similarità degli utenti e i  $k$  vicini più prossimi all'utente dato.
- Il **Social Neighborhood**: questo ambito viene costruito a partire dalla rete sociale dell'utente sulla quale egli ha espresso delle valutazioni di “fiducia”. È l'utente stesso quindi che esprime valutazioni di fiducia sui suoi amici,

indicandole su una scala che va da 1 (scarsa fiducia) a 5 (massima fiducia). Sulla base di queste valutazioni si può ricavare la rete sociale che in termini di fiducia rappresenta “il vicinato” dell'utente in questione.

- Infine il **Semantic Neighborhood**: questo ambito viene costruito a partire dall'interesse che l'utente esprime sul contenuto di generici *Item*. Ovviamente in questo caso è necessario che sia disponibile una rappresentazione semantica del contenuto degli *Item*.

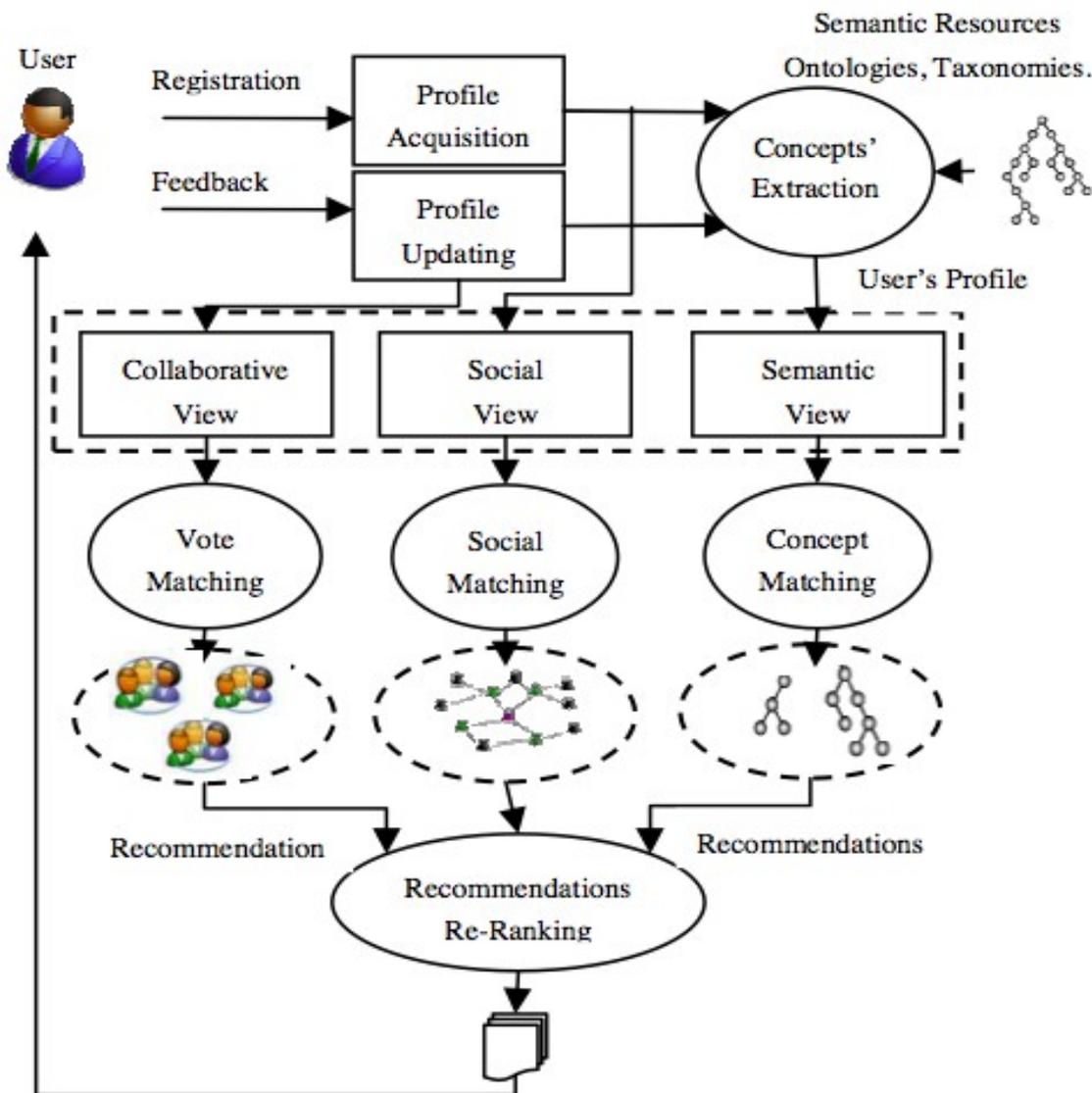


Figura 27: schema del multi view recommendation engine, tratto dal paper *Exploiting "Semantic Web Technologies for Recommender Systems A Multi View Recommendation Engine"*

## Ontology-based Web Recommendation from Tags

In questo lavoro di Nizar R. Mabroukeh e C. I. Ezeife della Università canadese di Windsor viene affrontato il tema dei *Web Recommendation Systems*, basati sulla valutazione di *Item* e sulla *browsing history* degli utenti del web. Più in particolare:

“...we discuss how an ontology-based WRS can utilize relations and concepts in an ontology, along with user-provided tags, to provide top-n recommendations without the need for item clustering or user ratings. For this purpose, we also propose a dimensionality reduction method based on the domain ontology, to solve the sparsity problem.”

Un *Web Recommendation System* (WRS) è un sistema capace di fornire raccomandazioni in tempo reale in base al comportamento dell'utente durante la navigazione web. Una volta riconosciuto un *pattern* di comportamento il WRS è in grado di predire le preferenze dell'utente e quindi produrre la raccomandazione. In questi termini il WRS non è altro che un sistema *content based*.

Molti WRS utilizzano la tecnica del raggruppamento in *cluster* sia degli utenti, sia degli *item*, quando questi posseggano caratteristiche che li accomunano, e quindi sono in grado di fornire ad un utente raccomandazioni che si basano sui gusti di altri utenti. E questo può essere considerato un WRS basato sul *collaborative filtering*.

Nel lavoro di Mabroukeh ed Ezeife viene proposto un approccio basato sul *tagging* che gli utenti spesso provvedono in numerosissimi sistemi nati col web 2.0 e con le reti sociali.

In questo sistema i *tag* vengono mappati sui concetti di una ontologia di dominio cosicché il processo di raccomandazione diventa in pratica un processo in cui: si effettua la ricerca dei concetti che hanno qualche affinità coi *tag* che vengono realizzati dall'utente; si raccomandano i primi *n Item* che condividono quei concetti.

Più in particolare il sistema proposto si sviluppa nei seguenti tre passaggi:

- *Clickstream mapping*: il *web log* acquisito, contiene coppie di *Item* coi *tag* loro associati. Viene computata la similarità tra ciascun *Item* e ciascun concetto “foglia” della ontologia. Tutte le similarità sono memorizzate in una matrice  $item \times concetto$  la cui dimensione dipende dal numero di concetti foglia della

ontologia.

- *Active recommendation*: quando l'utente arriva ad una data pagina web (ovvero compra un certo prodotto) i *tag* associati a questo *Item* vengono recuperati e viene generato un vettore (simile a una riga della matrice). Questo vettore rende conto della similarità tra i *tag* attivi di quell'utente e ciascun concetto della ontologia. Dopodiché il vettore viene comparato con ciascuna riga della matrice e i migliori *n matching Item* sono utilizzati per la raccomandazione.
- *Expanding recommendation*: il set delle raccomandazioni può essere ulteriormente espanso aumentando *n*, come è ovvio, oppure aumentando la dimensione del vettore, utilizzando le relazioni semantiche della ontologia, con lo scopo di includere un maggior numero di concetti. Tali ulteriori concetti, non essendo presenti nella matrice, dovranno prevedere il disegno di una raccomandazione *ad hoc*.

#### 5.5.2.4 TECNICHE CONVERSAZIONALI

Le tecniche conversazionali hanno lo scopo di creare una interazione con l'utente di modo che, dai suoi comportamenti, sia possibile affinarne il profilo e definire la raccomandazione affinché questa possa rifletterne il più possibile i *desiderata*.

Una delle tecniche conversazionali di base e di maggior efficacia è la “domanda”: questa infatti, se posta correttamente, può guidare l'attenzione dell'utente inducendo una percezione nuova e diversa di una data situazione. Probabilmente in molti casi non è nemmeno importante ottenere una risposta o una data risposta poiché il semplice fatto di nominare determinate parole attiva nell'interlocutore sensazioni o bisogni che possono orientare le sue scelte. Se all'utente chiedo: “ti piacerebbe vedere un buon concerto? Tra mezz'ora nel tal luogo suona il tale artista” non mi aspetto che mi sia risposto “sì” ma semplicemente desidero suggerire all'utente una possibilità: se poi egli effettivamente andrà al concerto questa informazione sarà utile per affinarne il profilo.

Una esperienza degna di nota, di approccio conversazionale per la raccomandazione

di Film, è stata recentemente realizzata da Eoin Hurrell e Alan F. Smeaton<sup>39</sup> della *Dublin City University*:

*“we explore a new approach to conversation within recommendation as applied to visual item, namely movies. We have developed a way to generate conversation around a large dataset, letting users navigate their recommendations in situations where metadata about items is not present. An application called MovieQuiz, which allows users to quickly navigate the dataset to alter the initial recommendation given to them, is used as a basis for an evaluation of our approach. We recorded user interactions, ratings and responses to a follow-up survey for the purposes of evaluation and we show the ways in which our interactions can improve a user’s ability to find good recommendations.”*

L'approccio proposto da Hurrell e Smeaton ruota attorno all'idea di utenti che definiscono progressivamente la loro area di maggior interesse. Essi ipotizzano che utilizzando solamente il numero di *rating* e il *rating* medio di un dato *Item*, possono riuscire a ridurre il *set* di elementi da raccomandare fornendo in tal modo dei suggerimenti più mirati.

Per affinare l'area da cui estrarre gli *Item* da suggerire essi producono dei *feedback* basati su reazioni, immediate o ragionate, del tipo: “non credo che mi piacerebbe..” ovvero “sono interessato a..”. Sebbene una modalità di interazione così netta possa ritenersi imprecisa e difficile da catturare, tuttavia essa viene utilizzata per due motivi:

- una parte importante del processo decisionale di una persona è basato su interazioni di questo tipo;
- a differenza della modalità di conversazione *case-based* (la più frequente) in cui si esprimono valutazioni basate su liste del tipo “sono più interessato a questo che a quello”, questo approccio ha probabilmente il vantaggio di poter essere utilmente impiegato in sessioni interattive dove c'è bisogno di acquisire velocemente alcune indicazioni da parte dell'utente.

---

<sup>39</sup> “*A Conversational Collaborative Filtering Approach to Recommendation*”

Per implementare questa tecnica vengono utilizzati i *rating* sugli *Item* per formare dei *cluster* di utenti o trovare *Item* simili a quelli che l'utente sta cercando. In pratica la tecnica funziona così:

- in base ai *rating* viene calcolata la popolarità di ciascun *Item* e la misura di quanto quel *Item* sia ben valutato (*rating* medio). In base a queste valutazioni l'insieme degli *Item* può essere rappresentato su un grafico con la valutazione media in ascisse e la popolarità in ordinate;
- all'utente vengono proposti due *Item* a caso (comunque appartenenti ad un'area di suo interesse): uno con valore medio di *rating* e uno con popolarità elevata;
- all'utente viene chiesto di esprimersi “quale dei due preferisci?” (a margine vengono comunque proposte delle alternative per dare all'utente la possibilità di fare nuove scelte);
- l'*Item* scartato rappresenta il confine inferiore dell'insieme degli *Item* accettabili;
- il procedimento può essere reiterato finché l'utente non interrompe o se vengono raggiunti dei limiti preimpostati.

# TERZA PARTE

## *LO SVILUPPO SPERIMENTALE*

## **6 RACCOMANDAZIONE PER SERVIZI TURISTICI AVANZATI**

Questa terza parte della tesi è dedicata allo sviluppo di una sperimentazione che tenta di mettere a frutto le esperienze fatte in questi anni proponendo un modello di sistema di raccomandazione per l'erogazione di servizi avanzati e innovativi per il turismo.

Nella fattispecie, essa stessa viene suddivisa in due parti: una prima dedicata ad analizzare le problematiche di impianto di un sistema di raccomandazione nel momento in cui si pensasse, per esempio, di avviare una attività di impresa per l'erogazione di tali servizi in ambito “turismo”; una seconda che cerca di sviluppare una idea di servizio di raccomandazione cosiddetto “evocativo” per la guida di un generico turista alla visita delle bellezze nascoste di una città o di una qualsiasi località di vacanza.

### **6.1 STUDIO PER L'IMPIANTO DI UN SISTEMA DI RACCOMANDAZIONE**

Il modello che viene presentato in questo capitolo è stato studiato nel 2012 con lo scopo di presentare al Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) un progetto di prevalente ricerca industriale sul tema dello “sviluppo di un servizio di raccomandazione per il turista culturale; il titolo della ricerca è *Recommending to People* da realizzare nel settore di attività ICT; la durata preventivata del progetto è di 30 mesi”.

L'ottenimento dell'eventuale finanziamento, presupponeva che venisse costituita una società a responsabilità limitata, spin-off dell'Università Iuav, per la conduzione del progetto e il suo successivo sviluppo industriale.

La domanda, presentata nel giugno 2012, non è mai stata valutata compiutamente in quanto dopo pochi mesi la linea di finanziamento è stata soppressa, dopo l'insediamento del governo Monti.

Tuttavia il lavoro svolto per la presentazione della domanda può ritenersi ancora sufficientemente valido per essere riproposto, con qualche piccolo aggiustamento e

aggiornamento, anche in questo contesto di tesi.

### 6.1.1 DEFINIZIONE DEGLI OBIETTIVI GENERALI

Obiettivo generale del progetto è quello di realizzare un ecosistema digitale nel quale possano interagire diverse componenti (uomini, macchine, sensori, servizi...) con lo scopo di contestualizzare il turista culturale (ma più in generale qualunque turista) rispetto all'ambiente e allo scenario culturale in cui si trova e fornirgli suggerimenti e informazioni anche in base alle sue normali aspettative.

Il processo che sta alla base del servizio proposto, può essere così schematizzato:

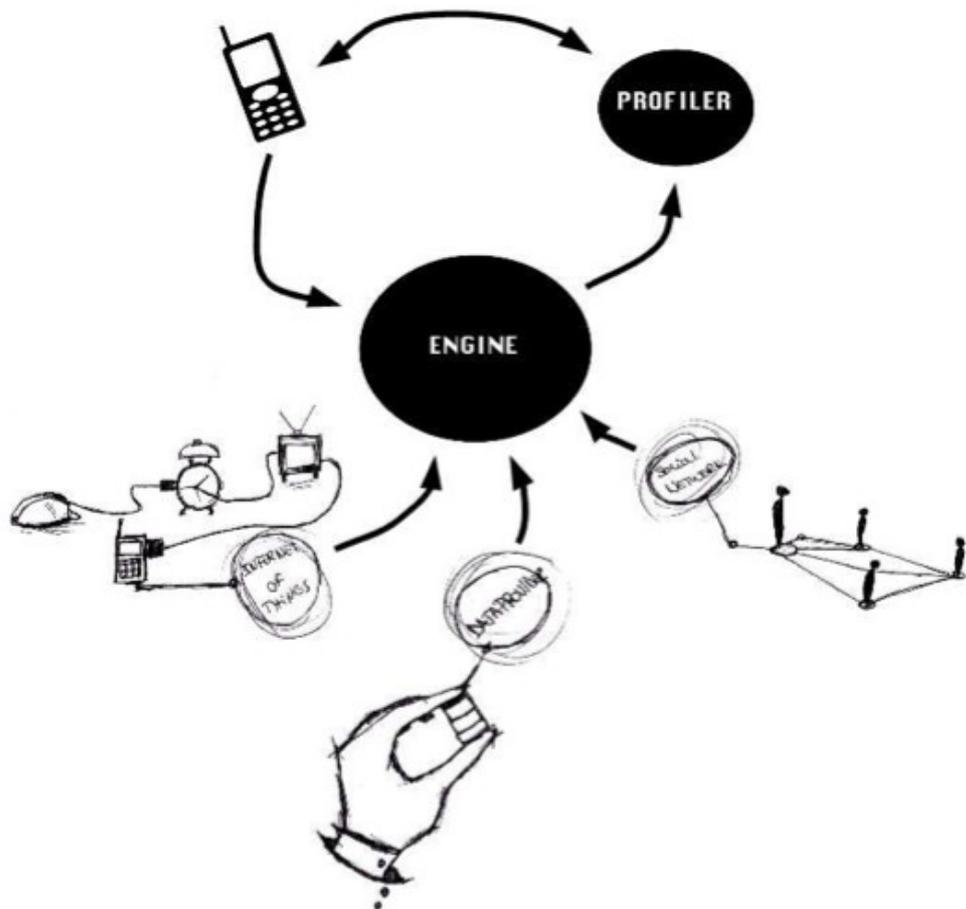


Figura 28: schema generale del sistema di raccomandazione proposto

Nel modello proposto i dati di contesto e le informazioni provenienti dalla *collective intelligence*<sup>40</sup> e dal web sono raccolti, omogeneizzati ed organizzati utilizzando un *Engine* attraverso il quale essi vengono caratterizzati anche dal punto di vista semantico. L'output del *Engine* alimenta un modulo *Profiler* che ne caratterizza i contenuti in funzione del profilo e dei comportamenti dell'utilizzatore che ha richiesto il servizio. I suggerimenti e le raccomandazioni sono quindi inviate in maniera asincrona all'utilizzatore stesso che ne può usufruire attraverso una interfaccia multimodale ed eventualmente attraverso diversi dispositivi. Una caratteristica ritenuta importante e che il servizio dovrebbe garantire, nell'ipotesi in esame, è quella della "discrezione" della raccomandazione, con lo scopo di non modificare le abitudini e gli usi tipici del dispositivo da parte dell'utente.

Il dominio di applicazione del modello appena descritto, rispetto alla ricerca proposta, riguarda il turismo culturale, limitatamente al contesto ben definito e conosciuto della città storica di Venezia.

Nel contesto veneziano, dunque, i numeri dei potenziali utilizzatori del servizio sono, *in primis*, quelli legati ai flussi turistici medi (si veda il rapporto "Turismo sostenibile a Venezia", redatto nel 2011 dall'Ufficio di Piano del Magistrato alle Acque, sui flussi turistici a Venezia: nel 2008 12 ML di escursionisti e 7 ML di pernottamenti nella Venezia insulare) e più in particolare a coloro che si appoggiano alla piattaforma di connettività e servizi *wireless* "Venice Connected" per la prenotazione e la vendita di servizi pubblici della città (oltre 190.000 ticket commercializzati nel 2011 a turisti provenienti da 188 paesi). Il sistema dovrà essere quindi opportunamente dimensionato per fare fronte al potenziale traffico generato dai numeri appena citati. In ogni caso esso sarà studiato affinché presenti caratteristiche di scalabilità tali da non comprometterne la funzionalità nel caso il numero degli utenti dovesse aumentare in maniera sostanziale rispetto a quanto inizialmente ipotizzato.

---

40 Definita come l'intelligenza che emerge dalla collaborazione, dalla competizione e dallo scambio di informazioni di individui all'interno di un gruppo, per esempio all'interno di un *social network*

### 6.1.2 PRINCIPALI PROBLEMATICHE DI RICERCA E SVILUPPO

Le principali problematiche di Ricerca e Sviluppo che dovranno essere affrontate, possono essere elencate in maniera sommaria come segue:

#### **problematiche di natura tecnico-scientifica**

- precisione della localizzazione all'interno di una struttura urbana complessa, che risulta particolarmente sfidante e scientificamente rilevante nel contesto urbano di Venezia, con strette strade (calli) ed alti edifici, dove i segnali (sia satellitari che *wireless*) faticano a propagarsi correttamente
- acquisizione ed organizzazione di dati disomogenei strutturati/non strutturati provenienti da fonti diverse (*data fusion*)
- armonizzazione di dati strutturati (p.e. dati provenienti da sistemi diversi (trip advisor, google place, EuroMuse)) ma in grado di fornire informazioni sovrapponibili (per esempio informazioni sugli hotel o sui musei)
- *harvesting*<sup>41</sup> e analisi di dati non strutturati, raccolti dalla *collective intelligence* (in particolare commenti e opinioni dai *social network*)
- studio di uno strumento per la “cattura” dei comportamenti e delle azioni dell'utente, su base geografica e temporale
- caratterizzazione semantica dei dati di contesto, studio di una ontologia di dominio e impiego di thesauri e vocabolari

#### **problematiche di natura tecnologica**

- impiego di architetture di tipo NoSQL per la memorizzazione di grandi moli di dati secondo modelli di scalabilità orizzontale
- impiego di tecniche statistiche di *data summarization* e database temporali probabilistici per gestire la enorme quantità di dati ridondanti generata
- studio e simulazione delle performance del sistema, con lo scopo di garantire l'erogazione di un servizio in grado di soddisfare i requisiti richiesti

---

<sup>41</sup> raccolta di dati utilizzando procedure software

- applicabilità di interfacce multimodali per una migliore e più completa interazione tra uomo e macchina; tra le interfacce di sicuro interesse potrebbero esserne studiate alcune specifiche per non vedenti, in grado di indirizzarli verso luoghi di interesse mediante comunicazioni sonore/audio.

### 6.1.3 LUOGO DI SVOLGIMENTO DEL PROGETTO

Il progetto, pur presentando caratteristiche generalmente applicabili, sarà realizzato in prima battuta nella città di Venezia, con lo scopo di verificare i risultati della ricerca e di attivare il servizio a partire dalla scala locale. Questa scelta è motivata da alcune considerazioni:

- si vuole studiare, come detto, il modello per applicazioni a livello locale;
- si desidera coinvolgere alcuni operatori locali che possano garantire, a regime, lo sviluppo industriale del progetto.
- Venezia può rappresentare un banco di prova eccellente per l'impianto del sistema proposto, infatti:
  - possiede una spiccata vocazione turistica;
  - dispone di un patrimonio culturale vastissimo e unico al mondo;
  - ospita una infrastruttura tecnologica di prim'ordine (rete in fibra, wi-fi pubblico, servizio *Venice Connected*).

Ciò consente di ipotizzare una effettiva buona riuscita del progetto e la fornitura di un servizio ad elevato valore aggiunto per il turista. Questa tesi è avvalorata dal fatto che l'iniziativa viene sostenuta da alcuni operatori locali e in particolare dalla società che ha in gestione l'infrastruttura tecnologica della città.

#### 6.1.4 OBIETTIVI SPECIFICI, ATTIVITÀ E TEMPISTICHE

Riprendendo in esame lo schema della “Figura 29” presentato in precedenza, viene ulteriormente dettagliato il processo che porterà alla erogazione del servizio proposto.

Il dispositivo dell'utente trasmette alcuni dati di contesto, e più in generale dati utili per la profilazione dell'utente, al *Engine*. Questi dati sono generati dai comportamenti dell'utente e possono comprendere una raccolta dati effettuata con modalità da individuare (p.e. registrandosi al portale di *Venice Connected*). La profilazione dei comportamenti dell'utente è una delle attività più delicate oggetto della ricerca e potrà comprendere, a puro titolo di esempio: come e con quale frequenza viene utilizzato lo *smartphone*, come e dove si sposta, orari dedicati a svolgere particolari attività etc.. In questo modo viene realizzata la contestualizzazione dell'utente dal punto di vista comportamentale. Tale contestualizzazione non richiede che le elaborazioni siano effettuate in tempo reale.

Oltre ai dati di contesto rilevati attraverso il dispositivo, all'interno del *Engine* confluiscono dati e informazioni di diversa natura e provenienza:

- dati da sensori distribuiti sul territorio – che all'interno del grande tema della *Smart City* costituiscono la cosiddetta *Internet of Things* e che possono essere per esempio webcam e sensori ambientali; in questo caso andranno valutate approfonditamente problematiche etiche e di *privacy*;
- dati forniti da sistemi locali – poiché il tema è quello del turismo culturale ci si riferisce certamente alle banche dati dei musei, alle comunicazioni della società Venezia Marketing & Eventi e della Biennale ma anche a dati meteo locali resi disponibili dal Centro di Segnalazione e Previsione delle Maree, dati su cantieri e interruzioni stradali messi a disposizione principalmente dal Comune di Venezia e dalle società Insula Spa e da Veritas Spa, dati sui trasporti pubblici messi a disposizione da ACTV Spa, etc...
- altri dati provenienti dai grandi provider e recuperabili tramite API<sup>42</sup> e Web

---

<sup>42</sup> Col termine API (*Application Programming Interface*) si indica un insieme di funzionalità software messe a disposizione da un programma per essere utilizzate da altri programmi

Services<sup>43</sup> – includendo in questa categoria anche informazioni provenienti dai *social network* (commenti, opinioni, suggerimenti etc..) e più in generale dalla cosiddetta *collective intelligence*.

- Il *Engine* acquisisce tutti questi dati, li aggrega, li omogeneizza e li organizza facendoli confluire in una struttura comune di archiviazione, modellata attraverso una ontologia di dominio, nel nostro caso relativa all'ambito del turismo culturale. L'ontologia di dominio ha lo scopo fondamentale di fornire una base di omogeneizzazione dei dati raccolti fornendo nel contempo una struttura semantica di riferimento. I dati così trattati a questo punto sono pronti per essere elaborati in relazione ai dati di contesto e ai dati di profilo utente, attraverso il modulo *Profiler*.

I dati di localizzazione dell'utente sono inviati direttamente al *Profiler* con lo scopo di realizzare, in tempo reale, la contestualizzazione dell'utente dal punto di vista geografico e temporale. La localizzazione dell'utente, che in talune zone di Venezia può risultare particolarmente difficile, può essere realizzata attraverso l'impiego di sistemi ibridi, capaci di ottimizzare in base al contesto d'uso e allo stato del dispositivo (p.e. stato di carica della batteria) diverse tecnologie di localizzazione (wifi, cella telefonica, Gps). Un tema importante di ricerca in questo caso riguarda la definizione di una metodologia di tracciamento che non comporti oneri eccessivi per l'utilizzatore, sia in termini di energia spesa dal dispositivo sia in termini di traffico dati (e quindi di costi) generato, ma che consenta di ottenere una precisione tale da garantire la corretta erogazione dei servizi.

I dati che raggiungono il *Profiler*, infine, sono elaborati con lo scopo di produrre le raccomandazioni e i suggerimenti da esporre in maniera asincrona all'utente attraverso un'interfaccia multimodale.

---

43 Secondo la definizione data dal W3C un *Web Service* è un sistema software progettato per supportare l'interoperabilità tra diversi elaboratori su di una medesima rete

6.1.4.1 *OBIETTIVI REALIZZATIVI DEL PROGETTO*

Definito dunque il processo in grado di supportare il servizio in questione, possiamo passare a descrivere gli Obiettivi Realizzativi (OR) del progetto. Ciascun OR è contraddistinto da una serie di attività che dovranno essere realizzate con lo scopo di raggiungere tali obiettivi e da una descrizione dei risultati attesi. Per ciascun OR è inoltre indicato il tipo di attività prevalente (Ricerca Industriale (RI) ovvero Sviluppo Precompetitivo (SP)) e una stima delle risorse richieste in termini di mesi/persona.

<b>n.</b>	<b>Titolo dell'OR</b>	<b>Tipologia di attività</b>	<b>Tempo totale previsto (mesi/persona)</b>
1	Acquisizione delle Informazioni di Localizzazione	Ricerca Industriale	6
2	Acquisizione delle risorse del web		
2.1	Acquisizione di risorse web disponibili sia da sistemi locali sia da provider esistenti	Ricerca Industriale	6
2.2	Acquisizione di risorse disponibili all'interno del cosiddetto "Internet of Things"	Ricerca Industriale	6
2.3	Acquisizione di risorse provenienti dalla cosiddetta "Collective Intelligence"	Ricerca Industriale	6
3	Archiviazione, organizzazione e caratterizzazione delle informazioni di contesto	Ricerca Industriale	18
4	Profilazione dell'utente, Raccomandazioni	Ricerca Industriale	12
5	Realizzazione di un prototipo applicativo	Sviluppo precompetitivo	18
6	Studio e progettazione di una applicazione su scala locale a Venezia		

6.1	Studio metodologico per l'applicazione su scala locale del sistema	Ricerca Industriale	4
6.2	Contatto con gli operatori turistici e culturali e coi fornitori di dati locali	Sviluppo precompetitivo	4
6.3	Meccanismi di affiliazione e di fidelizzazione dell'utente	Ricerca Industriale	4
6.4	Studio delle interfacce	Ricerca Industriale	10
6.5	Sviluppo di un prototipo del sistema	Sviluppo precompetitivo	12
6.6	Testing e revisione del sistema	Sviluppo precompetitivo	4

Complessivamente 72 mesi uomo di ricerca industriale e 38 mesi uomo di sviluppo precompetitivo

#### 6.1.4.2 DESCRIZIONE DEGLI OBIETTIVI REALIZZATIVI

<b>Titolo dell'OR 1</b>	Acquisizione delle informazioni di localizzazione	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	6
<b>Obiettivi</b>	Studiare le diverse modalità di localizzazione di un utente di smartphone all'interno di un'area urbana e in particolare all'interno del centro storico. Definire le migliori strategie di risparmio energetico degli apparati. Valutare le nuove opportunità tecnologiche offerte dalla futura attivazione della rete europea di posizionamento "Galileo"				
<b>Attività</b>	Ricognizione dello stato dell'arte dei sistemi di posizionamento ibridi per smartphone. Questa attività dovrà evidenziare quali siano le tecnologie all'avanguardia in grado di garantire la massima precisione di localizzazione e il minimo consumo energetico. Nello specifico caso di Venezia, poiché la città è dotata di una infrastruttura a larga banda con rete WiFi, sarà studiata la possibilità di utilizzare tale rete (ed				

	<p>eventualmente suggerirne il potenziamento) per far fronte alle necessità di precisione nel posizionamento degli utenti in un'area urbana complessa come quella veneziana. Saranno studiate le modalità di integrazione con sistemi di localizzazione ibridi esistenti. Saranno valutate le potenzialità e le opportunità di integrazione offerte dal nuovo sistema di posizionamento europeo “Galileo” del quale è prevista la messa in funzione nei prossimi anni.</p>
<b>Risultati attesi</b>	<p>Definire le migliori strategie operative per la localizzazione di un utente smartphone all'interno di un centro storico con lo scopo di massimizzare la precisione di localizzazione e di minimizzare il consumo energetico degli apparati.</p>

<b>Titolo dell'OR 2</b>	Acquisizione delle risorse del Web	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	
<b>Obiettivi</b>	<p>Studiare le modalità più efficienti per l'acquisizione delle risorse web necessarie a contestualizzare l'utente rispetto all'ambiente in cui si trova.</p>				
<b>Attività</b>	<p>L'obiettivo realizzativo n. 2 sarà perseguito attraverso la realizzazione di tre attività descritte in altrettanti sotto obiettivi realizzativi:</p> <p><i>OR 2.1: acquisizione di risorse web disponibili sia da sistemi locali sia da provider esistenti</i></p> <p><i>OR 2.2: acquisizione di risorse disponibili all'interno del cosiddetto “Internet of Things”</i></p> <p><i>OR 2.3: acquisizione di risorse provenienti dalla cosiddetta “Collective Intelligence”</i></p>				
<b>Risultati attesi</b>	<p>Rassegna delle risorse disponibili nel web per l'alimentazione del sistema e definizione di una serie di procedure per l'utilizzazione di tali risorse con lo scopo di contestualizzare l'utente rispetto all'ambiente in cui si trova.</p>				

<b>Titolo dell'OR 2.1</b>	Acquisizione di risorse web disponibili sia da sistemi locali sia da provider esistenti	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	6
<b>Obiettivi</b>	Studio delle modalità più efficienti per l'acquisizione di risorse web simili ma strutturate in modo disomogeneo e provenienti da fonti locali o da data provider.				
<b>Attività</b>	<p>Sarà condotta una approfondita rassegna di possibili fornitori sia per quanto riguarda risorse locali (p.e. Venezia marketing &amp; eventi, musei, biennale per gli eventi culturali, ACTV per i trasporti, Centro previsione e segnalazione maree per il meteo locale e le previsioni di alta marea etc..) sia per quanto riguarda <i>data provider</i> (p.e. Google Places, Trip Advisor, Eventbrite API).</p> <p>Saranno definite le procedure di acquisizione e di integrazione di tali risorse che in linea di massima si possono considerare tra loro disomogenee ovvero strutturate in maniera non omogenea (p.e. più risorse possono fornire la stessa informazione ovvero informazioni simili ma strutturate in maniera diversa con un maggiore o minore grado di dettaglio).</p> <p>Sarà studiata la fattibilità di una applicazione in grado di realizzare il servizio di <i>harvesting</i> sulle risorse individuate e di <i>mash-up</i> in grado di includere dinamicamente informazioni o contenuti provenienti da più fonti.</p>				
<b>Risultati attesi</b>	Definire le migliori strategie operative e procedurali per l'acquisizione di risorse disomogenee dal web.				

<b>Titolo dell'OR 2.2</b>	Acquisizione di risorse disponibili all'interno del	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	6
---------------------------	---	------------------------------	---------------------	---	---

	cosiddetto “Internet of Things”				
<b>Obiettivi</b>	Studio delle modalità più efficienti per l'acquisizione di risorse web provenienti da sensori distribuiti sul territorio.				
<b>Attività</b>	Sarà condotta una approfondita rassegna della sensoristica presente nell'area veneziana e di potenziali fornitori di dati da sensori. Saranno definite le procedure di acquisizione e di integrazione di tali risorse che in alcuni casi potranno essere disomogenee (p.e. due sensori di temperatura potranno fornire l'informazione in °C o °F). Sarà studiata la fattibilità di una applicazione in grado di realizzare il servizio di <i>harvesting</i> sulle risorse individuate.				
<b>Risultati attesi</b>	Definire le migliori strategie operative e procedurali per l'acquisizione di risorse disomogenee provenienti da sensori distribuiti sul territorio.				

<b>Titolo dell'OR 2.3</b>	Acquisizione di risorse provenienti dalla cosiddetta “Collective Intelligence”	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	6
<b>Obiettivi</b>	Studio delle modalità più efficienti per analizzare la cosiddetta collective intelligence e derivarne contenuti utili a completare la contestualizzazione dell'utente.				
<b>Attività</b>	Ricognizione delle possibili risorse web da utilizzare come fonti della cosiddetta collective intelligence (intelligenza condivisa derivante dalla collaborazione e dalla competizione dei singoli individui). In linea di massima tali fonti sono individuabili nei <i>social network</i> (Twitter, Facebook, ...) attraverso i quali gli utenti commentano e si scambiano informazioni su eventi, prodotti, servizi etc.. Saranno definite le procedure di acquisizione e di integrazione di tali risorse, sempre tenendo in dovuta considerazione le problematiche e le				

	leggi sulla privacy. Sarà studiata la fattibilità di una applicazione in grado di realizzare il servizio di <i>harvesting</i> sulla <i>collective intelligence</i> .
<b>Risultati attesi</b>	Definire le migliori strategie operative e procedurali per l'acquisizione di risorse dalla <i>collective intelligence</i> .

<b>Titolo dell'OR 3</b>	Archiviazione, organizzazione e caratterizzazione delle informazioni di contesto	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	18
<b>Obiettivi</b>	Studio di un sistema per l'archiviazione delle informazioni acquisite e per la loro elaborazione con lo scopo di utilizzarle nella contestualizzazione dell'utente.				
<b>Attività</b>	Sarà studiata la soluzione più efficiente per l'archiviazione e l'organizzazione delle risorse acquisite. Saranno valutate le diverse tecnologie di data base e modelli dati disponibili, anche di tipo non relazionale (NoSql Data Base) e definite le architetture a supporto del sistema includendo anche servizi di <i>Cloud Computing</i> . Le risorse acquisite saranno armonizzate e caratterizzate dal punto di vista semantico e modellate attraverso una ontologia di dominio (turismo culturale) che sarà studiata e realizzata allo scopo. Tale modello terrà conto dello stato dell'arte delle ontologie e dei thesauri disponibili nel contesto dei linked data.				
<b>Risultati attesi</b>	Definire il modello di archiviazione e di elaborazione delle informazioni e definire il sistema di archiviazione delle informazioni. (Modulo <i>Engine</i> nel modello concettuale)				

<b>Titolo</b>	Profilazione	<b>Tipologia</b>	Ricerca	<b>Tempo totale</b>	12
---------------	--------------	------------------	---------	---------------------	----

<b>dell'OR 4</b>	dell'utente, Raccomandazioni	<b>di attività</b>	Industriale	<b>previsto</b> <b>(mesi/persona)</b>	
<b>Obiettivi</b>	Studio di un modello di profilazione dell'utente e del motore delle raccomandazioni.				
<b>Attività</b>	<p>Studio delle modalità di raccolta di informazioni per la profilazione dell'utente: queste in linea di massima possono essere raccolte attraverso dei processi di registrazione (p.e. per poter accedere alla rete WiFi di Venezia), oppure raccolte analizzando i comportamenti e le abitudini dell'utilizzatore (p.e. tra le 7 e le 8 di mattina vado a correre al parco) oppure ancora, raccolte interpellando l'utente in situazioni particolari (p.e. non ti piacerebbe ascoltare un concerto jazz?).</p> <p>Studio di una procedura per la composizione del profilo di un utente.</p> <p>Studio di fattibilità di un modulo applicativo per la produzione di raccomandazioni personalizzate sia legate al profilo personale dell'utente (gusti, preferenze, abitudini...), sia legate al contesto in cui l'applicazione viene utilizzata (ora, luogo, condizioni meteo, ...).</p>				
<b>Risultati attesi</b>	Definire il modulo <i>Profiler</i> del modello concettuale.				

<b>Titolo dell'OR 5</b>	Sviluppo di un prototipo applicativo	<b>Tipologia di attività</b>	Sviluppo precompetitivo	<b>Tempo totale previsto</b> <b>(mesi/persona)</b>	18
<b>Obiettivi</b>	Realizzare un prototipo che consenta di realizzare il testing dei risultati di OR3-OR4				
<b>Attività</b>	<p>Realizzazione di un prototipo di <i>Engine</i> da utilizzare con risorse e fonti dati predefinite.</p> <p>Realizzazione di un prototipo di <i>Profiler</i> da utilizzare con dati di test.</p>				
<b>Risultati attesi</b>	<i>Testing</i> della applicabilità degli studi effettuati.				

<b>Titolo dell'OR 6</b>	Studio e progettazione di una applicazione su scala locale a Venezia	<b>Tipologia di attività</b>	Ricerca Industriale e sviluppo precompetitivo	<b>Tempo totale previsto (mesi/persona)</b>	
<b>Obiettivi</b>	Studio e progettazione di una applicazione su scala locale a Venezia				
<b>Attività</b>	L'obiettivo realizzativo n. 6 sarà perseguito attraverso la realizzazione di sei attività descritte in altrettanti sotto obiettivi realizzativi: <i>OR 6.1 Studio metodologico per l'applicazione su scala locale del sistema</i> <i>OR 6.2 Contatto con gli operatori turistici e culturali e coi fornitori di dati locali</i> <i>OR 6.3 Meccanismi di affiliazione e di fidelizzazione dell'utente</i> <i>OR 6.4 Studio delle interfacce</i> <i>OR 6.5 Sviluppo di un prototipo del sistema</i> <i>OR 6.6 Testing e revisione del sistema</i>				
<b>Risultati attesi</b>	Si vedano i singoli OR 6.1-6.6				

<b>Titolo dell'OR 6.1</b>	Studio metodologico per l'applicazione su scala locale del sistema	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	4
<b>Obiettivi</b>	Definire di una metodologia generale per l'applicazione del sistema su scala locale				
<b>Attività</b>	Definizione della metodologia generale. Raccolta requisiti specifici del caso di studio veneziano. Applicazione al caso di studio veneziano.				
<b>Risultati attesi</b>	Redazione di una metodologia generale per l'applicazione del sistema su scala locale				

<b>Titolo</b>	Contatto con gli	<b>Tipologia</b>	Sviluppo	<b>Tempo totale</b>	4
---------------	------------------	------------------	----------	---------------------	---

<b>dell'OR 6.2</b>	operatori turistici e culturali e coi fornitori di dati locali	<b>di attività</b>	precompetitivo	<b>previsto (mesi/persona)</b>	
<b>Obiettivi</b>	Coinvolgere gli operatori turistici e culturali				
<b>Attività</b>	Coinvolgimento degli operatori turistici e culturali coi quali si dovrà collaborare sia nella fase di attivazione del sistema, sia in fase di manutenzione. In linea di massima il contatto avverrà attraverso associazioni di categoria e associazioni culturali. Saranno istituiti dei <i>focus group</i> con lo scopo di individuare le modalità più efficaci di presentazione dell'offerta attraverso il sistema delle raccomandazioni.				
<b>Risultati attesi</b>	Coinvolgimento degli operatori e affinamento del sistema delle raccomandazioni				

<b>Titolo dell'OR 6.3</b>	Meccanismi di affiliazione e di fidelizzazione dell'utente	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	4
<b>Obiettivi</b>	Studiare le modalità di affiliazione e di fidelizzazione degli utenti				
<b>Attività</b>	Studio delle diverse modalità di affiliazione e di fidelizzazione degli utenti affinché siano incentivati alla installazione della applicazione e alla sua utilizzazione.  Studio di meccanismi di incentivazione e di sensibilizzazione all'ascolto della raccomandazione.				
<b>Risultati attesi</b>	Affiliazione e fidelizzazione dell'utente				

<b>Titolo dell'OR 6.4</b>	Studio delle interfacce	<b>Tipologia di attività</b>	Ricerca Industriale	<b>Tempo totale previsto (mesi/persona)</b>	10
<b>Obiettivi</b>	Studio delle interfacce				

<b>Attività</b>	Studio delle interfacce attraverso cui implementare il sistema di raccomandazione. In linea di massima saranno realizzate interfacce multimodali (video e voce) e le modalità di accesso alla applicazione ovvero di erogazione delle raccomandazioni sarà semplificata al massimo. L'idea che si vorrebbe realizzare è quella di uno <i>smartphone</i> “amico” che ogni tanto ti da qualche buon suggerimento, ma senza assillarti e possibilmente senza che tu debba premere pulsanti o fare qualsiasi tipo di operazione.
<b>Risultati attesi</b>	Definire le funzionalità e le caratteristiche di una interfaccia “ <i>very friendly</i> ”

<b>Titolo dell'OR 6.5</b>	Sviluppo di un prototipo del sistema	<b>Tipologia di attività</b>	Sviluppo precompetitivo	<b>Tempo totale previsto (mesi/persona)</b>	12
<b>Obiettivi</b>	Sviluppo di un prototipo completo di sistema				
<b>Attività</b>	Realizzazione di tutte le attività necessarie ad allestire un prototipo completamente funzionante nell'area veneziana, inclusa la implementazione dei meccanismi di affiliazione e di fidelizzazione e la realizzazione della interfaccia.				
<b>Risultati attesi</b>	Realizzare un prototipo di sistema completamente funzionante a Venezia				

<b>Titolo dell'OR 6.6</b>	Testing e revisione del sistema	<b>Tipologia di attività</b>	Sviluppo precompetitivo	<b>Tempo totale previsto (mesi/persona)</b>	4
<b>Obiettivi</b>	Verificare il buon funzionamento ed attuare i correttivi necessari				

<b>Attività</b>	Saranno realizzate tutte le attività necessarie a verificare il buon funzionamento del sistema attraverso una serie di test su: <ul style="list-style-type: none"><li>• affidabilità e precisione della localizzazione e durata della batteria su diversi apparati;</li><li>• verifica della bontà delle procedure di acquisizione e di armonizzazione dei dati (test di laboratorio);</li><li>• verifica della bontà delle procedure di caratterizzazione semantica (test di laboratorio);</li><li>• test sul campo del sistema.</li></ul>
<b>Risultati attesi</b>	Verifica dell'esito della ricerca

## 6.1.5 VERIFICA DELL'ESITO DEL PROGETTO DI RICERCA

### 6.1.5.1 RISULTATI DISPONIBILI A FINE ATTIVITÀ

I risultati disponibili a fine attività saranno riportati in una serie di documenti e relazioni che descriveranno i risultati ottenuti negli studi effettuati. Tali documenti includeranno informazioni di valore commerciale ed eventualmente brevettabili e porranno le basi operative per lo sviluppo dei prototipi. Oltre alla redazione dei documenti succitati, il progetto ha come obiettivo la realizzazione di due prototipi: il primo includerà i moduli del *Engine* e del *Profiler* e sarà utilizzabile con risorse e fonti dati predefiniti (definibile come “prototipo da laboratorio”); il secondo (definibile come “prototipo sul campo”), che includerà anche l'interfaccia multimodale “*very friendly*” e i metodi di fidelizzazione dell'utente, consentirà di effettuare un testing completo dei risultati della ricerca, su un campione significativo seppur limitato di utenti a Venezia. Lo scopo di questi prototipi è la verifica fattiva delle soluzioni proposte e l'identificazione dei loro limiti e delle potenzialità. Va evidenziato come il caso di studio individuato (la città di Venezia) presenti peculiarità uniche (per esempio, la criticità intrinseca dei metodi di localizzazione dei dispositivi mobili, o la grande varietà dei potenziali utenti) che renderanno il sistema, una volta ingegnerizzato, innovativo rispetto ai prodotti presenti sul mercato e potenzialmente di successo.

In linea di massima gli obiettivi della ricerca che saranno raggiunti al termine dei 30 mesi di progetto, sono quelli elencati nel capitolo dedicato alle “principali problematiche di Ricerca e Sviluppo” e più compiutamente descritti nel capitolo dedicato agli obiettivi realizzativi. Più in particolare saranno raggiunti e verificabili i seguenti obiettivi di studio:

- soluzioni per gestire problematiche di precisione della localizzazione all'interno di una struttura urbana complessa, tenendo in considerazione le opportunità tecnologiche offerte dalla futura attivazione della rete europea di posizionamento “Galileo”;

- procedure per l'acquisizione di risorse web disomogenee e per una loro organizzazione e armonizzazione con lo scopo di utilizzarle nella contestualizzazione dell'utente;
- strumenti per la cattura dei comportamenti e delle azioni dell'utente, su base geografica e temporale;
- metodi e procedure di caratterizzazione semantica dei dati di contesto, incluso lo studio di una ontologia di dominio relativa al turismo culturale.

Il raggiungimento degli obiettivi citati sarà verificabile già a valle dello sviluppo e del test del primo prototipo applicativo (OR5) il cui completamento è previsto dopo circa 20 mesi.

Altri obiettivi di studio saranno resi disponibili dopo il completamento del progetto e saranno verificabili attraverso il test del prototipo di sistema (OR6.6):

- architetture di tipo NoSQL per la memorizzazione di grandi moli di dati secondo modelli di scalabilità orizzontale;
- interfacce multimodali di tipo “*very friendly*”;
- affiliazione e fidelizzazione dell'utente;
- sistema delle raccomandazioni.

#### 6.1.5.2 MODALITÀ CON CUI SARÀ VERIFICABILE L'ESITO DELL'INTERA RICERCA

L'esito della ricerca sarà verificabile sul campo al termine dello sviluppo e nel corso delle attività di test del prototipo di sistema. La valutazione avverrà utilizzando criteri quantitativi e oggettivi di raggiungimento degli obiettivi.

Saranno realizzati i seguenti test:

1. **localizzazione** – sarà valutata sia in termini di accuratezza sia in termini di consumi.

- *Accuratezza* – saranno effettuati test con diversi tipi di apparati verificando il grado di accuratezza del posizionamento lungo percorsi predefiniti di cui siano noti i parametri di copertura del segnale GPS, rete wifi etc. I percorsi saranno definiti con lo scopo di verificare il comportamento degli apparati anche all'interno di aree particolarmente critiche.
  - *Consumi* – saranno valutati i consumi medi di diversi apparati in diverse situazioni anche critiche con l'obiettivo di individuare il giusto bilanciamento tra l'esigenza di accuratezza del tracciamento e l'esigenza di limitare al massimo i consumi dell'apparato.
2. **acquisizione e di armonizzazione dei dati** – sarà verificata la bontà delle procedure di acquisizione e di armonizzazione dei dati raccolti, effettuando dei test comparativi su dati preimpostati, valutando le risposte del sistema rispetto ai risultati attesi.
  3. **caratterizzazione semantica dei dati** – sarà verificata la bontà delle procedure di caratterizzazione semantica dei dati raccolti, effettuando dei test comparativi su dati preimpostati, valutando le risposte del sistema rispetto ai risultati attesi.
  4. **test del sistema di raccomandazione e dell'intero sistema** – sarà effettuato attraverso una verifica diretta sul campo così concepita: un gruppo di volontari (di diversa estrazione sociale, cultura e lingua) sarà fornito di smartphone di almeno due diverse tipologie e con almeno due diversi sistemi operativi. I volontari effettueranno la normale procedura di *startup* come avverrebbe per un qualsiasi turista (scaricamento dell'App, accreditamento, registrazione, acquisto ticket etc..) e per una settimana intera utilizzeranno l'applicazione in assoluta libertà. Per la verifica dei risultati tutte le operazioni eseguite e le risposte del sistema saranno memorizzate in file *log* che potranno essere analizzati a valle del periodo di test. I volontari, al termine del test, saranno chiamati ad esprimere il proprio giudizio su: facilità di uso dell'applicazione, utilità dell'applicazione, pertinenza e utilità delle

raccomandazioni e quant'altro possa essere ritenuto utile affinché il sistema risponda effettivamente alle aspettative dell'utilizzatore.

## 6.1.6 INTERESSE TECNICO-SCIENTIFICO

### 6.1.6.1 NOVITÀ E ORIGINALITÀ DELLE CONOSCENZE ACQUISIBILI

Nel seguito sono descritte alcune delle attuali tecnologie e soluzioni confrontabili con quella proposta, sviluppate e utilizzate a livello internazionale. A tal proposito è importante sottolineare che tali soluzioni, rispetto alle quali il progetto *Recommending to People* potrebbe apparire in competizione, sono viceversa da considerare come opportunità in grado di fornire “valore aggiunto” a questo stesso progetto. Questo perché i dati raccolti dai grandi *provider* internazionali serviranno essenzialmente a completare il processo di contestualizzazione del turista culturale, che sarà basato per buona parte su fonti di dati locali. Infatti, la scelta di realizzare il progetto a partire dalla scala locale ha come obiettivo anche quello di evitare la competizione coi grandi *provider* internazionali rispetto ai quali non potrebbe di fatto esserci alcuna competizione, per lo meno nella fase di start-up del progetto.

**Tripadvisor** è uno dei maggiori siti di viaggi *online*. Ha una banca dati (*business*) con commenti di inestimabile valore, che senz'altro sarà utilizzata in questo progetto attraverso le API messe a disposizione dallo stesso Tripadvisor. Aiuta i turisti di tutto il mondo a programmare la propria vacanza grazie a recensioni e opinioni, scritte dagli stessi utenti, relative ad alloggi, voli e ristoranti. Le informazioni riguardanti hotel, piuttosto che bar e ristoranti, vengono fornite insieme ai commenti e alle valutazioni dei clienti del servizio, facendo sì che ne derivi una descrizione accurata, con pregi e difetti.

**GooglePlaces** ha la più vasta banca dati di *business* geolocalizzati, che pure utilizzeremo per questo progetto. Google Places è uno strumento pubblicitario

gratuito che consente ai titolari di attività commerciali di aggiornare e gestire direttamente informazioni relative alle proprie attività commerciali. Queste informazioni possono essere utilizzate per aggiornare la propria scheda di attività in modo che sia visualizzata correttamente in Google Maps e in altri strumenti messi a disposizione da Google.

**Siri** è un sistema avanzato, basato sul riconoscimento vocale, integrato da Apple negli iPhone di ultima generazione. Si tratta di una applicazione complessa che negli USA utilizza circa 1800 *data provider*. Possiede alcune somiglianze con questo nostro progetto ma non è un *Recommender system*. E' in grado di imparare archiviando ed elaborando le domande e le risposte degli utenti e ha una discreta capacità di contestualizzare le richieste dell'utente grazie all'impiego di una base semantica. Ciononostante la proposta presentata in questo progetto include anche dati da sensori distribuiti sul territorio ed è calata sulla realtà di Venezia che, come detto in precedenza, presenta peculiarità uniche e può essere presa a modello della città storica italiana.

**Facebook Graph Search** è un motore di ricerca semantico che è stato introdotto da Facebook nel marzo 2013. È progettato per dare risposte alle domande degli utenti sotto forma di linguaggio naturale, piuttosto che come lista di link. La funzione *Graph Search* combina i dati acquisiti dagli utenti di Facebook e i dati esterni in un unico motore di ricerca che fornisce risultati di ricerca specifici dell'utente, recuperando le informazioni all'interno di una rete di utenti amici.

**Foursquare** è un *social network* basato sulla geolocalizzazione degli utenti. Ha come scopo la collezione di *check-in* in luoghi condivisi o meno. Il *check-in* avviene usando l'applicazione installata sul proprio smartphone e può consentire di accumulare punti o buoni sconto. Nella ultima versione è sostanzialmente un *Recommender System*, ma basato su una comunità 2.0 dove la partecipazione è esplicita. Per gli scopi di questo

progetto potrà costituire una fonte molto interessante di informazioni.

**Suggesto for travel** si tratta di un prodotto italiano, realizzato dalla società eCTRLSolution di Trento anche in versione *mobile*, in grado di aiutare l'utente nella pianificazione del proprio viaggio. E' un sistema di raccomandazione ma che parte da presupposti diversi rispetto a quelli ipotizzati nel progetto *Recommending to People* e non è specificatamente studiato per operare in tempo reale attraverso il tracciamento dell'apparato. *Suggesto for Travel* è stato utilizzato nel progetto OmniTurist.

#### 6.1.6.2 UTILITÀ DELLE CONOSCENZE ACQUISIBILI

In questo capitolo si cerca di evidenziare l'utilità delle conoscenze acquisibili con lo sviluppo del progetto, per innovazioni di prodotto/processo/servizio che accrescano la competitività e favoriscano lo sviluppo della nel settore di riferimento.

Joseph Schumpeter, economista austriaco della prima metà del '900, nell'opera *The theory of economic development* ha definito l'innovazione di prodotto come:

*“The introduction of a new good - that is, one with which consumers are not yet familiar - or a new quality of a good”.*

Nel progetto in questione, questa affermazione è senza dubbio perfettamente attinente: non esiste, allo stato attuale, un prodotto con caratteristiche paragonabili a quelle che vengono proposte<sup>44</sup>. Si possono trovare, per la verità, molti prodotti in grado di realizzare alcune delle cose proposte (in particolare tra quelli citati nel precedente paragrafo), ma nessuno in grado di realizzare un servizio complessivamente riconducibile a quello immaginato con questo progetto.

Quasi tutti i servizi attuali - mai concepiti per essere utilizzati su scala locale - cercano normalmente di sviluppare due aspetti:

- l'aspetto sociale: contatto gli amici (amici della rete) che mi stanno intorno,

---

<sup>44</sup> Per la verità ciò poteva essere vero al tempo della presentazione della domanda: successivamente questa stessa idea è stata recuperata per nuove proposte (i.e. il progetto Mozia presentato assieme alla Università di Palermo o il progetto OmniTurist attualmente in fase di completamento)

scambio informazioni, ascolto i loro suggerimenti, ne fornisco io a loro;

- l'aspetto pubblicitario: raccomando la visita di locali, fornisco incentivi e buoni sconto etc.

Questo progetto viceversa parte da un presupposto diverso: fornire all'utente un vero servizio di raccomandazione che possa includere molti altri aspetti oltre a quelli *social* e di pubblicità. Il *target* infatti è quello del turista culturale e in generale del turista che quando visita una città è senz'altro interessato a molte altre cose oltre che a mangiarsi una pizza: se è una bella giornata verso sera e sono nella parte ovest della città, perché non vedere un bel tramonto sulla laguna? Se sono le tre del pomeriggio e sono nei pressi di piazza San Marco, meglio non visitare il museo Correr per il quale serve almeno una giornata a disposizione, preferirò essere indirizzato verso una attività meno impegnativa; se sono allergico al polline del tiglio ed è primavera, eviterò la zona dei giardini; se sono interessato all'arte contemporanea vorrei essere indirizzato di conseguenza. Su questi aspetti si basa quella che noi riteniamo essere la vera “innovazione di servizio” che oltretutto consentirà lo sviluppo della cosiddetta “coda lunga”: Venezia potrà vendere una gamma molto più ampia di prodotti/servizi rispetto a prima, in particolare sviluppando la visita delle zone più periferiche della città o per lo meno dei percorsi meno tradizionali.

Per ottenere questi risultati e per poter fornire un servizio articolato e di elevata qualità come quello immaginato, sarà necessario provvedere un sistema avanzato sia di profilazione dell'utente, sia di caratterizzazione semantica del contesto, sia di raccomandazione, ed è su questi tre aspetti che si gioca la vera “innovazione di prodotto” che potrebbe portare anche alla realizzazione di nuovi brevetti.

## 6.1.7 VALIDITÀ INDUSTRIALE DEL PROGETTO

### 6.1.7.1 COERENZA STRATEGICA E GESTIONE DEL PROGETTO

Per quanto riguarda lo sviluppo della società, una volta che il progetto sia entrato in produzione, e la coerenza delle attività sviluppate con gli obiettivi strategici dell'impresa, possono essere fatte alcune considerazioni.

La società nasce con l'obiettivo di sviluppare servizi di raccomandazione per il turismo culturale, così come sono stati descritti negli obiettivi di progetto. Tali servizi ne costituiranno il *business* principale.

Parallelamente a questi servizi tuttavia, potranno essere sviluppate anche altre opportunità, che in buona parte sono strettamente correlate o comunque riconducibili al *business* principale:

1. servizi per le istituzioni tesi a diffondere informazioni su eventi che le istituzioni stesse sono interessate a pubblicizzare (p.e. la biennale del cinema, i seminari delle Università etc..) ovvero servizi di raccolta di informazioni (traffico pedonale, distribuzione dei turisti e dei flussi turistici in città etc.);
2. servizi B2B, da sviluppare su richiesta, in occasione di fiere ed eventi organizzati da privati;
3. altri servizi legati al mondo del web e dello sviluppo di software e di applicazioni;
4. attività di ricerca su temi di interesse, sviluppati anche accedendo a finanziamenti comunitari, col duplice scopo sia di mantenere il sistema e l'impresa all'avanguardia, sia di conservare un contatto stretto col mondo dell'Università.

Più in particolare si ritiene possibile e auspicabile promuovere, fin dal II anno di attività, una serie di servizi tesi inizialmente a contribuire all'autofinanziamento della società ma che hanno lo scopo di cominciare a confrontare, fin dalle prime battute, la

società col mercato. In seguito potranno essere erogati alcuni servizi intermedi, principalmente nei confronti delle istituzioni cittadine interessate ad acquisire informazioni di cui al punto 2. del precedente elenco.

Dopo la conclusione del progetto e una prima fase di sperimentazione ed ingegnerizzazione dei prodotti, si potranno cominciare ad erogare i servizi di raccomandazione.

Naturalmente, è consigliabile e auspicabile cominciare presto a fare delle ipotesi di sviluppo dei servizi anche in altri contesti diversi da quello veneziano, in modo da trovarsi preparati ad affrontare una possibile, veloce crescita della società, nel momento in cui il servizio dovesse effettivamente soddisfare le aspettative.

#### 6.1.7.2 *COMPETITIVITÀ TECNOLOGICA*

Per quanto concerne gli aspetti di competitività tecnologica, come già accennato le prospettive di crescita appaiono interessanti:

- in primo luogo c'è da considerare che il mercato dei servizi *mobile* è in grande crescita e più nello specifico per il mondo *smartphone* si prevede una vera impennata nei prossimi anni;
- in secondo luogo questo progetto può tranquillamente essere inquadrato nell'ambito del vastissimo tema delle *smart-cities*, su cui attualmente c'è un grande fermento;
- infine le tecnologie di localizzazione satellitare sono in continua evoluzione e promettono, nel giro di pochissimi anni, un drastico miglioramento sia in termini di precisione, sia in termini di robustezza del segnale all'interno di un centro urbano.

Queste considerazioni fanno ritenere che nel breve-medio termine si assisterà ad un notevole ampliamento del mercato, indotto per lo più dalla innovazione tecnologica.

## 6.2 IL MODELLO ONTOLOGICO EVOCATIVO – SPERIMENTAZIONE

Questa sperimentazione vuol tentare di sviluppare la seguente idea:

- consideriamo uno *User* (per esempio un turista) alla visita di un Luogo (per esempio una città);
- supponiamo che lo *User* sia caratterizzato da un Profilo che lo descrive: per esempio, se si tratta di un turista, il profilo potrebbe riguardare i suoi principali interessi (l'arte piuttosto che le tradizioni o la gastronomia o il puro e semplice divertimento), il budget disponibile, le abitudini di viaggio e così via;
- supponiamo inoltre che della città siano state raccolte alcune “storie” relative ad *Item* di diversa natura che si trovano lungo un dato percorso (edifici, luoghi, persone, altri oggetti ...). Queste storie raccontano aspetti inusuali della città, che possono legarsi a leggende e a tradizioni e anche a persone che vivono o che sono vissute in quella città: storie, per intenderci, che potrebbe raccontarci chi è vissuto tutta una vita in quella città o in quel determinato quartiere;
- tutte queste storie conteranno esse stesse dei riferimenti a nuovi *Item* (ancora persone, luoghi, edifici etc..) che contribuiranno ad arricchire la storia originale lungo nuovi percorsi conoscitivi, mai predeterminati, il cui sviluppo potrebbe essere opportunamente legato a quelle che sono le aspettative dello *User* e le scelte da lui operate *in loco*, nel momento della visita; l'ampiezza e la qualità di tali percorsi diverrebbe indecidibile e potenzialmente senza soluzione.

Per realizzare questa idea si è proceduto realizzando i passi descritti nei prossimi capitoli.

## 6.2.1 COSTRUZIONE DEL MODELLO ONTOLOGICO EVOCATIVO “EvONTOLOGY”

Il primo passo ha riguardato la costruzione di una prima versione *beta* del Modello Ontologico Evocativo, denominato “EvOntology”.

EvOntology è stato realizzato in OWL, utilizzando l'editor di ontologie Protégé della Stanford University in California.

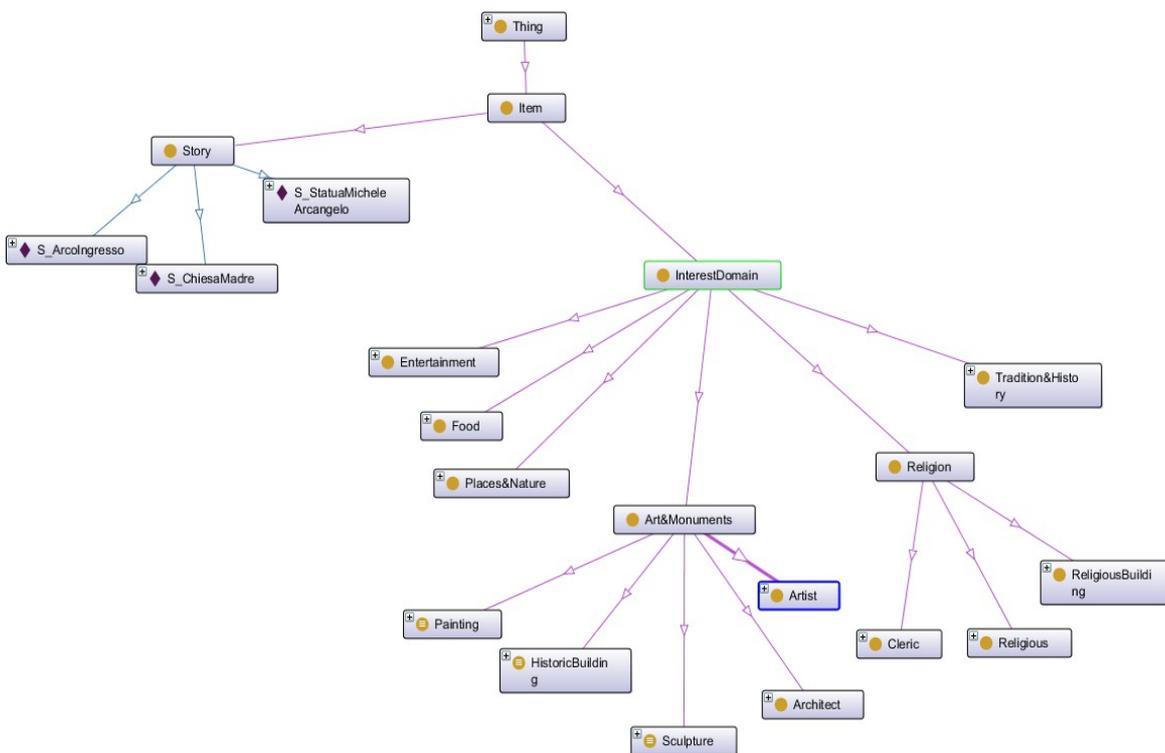


Figura 29: piccolo estratto della EvOntology integrata con Dbpedia Ontology e Foaf Ontology

Il modello è stato sviluppato in una prospettiva di interconnessione col *network* dei *linked data*; per questo motivo esso è stato integrato con la Dbpedia Ontology e la FOAF Ontology.

Considerata questa scelta, l'ontologia è stata sviluppata in questo modo:

- Il generico *Item* viene caratterizzato da una Storia (“Story”) e da un Dominio di Interesse (“InterestDomain”) su cui l'*Item* viene mappato con lo scopo di

poterlo poi riferire ad un profilo utente.

- Ciascun Dominio di Interesse – in questa versione *beta* ne sono stati individuati sei: *Entertainment*, *Food*, *Places&Nature*, *Art&Monuments*, *Religion*, *Tradition&History* - è a sua volta caratterizzato da una serie di sottoclassi che si riferiscono ad altrettante classi DBpedia o Foaf o anche appartenenti a Schema.org, quest'ultima richiamata da Dbpedia. Per esempio, la classe “Entertainment” di EvOntology comprende le sottoclassi “Event” da DBpedia, “Movie” da Schema.org.

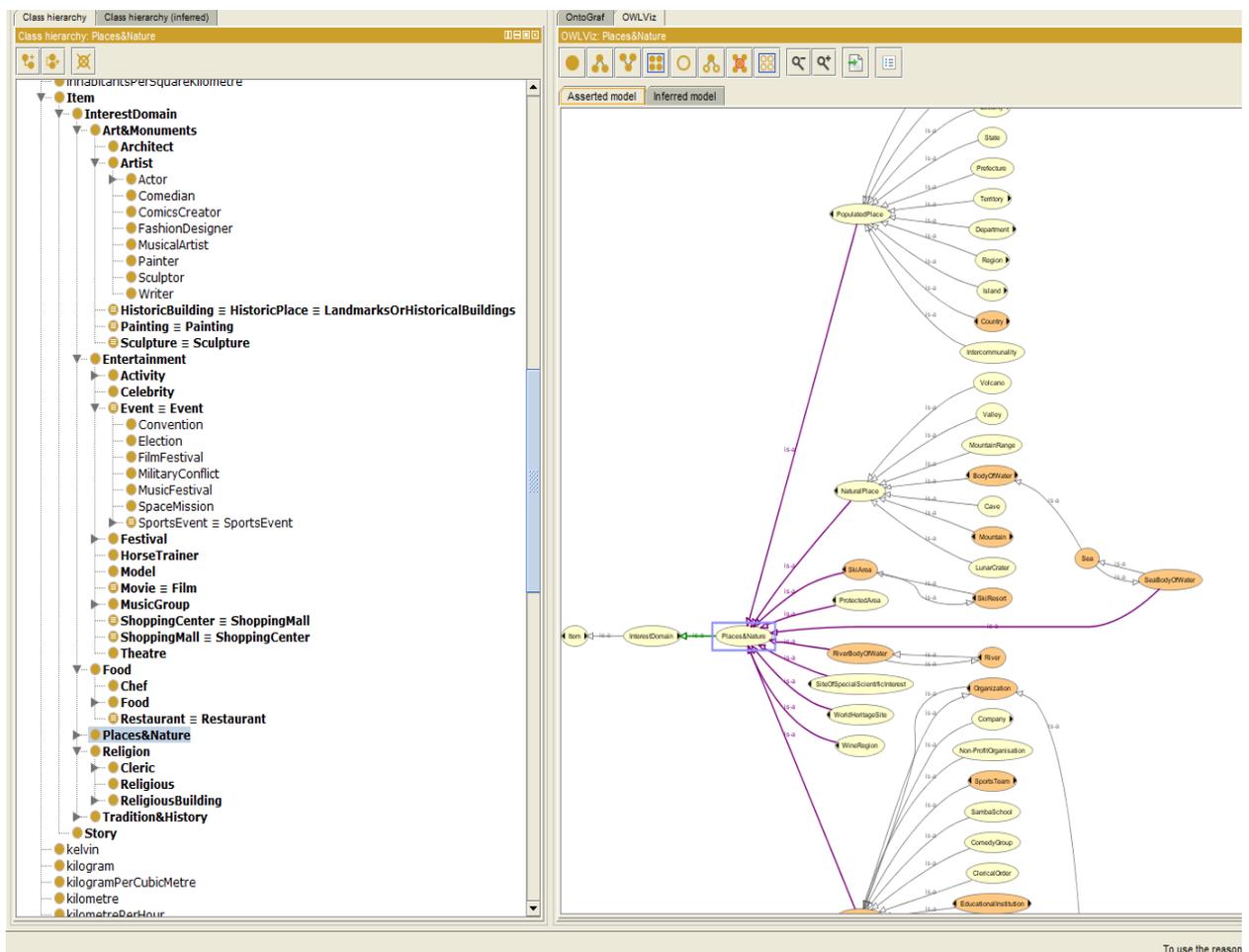


Figura 30: Sviluppo di alcune sottoclassi della EvOntology

## 6.2.2 ISTANZE E CARATTERIZZAZIONE DEGLI *ITEM* E GENERAZIONE DEL *DATA SET*

Su questo modello sono state realizzate le istanze di alcuni *Item* relativi alla città di Peschici in Puglia, sfruttando parte dei dati del progetto OmniTulist in corso di realizzazione in quella città.

Ciascuna di queste istanze fa riferimento ad una o più risorse “Storia” (istanze della classe “Story” sottoclasse di “Item”) che la descrive in termini evocativi.

Dalle storie associate agli *Item* (“ItemStory”) vengono quindi estrapolati dei riferimenti a nuovi *Item*, cercando di caratterizzare questi ultimi rispetto agli InterestDomain che abbiamo visto (arte, religione, tradizione, divertimento etc..) e che possono risultare altrettante aree di interesse per lo *User*, profilato in base a questi stessi InterestDomain.

Facciamo un esempio: l'istanza dell'arco di ingresso alla città di Peschici “ArcoIngresso” è caratterizzata dalla storia “S\_ArcoIngresso” (has\_itemstory S\_ArcoIngresso) di seguito riportata:

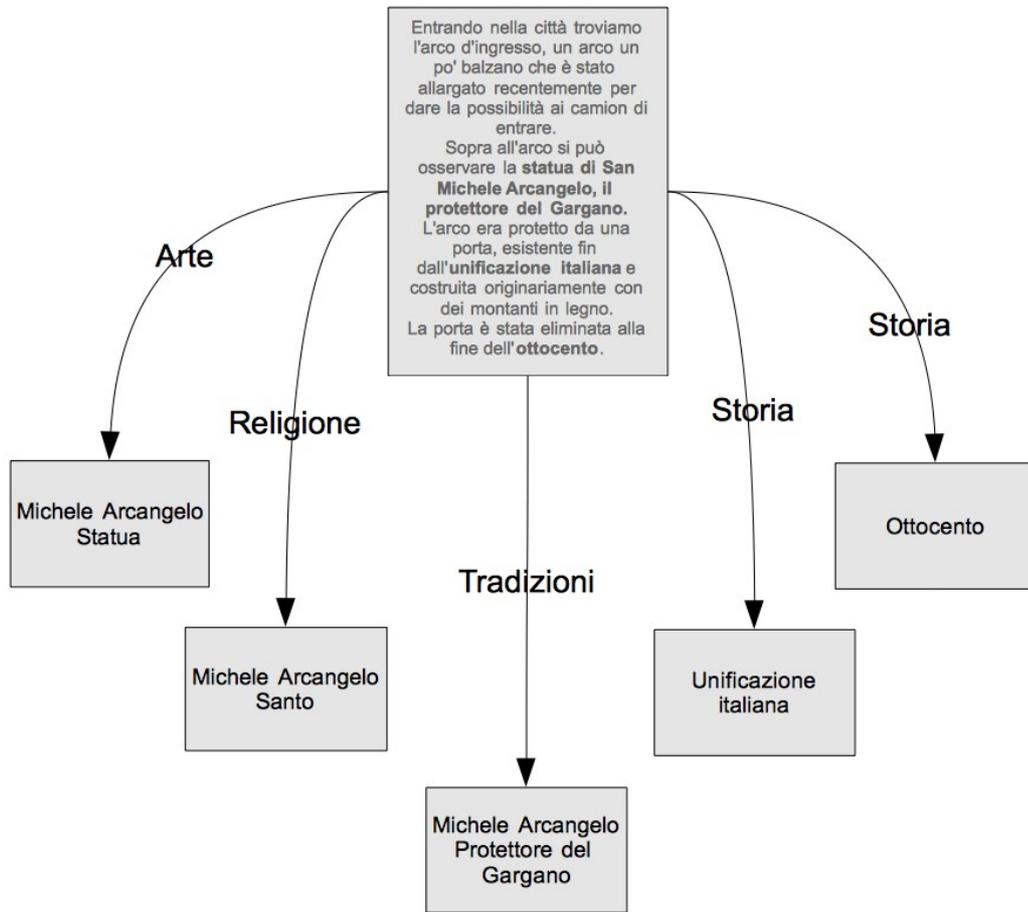
*Entrando nella città troviamo l'arco d'ingresso, un arco un po' balzano che è stato allargato recentemente per dare la possibilità ai camion di entrare.*

*Sopra all'arco si può osservare la **statua di San Michele Arcangelo, il protettore del Gargano.***

*L'arco era protetto da una porta, esistente fin dall'**unificazione italiana** e costruita originariamente con dei montanti in legno.*

*La porta è stata eliminata alla fine dell'**ottocento**.*

In questa storia possono essere individuati dei nuovi *Item*, caratterizzati da nuove storie, che possono riferirsi a diversi InterestDomain, come si può evincere dal grafico sotto riportato:



Ripercorrendo ora tutto il procedimento che porta alla costruzione del modello ontologico e aiutandoci con la “Figura 31” riportata più oltre per avere conferma dei diversi passaggi formali, abbiamo che (riferendosi comunque soltanto all'arco di ingresso):

- “ArcoIngresso” è caratterizzato dalla storia “S\_ArcoIngresso” (riquadro centrale in basso);
- “S\_ArcoIngresso” possiede la proprietà di generare essa stessa nuove storie, e infatti abbiamo che “S\_ArcoIngresso” (riquadro centrale in basso):
  - “has\_itemstory S\_SanMicheleArcangelo”
  - “has\_itemstory S\_StatuaMicheleArcangelo”
  - “has\_itemstory S\_MicheleArcangeloProtettoreGargano”
  - “has\_itemstory S\_Ottocento”

- “has\_itemstory S\_UnificazioneItaliana”
- ciascuna di queste storie è mappata in uno specifico “InterestDomain” (riquadro in alto a destra):
  - “S\_SanMicheleArcangelo” nella classe “Saint” appartenente al InterestDomain “Religion” ;
  - “S\_MicheleArcangeloProtettoreGargano” è mappato nella classe “Tradition&History” appartenente all'omonimo InterestDomain;
  - “S\_StatuaMicheleArcangelo” è mappato nella classe “Sculpture” appartenente al InterestDomain “Art&Monument”;
  - “S\_Ottocento” è mappato nella classe “Tradition&History” appartenente all'omonimo InterestDomain;
  - “S\_UnificazioneItaliana” è mappato nella classe “Event” appartenente al InterestDomain “Tradition&History”.
- Infine queste stesse storie possiedono la proprietà di riferirsi a delle risorse esterne (per esempio la risorsa Dbpedia che descrive l'Unificazione d'Italia o quella di San Michele Arcangelo) ovvero a delle risorse interne contenute nel *DataSet* “EvOntology.rdf”:
  - “S\_SanMicheleArcangelo” è caratterizzato da un ItemUri che si riferisce ad una risorsa Dbpedia  
 “[http://dbpedia.org/Michael\\_\(archangel\)](http://dbpedia.org/Michael_(archangel))”
  - “S\_UnificazioneItaliana” è caratterizzato da un ItemUri che si riferisce ad una risorsa Dbpedia (riquadro in basso a destra)  
 “[http://dbpedia.org/Italian\\_unification](http://dbpedia.org/Italian_unification)”
  - Le altre risorse (statua di S. Michele, Ottocento e protettore del Gargano) contengono invece dei riferimenti al *DataSet* “EvOntology.rdf”.

A questo punto, cosa dovremmo fare se volessimo continuare la nostra esplorazione?

In questo prototipo viene suggerito il seguente metodo:

- nel caso in cui la nostra storia dovesse contenere dei riferimenti a risorse interne a EvOntology.rdf, per sviluppare questi ultimi non dovremo fare altro che reiterare il procedimento precedentemente descritto (ampliando di conseguenza il nostro *DataSet*) fino a quando il percorso non dovesse terminare ovvero sfociare verso una risorsa esterna;
- nel caso in cui la nostra storia dovesse contenere dei riferimenti a risorse esterne, dovremo essere in grado di sviluppare il percorso conoscitivo direttamente nel web, connettendosi alla *network* dei *linked data* e tenendo in conto l'esigenza di rispettare lo *User InterestDomain*. Questo procedimento viene descritto nel prossimo capitolo.

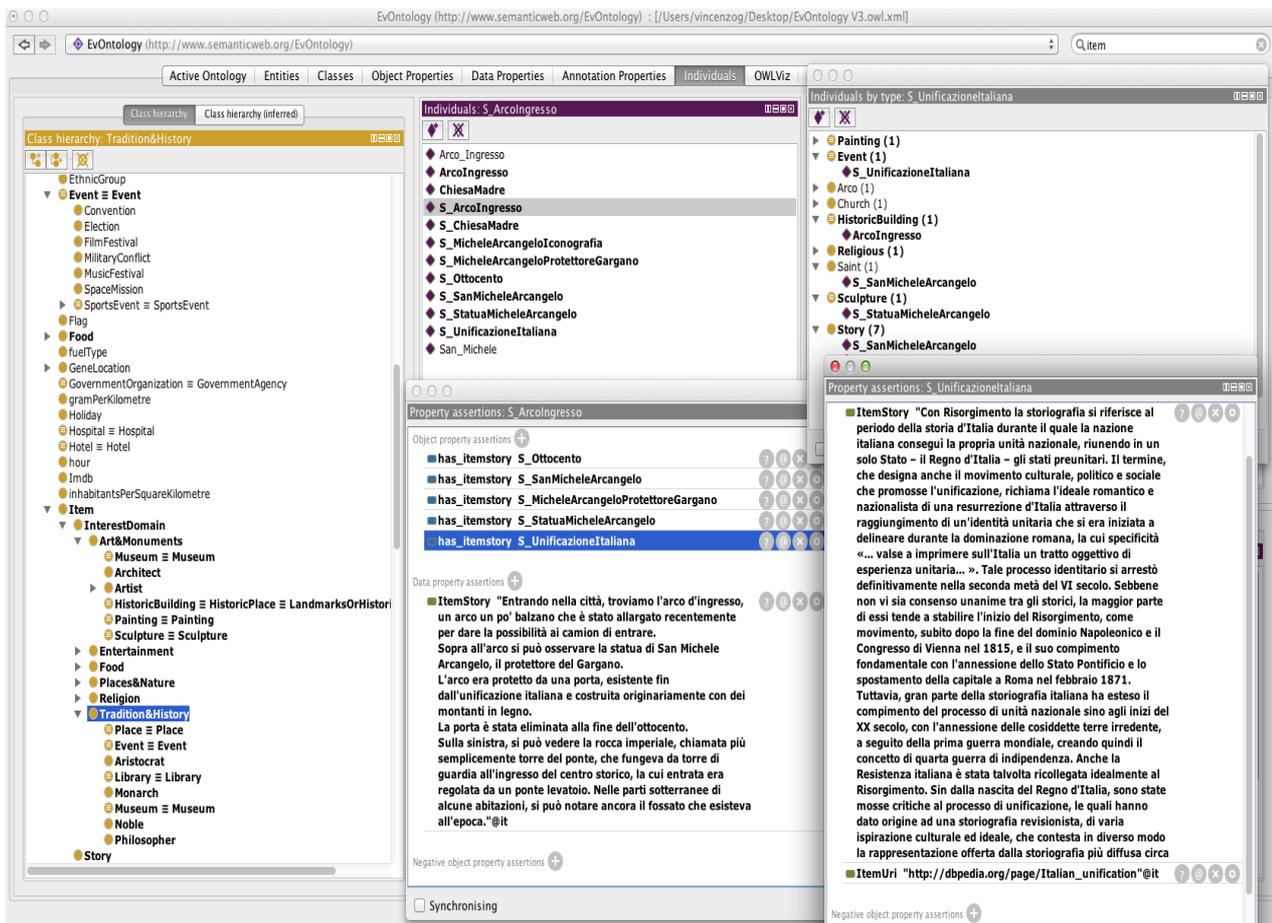


Figura 31: i passaggi che portano alla descrizione degli Item, delle loro storie e dei contenuti di queste ultime, in Protégé

```

//
// Individuals
//
////////////////////////////////////////////////////////////////////
-->

<!-- http://www.semanticweb.org/EvOntology#ArcoIngresso -->
<NamedIndividual rdf:about="&EvOntology;ArcoIngresso">
  <rdf:type rdf:resource="http://dbpedia.org/ontology/HistoricBuilding"/>
  <rdf:comment>fino a fine &apos;800 era protetto da una porta di ingresso</rdf:comment>
  <EvOntology:has_itemstory rdf:resource="&EvOntology;S_ArcoIngresso"/>
</NamedIndividual>

<!-- http://www.semanticweb.org/EvOntology#ChiesaMadre -->
<NamedIndividual rdf:about="&EvOntology;ChiesaMadre">
  <rdf:type rdf:resource="http://dbpedia.org/ontology/Church"/>
  <rdf:comment>Chiesa di Sant&apos;Elia, risalente al 1200</rdf:comment>
  <EvOntology:has_itemstory rdf:resource="&EvOntology;S_ChiesaMadre"/>
</NamedIndividual>

<!-- http://www.semanticweb.org/EvOntology#S_ArcoIngresso -->
<NamedIndividual rdf:about="&EvOntology;S_ArcoIngresso">
  <rdf:type rdf:resource="http://dbpedia.org/ontology/HistoricBuilding"/>
  <rdf:type rdf:resource="&EvOntology;Story"/>
  <EvOntology:ItemStory xml:lang="it">Entrando nella città, troviamo l&apos;arco d&apos;ingresso, un arco un po&apos; balzano che è stato allargato
recentemente per dare la possibilità ai camion di entrare.
Sopra all&apos;arco si può osservare la statua di San Michele Arcangelo, il protettore del Gargano.
L&apos;arco era protetto da una porta, esistente fin dall&apos;unificazione italiana e costruita originariamente con dei montanti in legno.
La porta è stata eliminata alla fine dell&apos;ottocento.
Sulla sinistra, si può vedere la rocca imperiale, chiamata più semplicemente torre del ponte, che fungeva da torre di guardia all&apos;ingresso del centro
storico, la cui entrata era regolata da un ponte levatoio. Nelle parti sotterranee di alcune abitazioni, si può notare ancora il fossato che esisteva
all&apos;epoca.</EvOntology:ItemStory>
  <EvOntology:has_itemstory rdf:resource="&EvOntology;S_SanMicheleArcangelo"/>
  <EvOntology:has_itemstory rdf:resource="&EvOntology;S_StatuaMicheleArcangelo"/>
</NamedIndividual>

<!-- http://www.semanticweb.org/EvOntology#S_ChiesaMadre -->

```

Figura 32: “S\_MicheleArcangeloProtettoreGargano” e “S\_StatuaMicheleArcangelo” contengono dei riferimenti al DataSet EvOntology.rdf



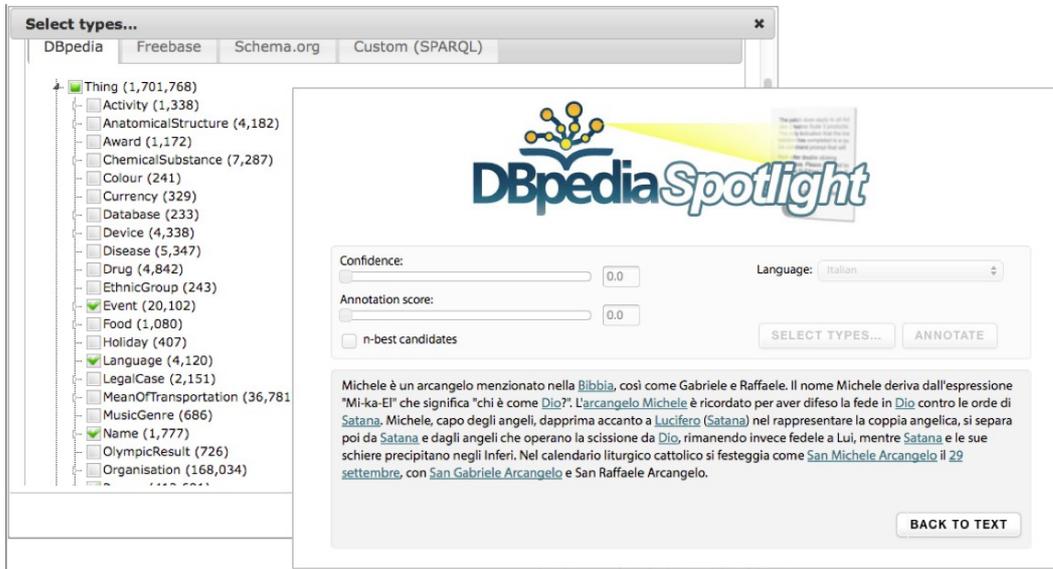


Figura 34: selezione del InterestDomain e parsing dell'abstract di San Michele Arcangelo

A partire dai nuovi *Item* individuati (che sono tutte risorse recuperabili all'interno della rete dei *Linked Open Data*) è possibile reiterare questa tecnica per produrre un percorso potenzialmente senza soluzione di continuità.

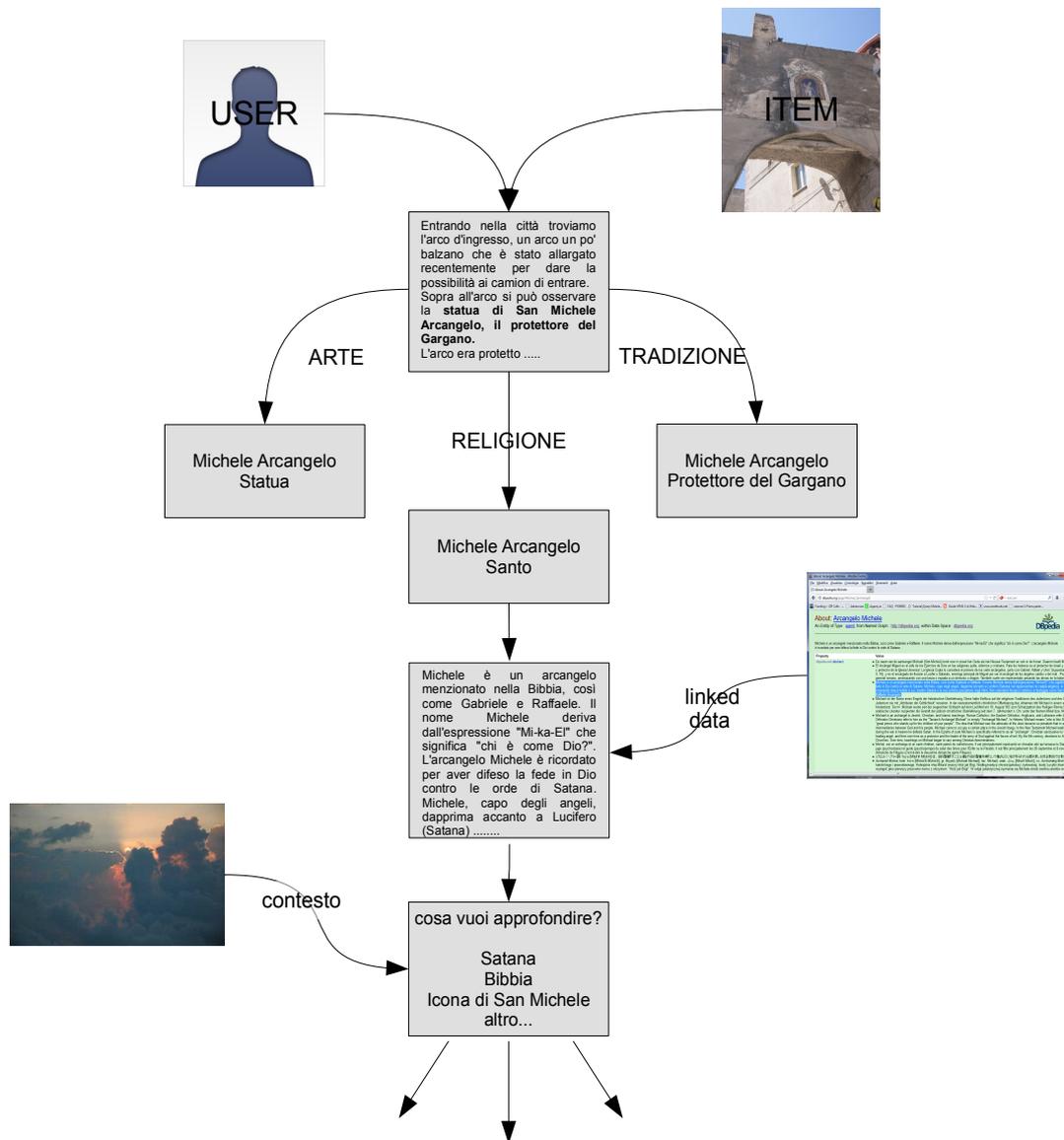
Volendo affinare i suggerimenti forniti allo *User* è possibile adottare tecniche di tipo conversazionale e di elaborazione del contesto che consentano di articolare il percorso evocativo secondo quelle che sono le sue più probabili aspettative: ed è questo che in effetti viene proposto nell'applicazione.

Sarebbe interessante, a questo proposito, studiare l'applicazione di modelli statistici con lo scopo di effettuare una analisi *ex post* delle condizioni che hanno portato un dato utente, con un dato profilo e in un determinato contesto, a sviluppare una storia in un certo modo piuttosto che in un altro<sup>45</sup>. Ciò potrebbe contribuire alla definizione di un sistema di raccomandazione in grado di suggerire non tanto un singolo *Item* quanto una vera e propria tipologia di percorso evocativo.

<sup>45</sup> una possibilità potrebbe essere quella di applicare il teorema di Bayes sulla "probabilità delle cause" (da in'idea di Domenico Patassini)

### 6.2.4 REALIZZAZIONE DELL'APPLICAZIONE “EVONTOLOGYREADER”

Per poter realizzare un test della procedura descritta, è stata sviluppata una applicazione Java, utilizzando il *framework* Jena più oltre descritto, in grado di condurre lo *User* lungo il percorso che maggiormente rispecchia i suoi interessi. In questa applicazione i dati di ingresso riguardano l'*Item* e il profilo (normalizzato) dello *User*. Gli *output* sono dati, *in primis* dalla storia del *Item*, quindi dalla storia di un nuovo *Item*, derivato dalla precedente storia in base al profilo dello *User*. Da questo momento si passerà ad interagire con lo *User*, e mediante l'applicazione di tecniche di tipo conversazionale ed eventualmente effettuando una valutazione del contesto, si svilupperà quel particolare percorso conoscitivo.



Più in dettaglio l'applicazione, in questa sua prima versione di test, funziona utilizzando il *framework* Apache Jena per l'esecuzione di interrogazioni su server Fuseki (motore SPARQL locale fornito dallo stesso sito-web di Apache Jena) ed alcune classi predefinite di Java per la gestione di eventuali errori e per l'utilizzazione di strutture in grado di gestire i vari oggetti su cui si andrà ad operare: come le risorse dell'itinerario, gli *InterestDomain* a cui queste stesse si riferiscono, gli *InterestDomain* propri di ciascun utente che usufruirà del servizio. Richiede inoltre che sia fornito in input il *dataset* su cui sarà operata la scelta delle risorse da raccomandare.

Il flusso dell'applicazione, della quale più oltre è riportato un estratto del codice, è il seguente:

1. si acquisisce la lingua dell'utente del dispositivo (italiana o inglese);
2. si acquisiscono i dati del profilo dell'utente;
3. si acquisisce l'*Item* di ingresso;
4. si recupera la storia riguardante l'*Item* di ingresso;
5. si analizzano gli *InterestDomain* delle risorse correlate alla risorsa recuperata;
6. si seleziona, fra tali risorse, il nuovo elemento da proporre all'utente;
7. si recupera la storia riguardante il nuovo elemento da proporre all'utente;
8. si analizza, utilizzando DBpedia Spotlight, la storia della nuova risorsa, ricavando nuovi riferimenti a risorse esistenti nei *linked data*;
9. si filtrano queste nuove risorse in base agli interessi dell'utente;
10. se una o più risorse vengono individuate si passa al punto 11 altrimenti si ritorna al punto 3, in attesa che sia fornito un nuovo *Item* di ingresso;
11. si presenta all'utente la lista delle nuove risorse e gli si richiede di esprimere una preferenza;
12. si reitera il procedimento a partire dal punto 7.

```

/*
 *          EvOntologyReader          *
 *          *                          *
 * Release Date:    10-2-2014        *
 * Revision Number: 3                 *
 *          *                          *
 *
 * To compile and run this project, you need to execute following commands:
 * > javac -classpath "/jena/path/to/lib/*" EvOntologyReader.java
 * > java -classpath "/jena/path/to/lib/*;" EvOntologyReader
 *
 * Note: the semicolon (";") includes the current folder.
 */

// librerie necessarie allo sviluppo utente
import java.io.BufferedReader;
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.NoSuchElementException;
import java.util.Vector;

// librerie necessarie al sistema d'interrogazione
import com.hp.hpl.jena.query.*;
import com.hp.hpl.jena.rdf.model.*;

// librerie necessarie alla scansione delle pagine-web
import java.net.URL;
import java.net.URLConnection;
import java.net.URLEncoder;

import org.jsoup.Jsoup;
// import org.jsoup.helper.Validate;
import org.jsoup.nodes.Document;
import org.jsoup.nodes.Element;
import org.jsoup.select.Elements;

public class EvOntologyReader {

    // definizione delle variabili che ospiteranno l'indirizzo del servizio remoto,
    // l'indirizzo dell'ontologia remota ed il suo IRI
    static String remoteService;
    static String remoteOntology;
    static String evOntology;

    public static void main(String[] args) {

        // inizializzazione delle variabili con l'indirizzo del servizio remoto,
        // l'indirizzo dell'ontologia remota ed il suo IRI
        remoteService = "http://localhost:3030/sparql";
        remoteOntology = "http://www.p3dclan.com/data/EvOntology.rdf";
        remoteOntology = "http://ts.p3dclan.com/data/EvOntology.rdf";
        evOntology = "http://www.semanticweb.org/EvOntology#";

        // chiediamo all'utente di impostare la preferenza sulla lingua (italiana o inglese)
        String language = getUserLanguage();

        // chiediamo all'utente di comunicarci le proprie classi d'interesse
        // questo metodo richiama un avviso che è stato soppresso con la seguente istruzione
        // @SuppressWarnings("unchecked")
        Vector<String> interests = getUserInterests();

        do {

            // chiediamo all'utente da quale oggetto ha intenzione di cominciare il suo itinerario
            // quando sarà pronta l'applicazione mobile questa informazione sarà passata da quest'ultima
            String startItem = getStartItem("ArcoIngresso");

            // preleviamo l'oggetto storia dell'oggetto principale (questo dovrebbe iniziare con "S.")

```

Figura 35: la parte iniziale del codice dell'applicazione EvOntologyReader

### 6.2.5 REALIZZAZIONE DELLA APPLICAZIONE “EvMOBILE”

In futuro sarà realizzata una semplice app *mobile* utilizzando il *framework* Titanium, in grado di localizzare lo *User*, individuare l'*Item* e di richiamare l'applicazione Jena di cui sopra fornendo in ingresso le informazioni necessarie alla elaborazione.

Questa applicazione è attualmente in fase di progettazione.



Nel prossimo capitolo saranno descritte le tecnologie impiegate nelle diverse fasi di questa realizzazione.

## 6.2.6 LE TECNOLOGIE IMPIEGATE

### 6.2.6.1 PROTÉGÉ – UN EDITOR DI ONTOLOGIE, OPEN SOURCE

Protégé è una *framework open-source* che supporta la comunità degli utenti che operano nel campo del *semantic web*, con una *suite* di strumenti per costruire modelli ontologici e applicazioni *knowledge-based*.

Protégé può essere personalizzato per realizzare ambienti *friendly* per la creazione di modelli di conoscenza e per l'inserimento dei dati e può essere inoltre esteso per mezzo di una architettura *plug-in* e di *Application Programming Interface* (API) *Java-based*.

L'editor OWL di Protégé consente di costruire ontologie per il *semantic web*, secondo gli standard più avanzati raccomandati dal W3C e mette a disposizione diversi strumenti per il *reasoning* e per la ricerca tramite SPARQL.

Protégé è stato sviluppato dalla Stanford University e precisamente presso lo *Stanford Center for Biomedical Informatics Research at the Stanford University School of Medicine* ed è rilasciato con licenza MPL (*Mozilla Public License*).

The screenshot shows the homepage of the Protégé project. At the top left is the Protégé logo and a navigation menu with links: HOME | OVERVIEW | DOCUMENTATION | DOWNLOADS | SUPPORT | COMMUNITY | WIKI | ABOUT US. A search bar is located at the top right. The main content area is titled 'welcome to protégé' and contains several paragraphs of text describing the project as a free, open source ontology editor and knowledge-base framework. It mentions that the platform supports modeling ontologies via a web client or a desktop client, and is based on Java. A 'community' section lists statistics: Registered Users (230,914), protege-announce list members (18,093), protege-discussion list members (2,374), and protege-owl list members (2,375). On the right side, there is a 'Tweets' section showing recent tweets from the Protege Project and Pavel Klinov. A 'news' sidebar on the left mentions a short course in March 2014.

Figura 36: home page del progetto protégé, all'indirizzo <http://protege.stanford.edu/>, a cui si rimanda per ulteriori approfondimenti

### 6.2.6.2 APACHE JENA – UN FRAMEWORK PER IL SEMANTIC WEB E I LINKED DATA

Apache Jena (spesso chiamato Jena per brevità) è un *framework* basato su Java, gratuito ed a sorgente aperto, per la creazione di applicazioni basate sul web semantico ed il mondo dei *linked data*.

Il *framework* è composto da diverse API che interagiscono fra di loro per l'elaborazione di triple RDF. All'interno del pacchetto che viene distribuito, oltre alla documentazione e ad alcuni esempi, è possibile trovare il cuore del *framework*, contenuto nella cartella “lib”.

Il pacchetto contiene tutte le classi compilate di Jena che permettono le più svariate operazioni nell'ambito del *semantic web* come la lettura, la scrittura, la modifica e la gestione di triple in tutti i formati più diffusi (RDF, N3, TURTLE...).

Jena presenta anche la possibilità di applicare un ragionatore al modello caricato e di effettuare un'inferenza sui dati rilevati.

Un modello può essere reperito in vari modi: il più comune è la lettura da un file, ma nulla osta che esso possa essere recuperato da un database, da un URL o da una combinazione fra questi ultimi due.

È inoltre possibile utilizzare un motore SPARQL esterno che permette di interrogare il modello e di ottenere risultati “intelligenti” dovuti al fatto che la struttura dei dati è organizzata semanticamente.

Apache Jena mette a disposizione un server denominato Fuseki per le interrogazioni SPARQL da locale, con la possibilità di interfacciare l'applicazione sviluppata con Jena al servizio di Fuseki.

Lo schema sottostante, riassume il funzionamento di Jena e il definisce il suo interfacciamento con il mondo esterno.

Come si può osservare, il codice dell'applicazione può comunicare con Fuseki per effettuare delle interrogazioni verso delle ontologie esterne.

Inoltre è composto dalle API suddivise principalmente in RDF per la gestione delle triple nei formati del *semantic web* e *Ontology* per la gestione delle ontologie

(entrambe comprendono i *parser* ed i *writer* per scandagliare le informazioni e per la loro scrittura).

Più in basso, troviamo le API di inferenza per l'applicazione di un ragionatore al *dataset* ed ancora nella parte più inferiore (caratteristica di rilievo in Fuseki) le API che permettono il salvataggio e la persistenza delle informazioni in memoria.

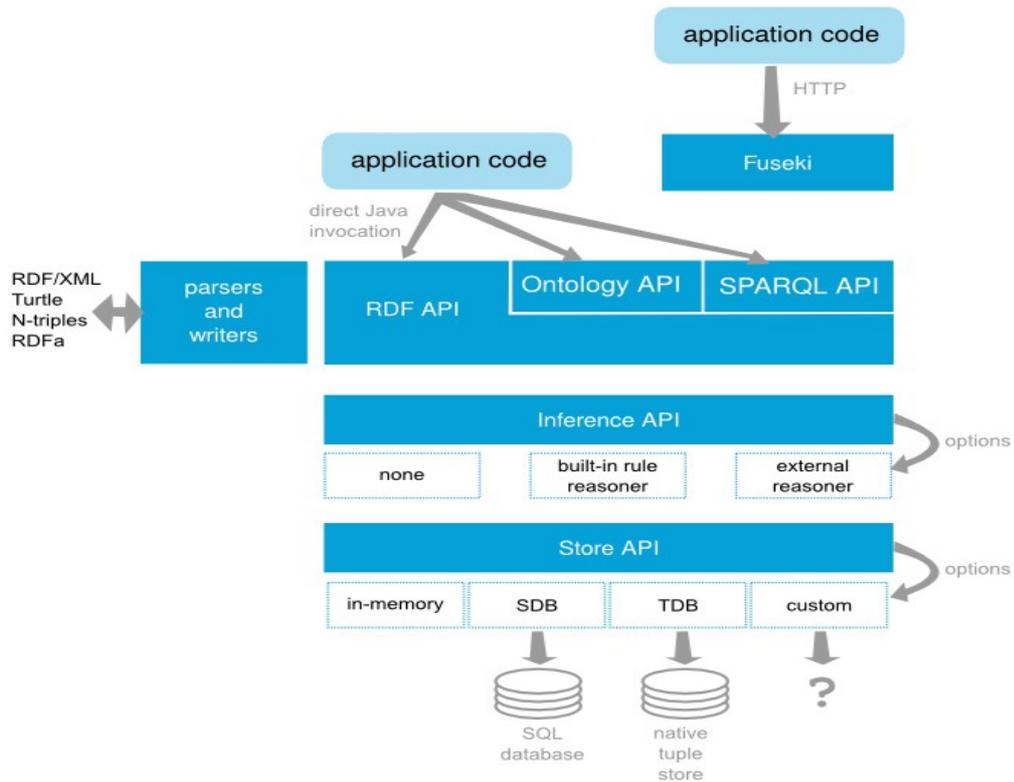


Figura 37: schema di Apache Jena

### 6.2.6.3 APPCELERATOR TITANIUM – UN AMBIENTE PER SVILUPPARE APP MOBILE

Appcelerator Titanium è una piattaforma per lo sviluppo di app orientate ai dispositivi mobili come smartphone e tablet utilizzando le tecnologie web più diffuse (HTML, CSS e JavaScript). Titanium è stata introdotta da Appcelerator Inc. nel dicembre del 2008 ed è tuttora mantenuta e supportata.

*“Appcelerator Titanium is an open, extensible development environment for creating*

*native apps across different mobile devices and OSs including iOS, Android, and BlackBerry, as well as hybrid and HTML5. It includes an open source **SDK** with over 5,000 device and mobile operating system APIs, **Studio**, a powerful Eclipse-based IDE, **Alloy**, an MVC framework and **Cloud Services** for a ready-to-use mobile backend.”*

Il supporto allo sviluppo per applicazioni su iPhone e Android è stato implementato solamente a giugno del 2009.

Il supporto per gli iPad è iniziato nel corso del 2010, ad aprile, precedente a quello per BlackBerry, a giugno dello stesso anno.

Il nuovo sistema Tizen (sviluppato dalla Linux e dalla LiMo *Foundation*) è stato aggiunto solamente nell'aprile 2013 ed attualmente il suo *software development kit* è alla versione 3.1.0.

Titanium è una delle molteplici applicazioni basate sulle tecnologie web che aiuta gli sviluppatori a produrre applicazioni per iPhone ed Android applicando delle competenze già acquisite e largamente diffuse come il linguaggio di *scripting* JavaScript.

Tutto il codice sorgente delle applicazioni è interpretato usando il motore JavaScript del dispositivo, dove l'app è stata distribuita in formato eseguibile.

La programmazione in Titanium risulta alquanto semplice per chi abbia dimestichezza con il linguaggio di *scripting* e fa di questa piattaforma una soluzione altamente portabile e compatibile perché permette di scrivere lo stesso codice ed essere sicuri che funzionerà anche su dispositivi diversi dotati di sistemi operativi differenti. Attualmente più di 55.000 applicazioni sviluppate in Titanium sono state distribuite sui vari *app store*.

A giugno del 2011, Appcelerator ha rilasciato Titanium Studio: Studio è un completo e libero IDE derivato dal precedente Aptana Studio che Appcelerator ha acquistato a gennaio dello stesso anno.

## 7 CONCLUSIONI E SVILUPPI FUTURI

Molte cose sono cambiate da quando ho iniziato diversi anni fa a lavorare su questi argomenti e in particolare sullo sviluppo di applicazioni per il *semantic web*: allora si trattava di tematiche di frontiera rispetto alle quali si intuivano grandi potenzialità ma che non mostravano un percorso chiaro da seguire, cosicché il timore diffuso era quello di intraprendere un ramo secco che non portasse a nulla di buono. Oggi le cose sono molto cambiate: la tecnologia è consolidata, diversi progetti hanno assunto un respiro globale e una moltitudine di nuovi progetti vengono pensati e finanziati anche per applicazioni industriali.

Il momento attuale è particolarmente importante perché sono proprio questo tipo di applicazioni che consentiranno il “salto di qualità” che tutti si attendono, infatti:

*“recent developments indicate that Semantic Web technologies are entering the industrial mainstream. Schema.org and the Facebook Open Graph Protocol are bringing metadata to bear on the Web large-scale. IBM's Watson and Apple's Siri incorporate Semantic Technologies. Google is revamping its search approach and is going more semantic in implementing their knowledge graph. And these are just a few of the prominent examples.”*<sup>46</sup>

In questo contesto, volendo fare qualche breve considerazione legata ai potenziali sviluppi dei sistemi semantici e di raccomandazione nel campo dei beni culturali e del turismo, possiamo affermare che le prospettive sono molto promettenti e soprattutto hanno un obiettivo generale molto chiaro: “guidare” il turista-visitatore alla scoperta di un mondo in la cui vastità e la eterogeneità dell'offerta, la natura multidisciplinare delle informazioni, la sconfinata *community* dei viaggiatori (del pianeta e del web), sembra al di fuori della portata delle sue misere capacità umane.

Per questo motivo alcuni degli studi più avanzati e interessanti prevedono lo sviluppo, nel prossimo decennio, del cosiddetto “*cognitive semantic web*” in cui :

---

46 Dal “Description and scope” del workshop “What will the Semantic Web look like 10 years from now?” tenutosi assieme alla “11th International Semantic Web Conference 2012” - Boston, USA

*“... if we could understand the way semantic web contents are elaborated and used by humans in their daily activities, we could also replicate those modalities and develop intelligent artificial agents capable of – first – generally interacting with humans and – then – specificall adapting to a single user’s profile ...”<sup>47</sup>*

Gli studi sul *cognitive semantic web* e altri studi visionari come quelli sulla condivisione di contenuti emozionali e sulla loro associazione al *network* dei *linked data*, sembrano prospettare la nascita, ancora una volta, di un nuovo web (il 4.0?) in cui la rete sociale sarà costituita da “computer e umani che lavorano in cooperazione”, dando finalmente corpo alla primitiva visione del web di sir Berners Lee.

In questo processo di innovazione la Unione Europea fa la sua parte ed è molto attenta a stimolare una certa vivacità di studi e di ricerche in un ambito in cui può vantare molte eccellenze e per il quale metterà in campo cospicue risorse anche nella prossima programmazione ICT 2014-2020.

Probabilmente è proprio nella prospettiva di proseguire sulla strada dei progetti europei che può essere trovato il maggior interesse per questo lavoro di tesi, che ha certamente molti difetti ma probabilmente anche il pregio di formulare alcune ipotesi che hanno un discreto potenziale di sviluppo: l'idea del percorso di visita evocativo, della raccomandazione serendipica, dello sviluppo di percorsi conoscitivi basati sulla “probabilità delle cause”, ritengo possano essere questioni da porre all'attenzione di future proposte per attività di ricerca: anche in questo caso infatti, come sempre, per poter dare corpo a queste idee (ammesso che possano avere un corpo), è necessario avere a disposizione un minimo di risorse, anche finanziarie, senza le quali si è condannati a rimanere confinati nel proprio piccolo laboratorio di casa.

---

47 Enabling the Cognitive Semantic Web – Alessandro Oltramari – Carnegie Mellon University

## 8 BIBLIOGRAFIA

- Vesin, Ivanović, Klašnja-Milićević and Zoran Budimac - “Ontology-Based Architecture with Recommendation Strategy in Java Tutoring System” (2013)
- Hurrell, Smeaton - “A Conversational Collaborative Filtering Approach to Recommendation” (2013)
- Iandelli, Giannotti, Di Prinzio - “Il monitoraggio e la gestione, attraverso servizi LBS, di eventi diffusi sul territorio” (2013)
- Condotta Massimiliano - "Using Controlled Vocabularies for a Creative Interpretation of Architectural Digital Resources". In: *Getty Research Journal, no 5, 2013*. Editors: Thomas W. Gaetgens, Lucy Bradnock. Published by the Getty Research Institute, Los Angeles. Getty Publications (The J. Paul Getty Trust), Los Angeles, 2013. Pagg. 157-163. (2013)
- Alessandro Oltramari - “Enabling the Cognitive Semantic Web” (2012)
- Gentner, van Harmelen, Hitzler, Janowicz, Kühnberger - “Cognitive Approaches for the Semantic Web” (2012)
- Abraham Bernstein - “The Global Brain Semantic Web – Interleaving Human-Machine Knowledge and Computation” (2012)
- Guerrini, Possemato - “Linked data: un nuovo alfabeto del web semantico” (2012)
- Sam Coppens, Ruben Verborgh, Miel Vander Sande, Davy Van Deursen, Erik Mannens, Rik Van de Walle - “A truly Read-Write Web for machines as the next generation Web?” (2012)
- Marko Grobelnik, Dunja Mladeni , Blaž Fortuna - Semantic Web in 10 years: Semantics with a purpose” (2012)
- Konstantinos N. Vavliakis, Georgios Th. Karagiannis, Pericles A. Mitkas - “Semantic Web in Cultural Heritage After 2020 “ (2012)
- Prokofyev, Boyarsky,Ruchayskiy, Aberer, Demartini,Cudre-Mauroux - “Tag Recommendation for Large-Scale Ontology-Based Information Systems“ (2012)

- Giannotti, Boscolo - "AdriaMuse.com, a new Communication Platform" (2012)
- Ricci, Rokach, Shapira, Kantor - "Recommender Systems Handbook" (2011)
- Marco Zanchetta - "Raccomender System con Intelligenza Collettiva" Tesi di Laurea (2011)
- Djallel Bouneffouf, Amel Bouzeghoub, Alda Lopes Gançarski - "Hybrid-e-greedy for Mobile Context-aware Recommender System" (2011)
- Liyang Yu - "A Developers Guide to the Semantic Web-Springer" (2011)
- Mabroukeh, Ezeife - "Ontology-based Web Recommendation from Tags" (2011)
- ISO 21127 - "A reference ontology for the interchange of cultural heritage information" - second edition (2010)
- Melville, Sindhwani - "Encyclopedia of Machine Learning (chapter 338) - "Recommender Systems" (2010)
- Ivano Boscolo Nale - "GeosemanticBrowser: dalla teoria alla pratica" tesi master (2010)
- Porrello, Bertasio, Tommarchi, Giannotti, Talone - "Cultural planning in small and middle towns Emerging themes and perspectives" (2009)
- Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, Ido Guy - "Make New Friends, but Keep the Old – Recommending People on Social Networking Sites" (2009)
- Chen, Geyer, Dugan, Muller, Guy - "Make New Friends, but Keep the Old – Recommending People on Social Networking Sites" (2009)
- Oufaida, Nouali - "Exploiting Semantic Web Technologies for Recommender Systems A Multi View Recommendation Engine" (2009)
- Vincenzo Giannotti - "Il sistema informativo semantico delle risorse culturali" (2007)
- Hepp, Martin; Siorpaes, Katharina - "Towards the semantic web in e-tourism: can annotation do the trick?" (2006)
- Schwartz, Ward - "Doing Better but Feeling Worse: The Paradox of Choice" (2004)

“A Semantic Web Primer” - Grigoris Antoniou and Frank van Harmelen – The MIT Press (2004)

Thompson, Goker, Langley - “A Personalized System for Conversational Recommendations” (2004)

O'Reilly - “Practical RDF” (2003)

Middleton, Shadbolt, De Roure - “Ontological User Profiling in Recommender Systems” (2001)

Resnick, Varian - “Recommender Systems” (1997)



# APPENDICE

## *LA RICERCA DEL TRIENNIO*

## 9 APPENDICE – LA RICERCA DEL TRIENNIO

Nei capitoli che seguono vengono presentati i cinque progetti che hanno costituito la base di questo lavoro di tesi:

1. Sistema Informativo Semantico delle risorse Culturali (SISC)
2. AdriaMuse
3. Monitoraggio e controllo partecipato della mobilità per l'erogazione di Location Based Services
4. Recommending to People
5. OmniTurist

Se si eccettua il progetto SISC, per il quale si è cominciato a lavorare nel lontano 2007, sono stati tutti realizzati nel corso del triennio di dottorato. Tuttavia anche SISC, che è stato completato nel periodo appena antecedente, è stato inserito in questa presentazione, in quanto è proprio da questo progetto che nasce l'interesse per le tematiche di ricerca descritte nella tesi ed è stato questo progetto un volano fondamentale per l'attivazione e la realizzazione dei successivi.

La presentazione di questi progetti viene effettuata attraverso una breve introduzione descrittiva, se necessaria, e quindi allegando alcuni dei prodotti dei progetti stessi (relazioni, report, paper etc..) che sono stati ritenuti utili per raccontarne i contenuti e i risultati più significativi: il tutto in costante riferimento alle tematiche di ricerca presentate nella tesi: applicazioni per il *semantic web*, sistemi di organizzazione della conoscenza, sistemi di raccomandazione, negli ambiti del turismo e della cultura.

AVVERTENZA: I contenuti allegati sono stati mantenuti il più possibile conformi agli originali, intervenendo con qualche modifica giusto per non appesantire troppo la linearità della esposizione. Alcuni di questi contenuti potrebbero risultare ridondanti rispetto a quanto riportato in tesi; altri sono stati mantenuti nella lingua inglese e non sono stati tradotti.

## **9.1 IL PROGETTO: SISTEMA INFORMATIVO SEMANTICO DELLE RISORSE CULTURALI (SISC)**

Frutto di una collaborazione tra Università Iuav e Regione Veneto, questo progetto di ricerca è stato realizzato tra il 2007 e il 2009, con una coda che si è protratta fino ai primi mesi del 2010.

L'obiettivo generale del progetto SISC era quello di mappare semanticamente le risorse culturali del Veneto utilizzando le tecnologie più all'avanguardia per l'organizzazione della conoscenza e la comunicazione dell'informazione. Lo scopo derivato dall'obiettivo generale era quello di attivare una rete di operatori culturali attraverso cui potessero essere attuate in maniera condivisa e coordinata politiche di valorizzazione del patrimonio culturale veneto.

Per arrivare a questo risultato fu progettato e sperimentato un primo nucleo della rete (nell'area del bellunese) nel quale le risorse non erano semplicemente geo-riferite attraverso delle coordinate, ma anche organizzate nel web attraverso uno schema ontologico.

La scelta di utilizzare le tecnologie del *semantic web* – nel 2007 abbastanza ben definite dal W3C ma non ancora supportate da una adeguata casistica di progetti – fu legata sia al tentativo di superare i limiti dei motori di ricerca del tempo, sia alla volontà di adottare un differente approccio, più articolato e versatile, per la diffusione della cultura, in linea con gli studi che in quel periodo venivano condotti sul tema del *cultural planning*:

*“The original concept of cultural mapping, [...], as a representation method for resources, networks and relationships in a group or a community – with their related geographic locations – has evolved consequently to the development of the ICTs, encouraging the shift of cultural mapping to the net. Here, an object (a resource) is characterized by an added reference: an address (URL) that individuates it univocally in the net.”*<sup>48</sup>

---

48 Cultural planning in small and middle towns: emerging themes and perspectives – Porrello, Bertasio, Tommarchi, Giannotti, Talone – 9<sup>th</sup> ESA Conference, Lisbon, September 2009

La prima fase del *cultural mapping* prevede un inventario delle risorse culturali mediante una rilevazione diretta. Tale rilevazione può essere associata o talvolta omessa, nel caso in cui siano presenti precedenti analisi, censimenti e indagini esaurienti. La successiva analisi qualitativa è la più specifica della pianificazione culturale, in quanto le caratteristiche del luogo – le relazioni sociali, gli stili di vita etc. – necessitano di dati qualitativi.

Poiché la mancata manifestazione della domanda richiede che si operi per la costruzione della domanda stessa, è necessario che questa sia compresa e integrata nelle scelte della politica culturale delle istituzioni locali e dell'azione degli *stakeholders*. Ciò può avvenire appunto, attraverso il *cultural mapping*, che nel SISC ha rappresentato la prima attività svolta, fondamentale per la successiva modellazione ontologica.

## ONTOLOGIA E DATI NEL SISC

### L'ONTOLOGIA DEL SISC (VERSIONE ALFA 2)

Viene presentata l'ultima versione della ontologia SISC (versione alfa 2), ricordando che essa è stata realizzata nell'ambito del dominio della “offerta culturale” con lo scopo di favorire la nascita di un “network degli operatori culturali”.

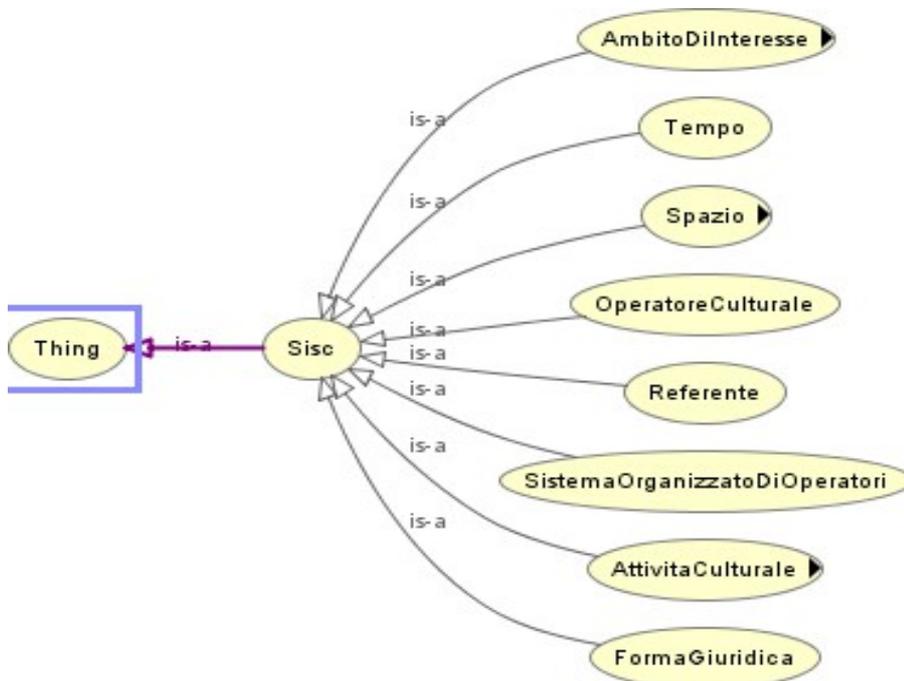


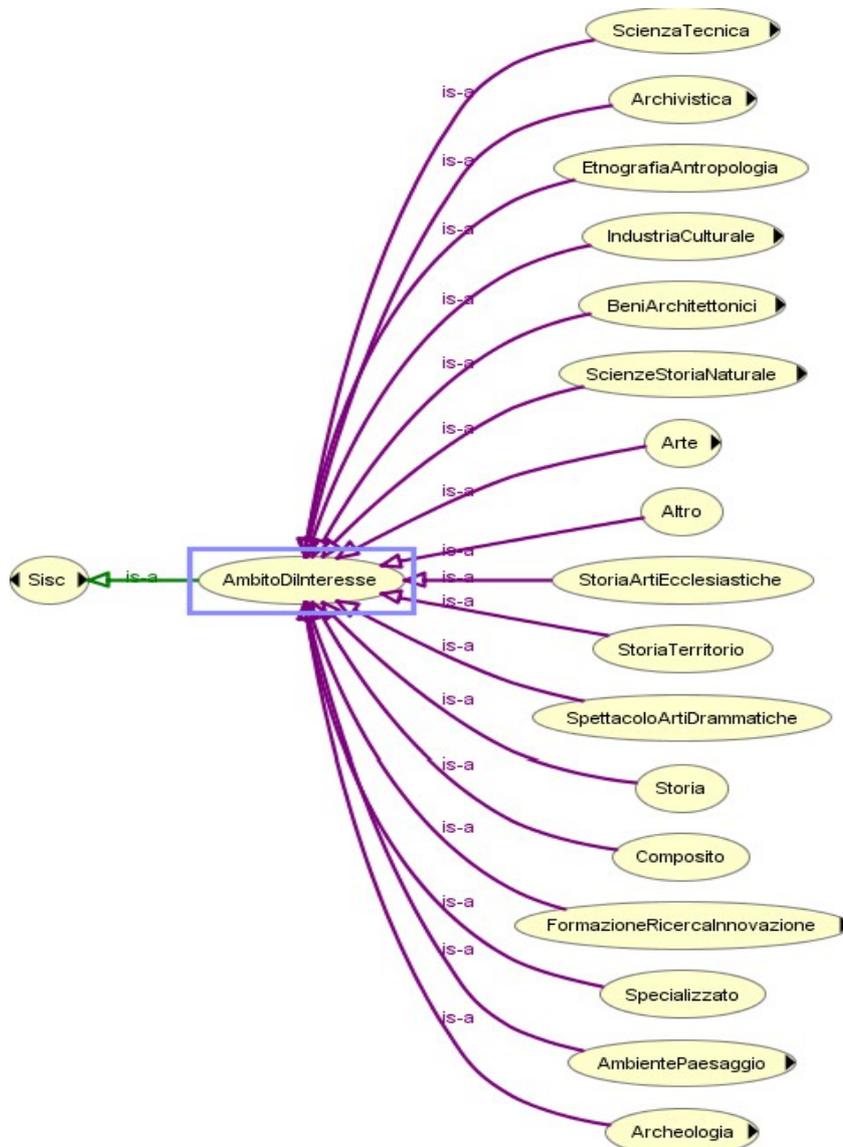
Figura 38: concetti principali dell'ontologia SISC

L'ontologia del SISC è stata implementata in OWL. OWL è una raccomandazione del W3C per lo sviluppo delle ontologie per il semantic web. Per garantire l'interoperabilità tra tutte le ontologie esistenti e future OWL definisce una radice comune (*Thing*) da cui tutte le ontologie debbono partire. L'ontologia sviluppata per il SISC, viene nominata con il nome della sua stessa radice *Sisc* scritto in minuscolo con l'iniziale maiuscola in modo da non confonderla con il nome del progetto che invece è un acronimo scritto in maiuscolo.

*Sisc* nasce da *Thing* ed è costituita da otto concetti principali<sup>49</sup>: *AmbitoDiInteresse*, *Tempo*, *Spazio*, *OperatoreCulturale*, *Referente*, *SistemaOrganizzatoDiOperatori*, *AttivitàCulturale* e *FormaGiuridica*.

Le proprietà che possono legare tra loro i concetti sono: *appartieneA*, *collabora*, *haReferente*, *eDiRiferimento*, *haCoordinate*, *haFormaGiuridica*.

Il concetto *AmbitoDiInteresse* è quello maggiormente sviluppato nell'ontologia *Sisc* ed è costituito da diciassette concetti derivati, alcuni dei quali sono sviluppati<sup>50</sup> in un



ulteriore livello di approfondimento. L'ambito di interesse serve per inquadrare il

<sup>49</sup> Il termine concetto in questo contesto viene sostituito al termine più tecnico di classe.

<sup>50</sup> Vedere la documentazione in HTML per avere una visione completa dell'ontologia SISC.

settore culturale all'interno del quale l'operatore svolge la propria attività.

Il concetto di *Tempo* è presente nell'ontologia solo per completezza e non è stato né ulteriormente sviluppato né utilizzato per la costruzione delle istanze attualmente implementate. Il tempo servirà in futuro, per modellare tutti quei fenomeni dotati di proprietà temporale che ruotano intorno al complesso mondo della cultura: mostre temporanee, spettacoli, manifestazioni ed eventi in generale.

Come per il tempo anche il concetto di *Spazio* è stato sviluppato in forma minima e contiene il solo concetto derivato di *Coordinate*. Le coordinate servono per rappresentare la posizione geografica di un oggetto che può essere modellato in un sistema di consultazione di mappe tramite un punto, come ad esempio il centroide delle sedi degli operatori culturali posizionato all'interno di un'immagine digitale del Veneto. Il concetto di spazio, in futuro, potrà essere sviluppato in modo da modellare altre possibili forme di localizzazione spaziale di un oggetto, come ad esempio l'indirizzo. Il concetto di coordinata dovrà essere sviluppato in modo da potersi interfacciare con i motori di elaborazione spaziale, come ad esempio i GeoRDBMS. Per poter fare questo sarà necessario sviluppare nell'ontologia un modello in grado di rappresentare le tre primitive di rappresentazione di feature geometriche (punto, linea, poligono) e il sistema di riferimento spaziale utilizzato<sup>51</sup>. Attualmente è stata definita una sola proprietà che lega l'operatore culturale alle proprie coordinate geografiche: *haCoordinate*.

Il concetto di *OperatoreCulturale* serve per raggruppare tutte le possibili istanze presenti e future di operatore culturale. A questo concetto è stata data all'interno dell'ontologia *Sisc* una definizione aperta: “*Chiunque produce un'offerta nel campo della cultura*” che permette di accogliere sviluppi anche inaspettati del concetto di operatore culturale<sup>52</sup>, come ad esempio l'artista oppure il consumatore attivo (prosumer) del web 2.0.

---

51 In letteratura esistono molti esempi da cui trarre ispirazione oppure interfacciarsi con l'ontologia del SISC.

52 Attualmente nel SISC abbiamo istanze di operatori culturali appartenenti a categorie istituzionali.

Nell'ottica di costruire una rete sociale che possa essere elaborata da un agente software è stata definita la proprietà *collabora*<sup>53</sup> che lega assieme due diversi operatori culturali. Tale proprietà, grazie alle interviste eseguite durante la prima fase del progetto, è stata arricchita con le seguenti voci:

- Ricerca
- Comunicazione e Promozione
- Organizzazione Eventi
- Prestiti beni
- Biglietto Unico Convenzione

Il concetto di *Referente* è legato alle persone fisiche di riferimento per un operatore culturale di tipo istituzionale ed è collegato all'operatore culturale nell'ontologia *Sisc* tramite la proprietà *haReferente*.

Il concetto di *SistemaOrganizzatoDiOperatori* serve per identificare quei particolari insiemi di operatori culturali che sono dati dalle reti o dai sistemi organizzati; per esempio: reti museali, poli museali, reti civiche etc. All'interno della banca dati OBSERVA tali istituzioni sono state mescolate assieme agli operatori culturali che ne fanno parte; una modellazione molto più efficace potrebbe essere fatta istanziando separatamente i sistemi organizzati e collegando ad essi gli operatori culturali tramite la proprietà *appartieneA* dell'ontologia *Sisc*.

---

<sup>53</sup> Purtroppo tale proprietà non è ancora stata utilizzata in modo adeguato all'interno delle istanze implementate per questa prima fase del progetto, attualmente non si è ancora trovato un modo per associare un nome univoco ed al contempo autoesplicativo agli operatori culturali per poterli collegare tra loro in modo automatico o semi automatico al momento della compilazione di un profilo di un operatore culturale.

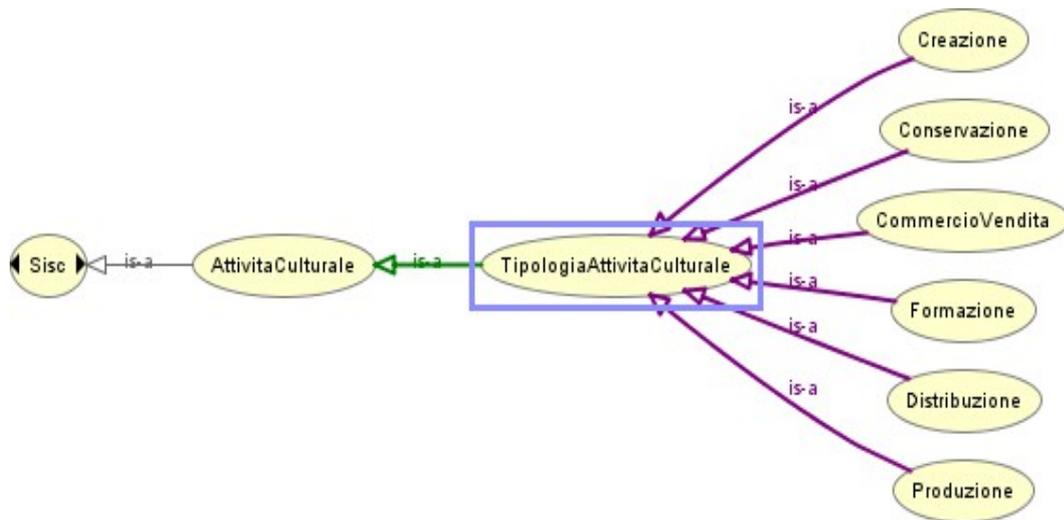


Figura 39: Il concetto di attività culturale nell'ontologia SISC

Il concetto di *AttivitaCulturale* è aperto a futuri sviluppi e contiene l'unico sotto-concetto di *TipologiaAttivitàCulturale* per il quale sono state utilizzate le funzioni presenti nello schema Leg (Cultural Statistics in Europe): *Creazione*, *Conservazione*, *CommercioVendita*, *Formazione*, *Distribuzione* e *Produzione*.

Il concetto di *FormaGiuridica* è attualmente utilizzato da tutte le istanze di operatore culturale, che al momento sono solo di tipo istituzionale, ma non ancora modellate all'interno dell'ontologia.

Le tipologie di forma giuridica, ricavate dalla banca dati OBSERVA e integrate attraverso la campagna di interviste, sono:

- Associazione riconosciuta
- Comune
- Comunità montana o isolana
- Consorzio o altra forma di cooperazione
- Consorzio di diritto pubblico

- Ente ecclesiastico o religioso
- Ente Pubblico
- Fondazione
- Istituto o ente di ricerca
- Istituto o scuola di ogni ordine e grado
- Privato cittadino
- Provincia
- Regione
- Società cooperativa
- Società di persone o capitali
- Soggetto privato
- Unione di Comuni
- Università non statale
- Università statale

## I DATI DEL SISC

I dati implementati in questa prima fase del SISC sono stati per lo più acquisiti attraverso la compilazione di apposite schede informative redatte dal gruppo di lavoro e pubblicate nel web e attraverso l'acquisizione di informazioni ricavate da banche dati esistenti.

Nel prosieguo del documento si illustra come tali dati siano stati raccolti per la pubblicazione in RDF, ma per una trattazione approfondita delle problematiche di raccolta dati e di costituzione del network, si rimanda ai relativi documenti che fanno parte del pacchetto di consegna finale.

## LE SCHEDE INFORMATIVE

Per l'acquisizione delle informazioni è stata sviluppata una semplice applicazione web che presenta due schede per acquisizione dati (*form*) distinte: una per gli operatori dell'alto bellunese e una per i musei del Veneto.

I *form* sono serviti sia agli operatori, che hanno potuto compilare in modo autonomo le proprie informazioni, sia al gruppo di lavoro che ha potuto inserire da remoto e in modalità asincrona le interviste effettuate faccia faccia con gli operatori culturali.

Le schede informative sono organizzate in maniera da poter acquisire in due diversi *form* le informazioni richieste:

- da un lato le informazioni anagrafiche dell'operatore culturale, le informazioni relative alla persona di riferimento e la forma giuridica dell'operatore;
- dall'altro, nel secondo *form*, vengono acquisite informazioni che riguardano: l'ambito culturale dell'operatore, la tipologia di attività che svolge<sup>54</sup>, l'appartenenza a reti e sistemi organizzati, la descrizione di eventuali rapporti di collaborazione con altri operatori, informazioni sul personale occupato e sui finanziamenti ottenuti.

---

<sup>54</sup> Per l'inquadramento delle attività sono state utilizzate le funzioni presenti nello schema Leg (Cultural Statistics in Europe).

REGIONE DEL VENETO

La Regionale Terzi Istituzionali Servizi alla Persona Ambiente e Territorio Economia

**Ambito di interesse**

Selezionare uno o più delle seguenti opzioni e correlarle da una breve descrizione che dettagli l'ambito culturale all'interno del quale l'operatore svolge le proprie attività.

**Attività**

Selezionare uno o più delle seguenti opzioni e correlarle da una breve descrizione che dettagli le macro attività (individuate secondo la classificazione europea LEG) su cui l'operatore può identificare le attività che svolge.

**Collaborazioni, reti e sistemi organizzati tra gli operatori**

Se l'operatore fa parte di una o più reti o sistemi organizzati, per esempio: reti museali, poli museali, reti civiche etc. Specificare la denominazione della rete/sistema organizzato assieme alla tipologia di rapporto all'interno della rete.

**Collaborazioni tra gli operatori**

Se l'operatore collabora con uno o più operatori, indicare le collaborazioni permanenti, periodiche o occasionali. Per esempio:

n°	Operatore	Oggetto della collaborazione
1	Istitut Ladin de la Dolomites	scambio di opere o beni per l'esposizione
2	Union Ladina de Fodom	cooperazione nell'organizzazione di eventi

**Personale**

Dove presente, indicare il numero di persone che hanno avuto nell'anno 2007 un rapporto di lavoro con l'operatore culturale, seguendo le tipologie date.

**Risorse finanziarie e fonti di finanziamento**

Indicare l'ammontare delle entrate dell'anno 2007 (compresi i contributi pubblici) suddiviso per le fonti date.

**Osservazioni del compilatore**

Difficoltà nella compilazione della scheda; considerazione sui contenuti della scheda; aspetti che avreste voluto venissero affrontati; etc.

Salva la scheda

© Regione del Veneto | Controlli 041.219.11.11 - informazioni - suggerimenti / commenti - URP - accessibilità - privacy

Figura 40: il I form di acquisizione dati per gli operatori dell'alto bellunese (I parte)

REGIONE DEL VENETO

La Regionale Temi Istituzionali Servizi alla Persona Ambiente e Territorio Economia

### Progetto SISC - Sistema informativo semantico delle risorse culturali

Nell'ambito del Progetto SISC, realizzato dall'Università IUAV di Venezia e coordinato dalla Regione del Veneto. La invitiamo a compilare la seguente scheda inserendo i Suoi dati identificativi, e quelli relativi agli ambiti e alle attività che svolge.

#### Anagrafica dell'operatore culturale

Compilare l'anagrafica relativa alla sede principale dell'operatore.

\* campi obbligatori

Denominazione: \*

Via/Viale/Piazza/: \*

Civico: \*

Comune: \*

Provincia:  ▼

Località: \*

Telefono:

Fax:

E-Mail: \*

Sito Web:

**Referente**

Nome: \*

Cognome: \*

#### Forma giuridica del titolare

Selezionare la tipologia di ente di appartenenza.

\*  Ente Pubblico     Ente Privato

Tipologia del titolare: \*  ▼

Nel caso di "Altro ente pubblico" o "Altro soggetto privato" è qui possibile specificare:

Accetta le condizioni e prosegui con la compilazione >>

**INFORMATIVA SUL TRATTAMENTO DEI DATI PERSONALI AI SENSI DELL'ART. 13 DEL D.LGS. 30 GIUGNO 2003 N. 196**

Il D.Lgs. n. 196/2003 prevede la tutela della persona e di altri soggetti rispetto al trattamento dei dati personali. Secondo la legge tale trattamento è improntato ai principi di correttezza, liceità e trasparenza, tutelando la Sua riservatezza e i Suoi diritti. I dati da Lei forniti saranno utilizzati all'interno del progetto SISC (Sistema Informativo Semantico delle Risorse Culturali), commissionato dalla Regione del Veneto allo IUAV, e potranno essere utilizzati per l'invio di avvisi e/o materiale informativo circa ulteriori iniziative/convegni su argomenti affini o correlati a al suddetto progetto. I dati, la cui gestione è informatizzata e manuale, sono custoditi in un data center della Regione del Veneto che potrà avvalersi della collaborazione di professionalità esterne qualificate per le attività di gestione e manutenzione informatica dello stesso. I dati, tranne quelli relativi al budget e ad informazioni sul personale, saranno pubblicati nel web. Il Titolare del trattamento dei dati è: Regione del Veneto / Giunta Regionale, con sede in: Venezia, Palazzo Balbi, Dorsoduro 3901. Il Responsabile del trattamento dei dati è la Dirigente della "U.C. progetti strategici e politiche comunitarie", con sede in Venezia, Palazzo Sceriman, Cannaregio 168. Come noto, Le competono tutti i diritti previsti dall'art. 7 del D.Lgs. n. 196/2003. Lei potrà, quindi, chiedere al Responsabile del trattamento la correzione e l'integrazione dei propri dati e, ricorrendone gli estremi, la cancellazione o il blocco.

U.C. progetti strategici e politiche comunitarie - Palazzo Sceriman - Cannaregio 168 - 30121 - Venezia

© Regione del Veneto | [Contattaci 041.279.21.11](#) - [Informazioni](#) - [suggerimenti / commenti](#) - [URP](#) - [accessibilità](#) - [privacy](#)

Figura 41: il form di acquisizione dati per gli operatori dell'alto bellunese (II parte)

## **INTEGRAZIONE CON BANCHE DATI ESISTENTI**

Le informazioni acquisite nel *form* online sono state integrate con quelle provenienti da precedenti campagne informative (OBSERVA Regione del Veneto 2006), all'interno di una banca dati Mysql<sup>55</sup>. Le fasi di integrazione e di generazione dei file RDF sono riportate di seguito:

- acquisizione e omogeneizzazione delle banche dati
- estrazione delle informazioni relative alle persone di riferimento
- attribuzione di un identificativo per tutte le risorse coinvolte
- generazione dei file in formato RDF.

### **Acquisizione e omogeneizzazione delle banche dati**

La banca dati OBSERVA e le informazioni pervenute tramite intervista oppure dalla compilazione dei form online sono state integrate in un'unica banca dati risolvendo i conflitti sugli identificativi ed eliminando le ridondanze. A tutti gli operatori sono state attribuite delle coordinate geografiche acquisite tramite il servizio di geocoding offerto da GoogleMaps<sup>56</sup>.

### **Estrazione delle informazioni relative alle persone di riferimento**

Da questa banca dati è stata estratta una prima lista di persone di riferimento. Le persone di riferimento sono state codificate secondo il modello ontologico standard FOAF.

### **Attribuzione di un identificativo per tutte le risorse coinvolte**

Tutte le risorse sono state identificate tramite un URI contenente il prefisso di

---

<sup>55</sup> All'interno della consegna "sequenza cinque" del progetto: operatori 20091127 1557.sql

<sup>56</sup> <http://code.google.com/intl/it-IT/apis/maps/documentation/services.html>

dominio

<http://www.regione.veneto.it/sisc#>

seguito da una stringa che distingue ciascuna risorsa in maniera univoca all'interno del dominio definito. Le stringhe utilizzate per i referenti sono state create tramite la concatenazione del campo *Nome* e *Cognome* eliminando spazi ed eventuali caratteri speciali, per esempio accenti, simboli di punteggiatura etc.

Le stringhe utilizzate per definire gli operatori culturali sono state create concatenando alla stringa *operatore\_* l'identificativo numerico presente nella banca dati. Non è stato possibile identificare in maniera univoca gli operatori tramite il campo denominazione poiché molti di questi utilizzano nomi simili (per esempio il nome *Musei Civici*): questo fatto potrebbe limitare l'accessibilità al dato in contesti più ampi dove i collegamenti tra le risorse divengono importanti (linked data<sup>57</sup>). Per il futuro sarà opportuno individuare una diversa modalità di definizione di tali risorse.

Infine, le coordinate geografiche sono state create concatenando alla stringa *coord\_* l'identificativo della risorsa a cui si riferiscono.

### **Generazione dei file in formato RDF**

La generazione dei file in RDF è stata eseguita in due modi:

- tramite l'esportazione in formato XML di alcune tabelle presenti nella banca dati e utilizzando una trasformata XSLT per convertire le tabelle in formato RDF/XML
- attraverso una procedura PHP sviluppata *ad hoc* che esegue una query nella banca dati e produce in output la stessa, etichettando in modo opportuno le risorse, in formato RDF/XML.

---

<sup>57</sup> I Linked Data riguardano la possibilità di utilizzare il Web per collegare dati correlati che prima non erano associati, utilizzando gli URI e l'RDF - <http://linkeddata.org/>

Tutte le informazioni raccolte tramite le applicazioni web e le campagne informative sono state uniformate all'interno della banca dati sopra citata. Le informazioni raccolte all'interno della medesima banca dati sono servite sia per creare le istanze delle prime risorse del SISC<sup>58</sup> sia per arricchire l'ontologia. I concetti e le proprietà dell'ontologia sviluppata (versione alfa 2), sommariamente descritti all'interno in questo documento al cap. 3.1, sono consultabili in maniera completa nel sito web della stessa, disponibile tra gli allegati (cartella html) della sequenza cinque.

---

<sup>58</sup> Riportate in RDF/XML negli allegati consegnati nella sequenza cinque.

## **ALTRE ONTOLOGIE UTILIZZATE PER ANNOTARE LE RISORSE**

Come più volte accennato, i dati che possono essere oggetto delle ricerche da parte del GeoBrowser sono pubblicati in RDF e hanno una struttura compatibile con la ontologia di riferimento. Le risorse individuate durante questa prima fase del progetto non solo sono state annotate secondo l'ontologia *Sisc* presentata in precedenza ma anche utilizzando ontologie e sistemi di metadati sviluppati da terze parti.

L'ontologia FOAF è stata utilizzata per definire la tipologia degli operatori culturali che nel caso specifico sono sia delle istanze del concetto *sisc:OperatoreCulturale* sia del concetto *foaf:Organization* e per la denominazione (*foaf:name*) dell'operatore culturale. FOAF inoltre è stata utilizzata anche per annotare le informazioni inerenti le persone di riferimento definendole come appartenenti al concetto *foaf:Person* e per indicare il nome (*foaf:Name*) e il cognome delle stesse.

Il Dublin Core è stato utilizzato per annotare le risorse del SISC che hanno una descrizione specifica (*dc:description*), come ad esempio le attività e le peculiarità degli ambiti di interesse degli operatori culturali frutto delle interviste ad essi dirette.

GeoRss è stato utilizzato per rappresentare le coordinate degli operatori culturali. Per le necessità emerse in questa prima fase del progetto è stata utilizzata solo la primitiva punto (*georss:point*) rappresentando le coordinate come coppia di valori decimali separate da uno spazio. Tra le codifiche possibili definite dal GeoRSS, quella scelta è la più semplice (*Simple*).

## PUBBLICAZIONE DELLE RISORSE

Nella prima fase del progetto SISC non è stato affrontato compiutamente il problema della pubblicazione e della diffusione delle risorse e dell'ontologia del SISC. In questo capitolo si cerca di definire, per lo meno in linea di principio, quali siano i requisiti fondamentali per la diffusione delle risorse e delle ontologie nell'ambito del semantic web.

Quando si pubblica e si diffonde una nuova ontologia è innanzitutto necessario che sia chiaramente esposto il suo contenuto: deve essere ben commentata – per questo si possono utilizzare le annotazioni *rdf:label* e *dc:description* - in modo che sia chiaro il significato di ogni concetto e proprietà. Una nuova ontologia deve essere accompagnata da un sito web che ne illustri le caratteristiche principali e i possibili scenari di utilizzazione e dovrebbe essere pubblicata anche in una versione *html* in modo che non sia necessario installare alcun particolare software sul computer di chi la vuole consultare.

Per poter essere utilizzata all'interno di applicazioni sviluppate da terzi parti è sufficiente che sia possibile accedere all'ontologia attraverso il protocollo *http*. Nel caso in cui la dimensione del file sia superiore ad un dato valore (coi normali valori di banda alcune centinaia di Kilobyte), conviene fare in modo che gli utilizzatori possano scaricare l'ontologia in modo da poterla utilizzare direttamente sul proprio computer.

A differenza delle ontologie che non dovrebbero mai superare alcune centinaia di Kilobyte<sup>59</sup>, i dati, che possono essere anche piuttosto pesanti, quando è possibile dovrebbero essere memorizzati in una banca dati apposita; in tal caso sarebbe conveniente fornire ai potenziali utenti una interfaccia su cui possano essere eseguite delle query SPARQL<sup>60</sup>.

---

<sup>59</sup> Riteniamo che in linea di massima sia sempre meglio sviluppare piccole ontologie su domini specifici eventualmente utilizzando diverse ontologie in funzione della applicazione che si vuole sviluppare.

<sup>60</sup> Tecnicamente un endpoint SPARQL ([http://semanticweb.org/wiki/SPARQL\\_endpoint](http://semanticweb.org/wiki/SPARQL_endpoint)).

La diffusione delle risorse nell'ambito del linked data, che è un sotto ambito del semantic web, deve seguire alcuni semplici principi per essere efficace. Per prima cosa bisogna distinguere tra due tipi di risorse: le risorse informative e le risorse non informative. Tutte le risorse che possiamo trovare in una pagina web tradizionale, come ad esempio documenti, immagini e altri file multimediali, sono risorse informative, mentre: la gente, i luoghi, i concetti scientifici, e così via sono risorse non informative. *“Come regola generale, tutti gli "oggetti del mondo reale" che esiste al di fuori del Web sono risorse non informative.”* Compreso questo possiamo elencare alcuni principi di base per la diffusione delle risorse non informative nel web:

- fornire un identificativo (URI) alla risorsa
- fornire un identificativo alla risorsa che sia adatto per i browser HTML (rappresentare la risorsa come pagina web)
- fornire un identificativo alla risorsa che sia adatto per i browser RDF (con una rappresentazione della stessa in RDF/XML).

Per esempio:

- <http://regione.veneto.it/sisc.it/resource/OperatoreCulturale>
- <http://regione.veneto.it/sisc.it/page/OperatoreCulturale>
- <http://regione.veneto.it/sisc.it/data/OperatoreCulturale>

Oppure:

- <http://id.regione.veneto.it/sisc.it/OperatoreCulturale>
- <http://page.regione.veneto.it/sisc.it/OperatoreCulturale>
- <http://data.regione.veneto.it/sisc.it/OperatoreCulturale>

## **IL SEMANTIC GEOBROWSER**

Le possibilità di utilizzazione del semantic web sono innumerevoli e difficilmente indagabili all'interno di questo documento: non si tratta solamente di immaginare dei filoni tematici, come può essere quello del network degli operatori culturali preso in esame dal SISC o quello dell'offerta culturale o della conservazione di beni culturali o della gestione del paesaggio, ma anche lo sviluppo di scenari applicativi come possono essere: la ricerca di informazioni, il cosiddetto “agente” o quello della organizzazione di un network collaborativo.

Un filone piuttosto che un altro può richiedere che sia sviluppato un idoneo strumento di indagine del semantic web e di gestione del filone stesso, anche in termini di organizzazione di risorse e strutture di supporto.

Nel capitolo successivo vengono prospettate due tipologie di scenario applicativo tra quelle prese in esame durante la fase di elaborazione del SISC. Lo scopo è quello di definire una ipotesi di ambito in cui collocare il progetto di massima del GeoBrowser Semantico.

## SCENARI APPLICATIVI

Il questo capitolo si illustrano due possibili scenari applicativi:

- ricerca di informazioni, all'interno del tema “conservazione dei beni culturali”
- agente culturale, per la costruzione di palinsesti/percorsi di visita di “eventi culturali”

## LA RICERCA DI INFORMAZIONI

### **Prima applicazione - Ricerca di monete catalogate nei musei del Veneto.**

Questo esempio descrive il flusso completo (in sei passi) per la realizzazione di una applicazione funzionante in forma di prototipo:

1. passo 1 - vengono individuati: il “dominio di interesse” e lo “scopo dell'applicazione”. In questo caso il dominio è quello della “conservazione di beni culturali” e lo scopo della applicazione è quello di recuperare all'interno del web semantico le schede di monete contenute nella banca dati dei beni culturali della Regione Veneto e archiviate secondo gli standard ICCD;
2. passo 2 - vengono analizzate le schede informative degli oggetti che si intendono recuperare;
3. passo 3 - viene definita la ontologia relativa al dominio di interesse e in funzione dello scopo applicativo: per questo esempio è stata definita una ontologia semplificata, con quattro concetti base ciascuno dei quali sviluppato al minimo, col solo obiettivo di far funzionare il prototipo. È interessante notare che sebbene questa ontologia sia molto semplice, consente comunque di effettuare delle ricerche molto precise.

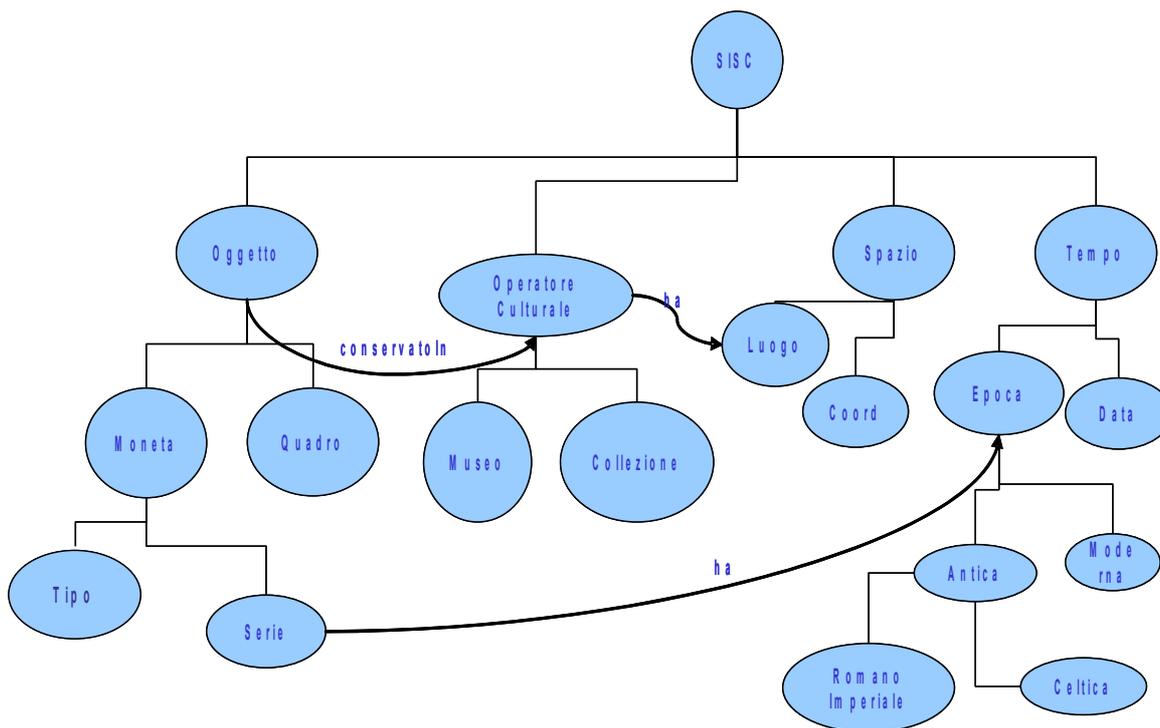


Figura 42: un semplice modello ontologico realizzato per la ricerca di monete

Oltre alle ovvie relazioni gerarchiche che legano i diversi concetti, su questo modello di ontologia c'è da dire quanto segue:

- Il concetto di oggetto (bene culturale) viene collegato al concetto di operatore culturale attraverso una relazione che serve ad affermare che: un oggetto è conservato da un operatore.
- Un operatore culturale è associato ad un luogo (normalmente il luogo in cui esso opera).
- La serie di una moneta viene associata al concetto di: epoca di conio.
- Spazio e tempo sono due concetti generali per i quali possono essere sviluppate ontologie *ad hoc*.

The screenshot shows a web browser displaying a museum catalog entry for 'Monete' (Coins) from the 'REGIONE DEL VENETO'. The entry includes a coin image, a dropdown menu for 'scegli l'immagine: 000359\_D', and descriptive text: 'Civico Museo "L. Bailo" - Treviso (TV) dracma (serie: celtica) Veneti Veneto 200 a.C. - 1 a.C. [Dritto] Reitia Testa di Reitia a destra con capigliatura caratterizzata da quattro grossi boccoli [Rovescio] Leone Leone gradiente a destra; sopra, pseudolegenda per <MASSA> AE (rame, bronzo, oricalco); diametro: 16; peso: 2.2'. Below the entry is a navigation bar with links from 'Precedente' to 'Successiva'. The browser's address bar shows 'http://beniculturali.regione.veneto.it/xway/application/crv/engine/crv.jsp'. To the right of the browser window is an ontology diagram with nodes: 'SISC', 'Spazio', 'Tempo', 'Luogo', 'Coord', 'Epoca', 'Data', 'Antica', 'Moderna', 'Romano Imperiale', and 'Celtica'. Arrows indicate relationships, such as 'ha' between 'Epoca' and 'Antica'. Below the browser window is a code editor showing the RDF/XML code for the entry, with line numbers 343 to 367. The code includes elements like <nu:DESM>, <nu:PVCC>, <sisc:conservatoIn rdf:resource="http://regione.veneto.it/sisc\_simulazione#museo\_138"/>, <nu:DTSI>, <nu:DTSF>, <nu:FTA\_IMG>, <rdf:type rdf:resource="http://regione.veneto.it/sisc\_simulazione#Moneta"/>, <rdf:Description>, <rdf:Description rdf:about="http://regione.veneto.it/nu#000359">, <nu:TSK>, <nu:OGTO>, <sisc:risaleA rdf:resource="http://regione.veneto.it/epoca#Celtica"/>, <nu:dritto>, <nu:rovescio>, <nu:DESM>, <nu:PVCC>, <sisc:conservatoIn rdf:resource="http://regione.veneto.it/sisc\_simulazione#museo\_51"/>, <nu:DTSI>, <nu:DTSF>, <nu:FTA\_IMG>, <rdf:type rdf:resource="http://regione.veneto.it/sisc\_simulazione#Moneta"/>, </rdf:Description>, and </rdf:RDF>.

4. passo 4 – le schede sono tradotte in formato RDF/XML<sup>61</sup> in funzione della ontologia di riferimento. Si passa da una rappresentazione esclusivamente accessibile ad un essere umano (la scheda da leggere) ad una rappresentazione accessibile alla macchina in cui il computer è in grado di riconoscere, relazionare ed elaborare i concetti contenuti nella scheda. La trasformazione delle schede in formato RDF/XML può essere fatta con tutti gli strumenti

61 Esistono molti formati per serializzare grafi RDF: N3, Turtle, etc. il formato XML/RDF è quello che è stato utilizzato per le risorse del SISC.

che sono stati sviluppati per l'XML<sup>62</sup>. Funzioni per la lettura, navigazione, ricerca e manipolazione di documenti XML sono disponibili per tutti i più moderni linguaggi di programmazione, in particolare per trasformare un documento XML con una data struttura in un documento XML con una struttura diversa. Il W3C ha definito l'XSL Transformations (XSLT)<sup>63</sup> che permette di descrivere in un documento formale le trasformazioni da un formato origine ad un altro. Sempre il W3C ha definito, questa volta nell'ambito del semantic web, il metodo Gleaning Resource Descriptions from Dialects of Languages (GRDDL) che serve per trasformare al volo (in tempo reale) un documento che ha al suo interno delle annotazioni, come può essere un documento XML. Tale documento può essere associato ad uno o più algoritmi di trasformazione (solitamente si usa XSLT) che permettono di rappresentarlo in RDF. La trasformazione al volo permette di mantenere la coesistenza dei diversi formati del medesimo documento. Una applicazione potrà in questo modo utilizzare il formato più idoneo per le proprie necessità.

5. passo 5 – a questo punto le schede tradotte in RDF/XML e l'ontologia redatta in OWL vengono caricate nel GeoBrowser assieme ad un ragionatore OWL che inferisce delle nuove istanze, come ad esempio:

- (la moneta x, haEpoca, Antica);
- Per tutte le monete che hanno epoca *RomanoImperiale e Celtica*.

<sup>62</sup> Il sito web della specifica è presente al seguente indirizzo: <http://www.w3.org/XML/>.

<sup>63</sup> Il sito web della specifica è presente al seguente indirizzo: <http://www.w3.org/TR/xslt>.

The screenshot displays the 'semantic geo browser' interface. At the top, it says 'semantic geo browser SISTEMA INFORMATIVO SEMANTICO DELLE RISORSE CULTURALI'. Below this is a search bar with 'Query', 'Mappa', and 'Lista' options. The main area shows a map of Northern Italy with a red dot indicating a coin location near Padua. To the right of the map is an ontology diagram with three blue circular nodes: 'Epoca' at the top, 'Moderna' in the middle, and 'Celtica' at the bottom, connected by lines. Below the map is a search results table with the following content:

1.	regione.veneto.it/nu#GRV-NU_0001232
Archeologico Nazionale di Adria	
sta	
terzio	
Imperiale	
[Dritto]	
vescio]	
leggibile/figura femminile stante a s.	
2.	regione.veneto.it/nu#GRV-NU_0001260
Archeologico Nazionale di Adria	
sta	
ondo	
Imperiale	
Archeologico Nazionale di Adria	
Rovescio: Pax [Rovescio]	
Descrizione: [leggibile/figura femminile a s. (Pax?)ha in mano ramo e cornucopia(?)	
	<a href="http://regione.veneto.it/nu#GRV-NU_0001679">http://regione.veneto.it/nu#GRV-NU_0001679</a>
Museo: Museo Archeologico Nazionale di Adria	
Oggetto: Moneta	
Tipologia: sesterzio	
Epoca: Romano Imperiale	
Dritto: Faustina II [Dritto]	

6. passo 6 – utilizzazione del GeoBrowser per la ricerca di informazioni. Si cercano le monete conservate nei musei del veneto in base alla “serie”. La serie è messa in relazione, tramite l'ontologia, all'epoca in cui la moneta è stata conosciuta. La relazione ontologica consente di cercare la moneta sia in base al contenuto informativo della scheda originale (richiedendo p.e. la serie o il museo che la contiene) sia in base ai concetti che compaiono nella ontologia (epoca di coniazione) ampliando notevolmente le possibilità espressive di chi effettua la ricerca.

## **Seconda applicazione - Ricerca di oggetti catalogati e utilizzazione dei vocabolari.**

Questo esempio descrive il flusso completo per la realizzazione di una applicazione funzionante in forma di prototipo.

1. passo 1 – vengono individuati: il “dominio di interesse” e lo “scopo dell'applicazione”. Anche in questo caso il dominio è quello della “conservazione di beni culturali” relativo ai seguenti oggetti: monete, reperti archeologici, beni demoetnoantropologici e fotografie, archiviati secondo gli standard ICCD nella banca dati dei beni culturali della Regione Veneto. Lo scopo della applicazione è quello di recuperare, all'interno del semantic web, le schede di tali oggetti, ampliando le possibilità espressive di chi effettua la ricerca attraverso l'impiego di MultiWordNet, un vocabolario on line.
2. passo 2 – vengono analizzate le schede informative degli oggetti che si intendono recuperare.
3. passo 3 – viene definita la Ontologia relativa al dominio di interesse, in funzione dello scopo applicativo.
4. passo 4 – le schede sono tradotte in formato RDF/XML in funzione della ontologia di riferimento.
5. passo 5 – viene eseguita una query sulla banca dati di MultiWordnet in modo da recuperare tutti i termini che hanno un legame con la parola immessa per la ricerca, per ogni termine ritrovato viene eseguita una richiesta all'interno dei dati a disposizione.
6. passo 6 – utilizzazione del GeoBrowser per la ricerca di informazioni. Si cercano oggetti conservati nei musei del veneto che rappresentino “la donna”. Il concetto di donna (che certamente potrebbe essere sviluppato a livello ontologico) è messo in relazione con concetti e termini affini, attraverso il vocabolario (donna, femmina, madonna, vecchia), ampliando anche in questo caso le possibilità espressive di chi effettua la ricerca.

The screenshot displays the 'semantic geo browser' interface. At the top, it reads 'semantic geo browser' and 'SISTEMA INFORMATIVO SEMANTICO DELLE RISORSE CULTURALI'. Below this, there are navigation options: 'Query', 'Mappa', and 'Lista'. A map of the Veneto region is shown with several items marked. Three detailed information cards are visible on the right side of the interface:

- Item 1:** [http://regione.veneto.it/bdm/CRV-BDM\\_0000729/](http://regione.veneto.it/bdm/CRV-BDM_0000729/)  
 Museo: Museo Archeologico Nazionale di Adria  
 Oggetto: Bene Demoetnoantropologico Materiale  
 Tipologia: cassetta per chiavi  
 Termine Dialettale: cassetta de i ciodi  
 Materiale: legno bronzo  
 Descrizione: L'oggetto ha forma di parallelepipedo a base rettangolare, internamente presenta due sezioni cave separate da un elemento verticale pure di sezione rettangolare. Sopra ad esso è alloggiato un elemento di forma riconducibile a un arco. In uno dei vertici superiori troviamo una breve catena inchiodata a una estremità mentre nell'altra è fissata una medaglia dal profilo esagonale con raffigurato da un lato San'Antonio da Padova e dall'altro la Madonna con Bambino e rosario.
- Item 2:** [http://regione.veneto.it/bdm/CRV-BDM\\_0001275/](http://regione.veneto.it/bdm/CRV-BDM_0001275/)  
 Museo: Museo dei Grandi Fiumi  
 Oggetto: Bene Demoetnoantropologico Materiale  
 Tipologia: dipinto  
 Classe: Madonna con Bambino  
 Materiale: carta vetro legno  
 Descrizione: quadro con cornice in legno e vetro; dipinto ad acquerello su precedente stampa
- Item 3:** [http://regione.veneto.it/nu#CRV-MI\\_0001232](http://regione.veneto.it/nu#CRV-MI_0001232)  
 Museo: Museo Archeologico Nazionale di Adria  
 Oggetto: Moneta  
 Tipologia: sestertio  
 Epoca: Romano Imperiale  
 Dritto: Adriano [Dritto]  
 Rovescio: [Rovescio]

## L'AGENTE

In questo esempio si descrive una possibile applicazione per la costruzione di percorsi e di palinsesti in maniera assistita.

### Terza applicazione - Percorso degli eventi dedicati a Giorgione

1. passo 1 – vengono individuati: il “dominio di interesse” e lo “scopo dell'applicazione”. In questo caso il dominio è quello degli “eventi culturali”. Lo scopo della applicazione è quello di consentire all'utente di costruirsi un percorso di visita, dato un determinato ambito territoriale (p.e. Il Veneto) e un determinato arco temporale (p.e. i primi sei mesi del 2010).
2. passo 2 – vengono individuati i concetti fondamentali che debbono essere trattati nelle schede informative degli eventi che si intendono trattare. Per esempio: titolo dell'evento, descrizione, periodo, luogo, organizzatore.
3. passo 3 – viene costruita la Ontologia relativa al dominio di interesse e in funzione dello scopo applicativo.
4. passo 4 – le schede sono implementate direttamente da chi propone l'evento. Il formato delle schede deve contenere i concetti richiesti. Le schede sono

tradotte in RDF/XML in funzione della ontologia di riferimento.

5. passo 5 – vengono caricate le schede in RDF/XML e l'ontologia sviluppata nel passo tre. Con l'impiego del ragionatore possiamo inferire nuove istanze.
6. passo 6 – utilizzazione del GeoBrowser per la ricerca di informazioni. L'utente effettua la query inserendo: oggetto della ricerca (Giorgione), periodo di visita, ambito territoriale. Il risultato può essere organizzato in diversi modi: lista, percorso, palinsesto. La ricerca - anche in funzione di come sono memorizzati i dati e di come viene costruita l'ontologia – può prevedere vincoli di qualsiasi tipo: budget, tempo atmosferico, gusti personali, mezzi per gli spostamenti etc..

## 9.2 IL PROGETTO: ADRIAMUSE

AdriaMuse è un progetto di cooperazione internazionale cofinanziato dalla Unione Europea, che coinvolge i paesi si affacciano sulle sponde del mare Adriatico. La descrizione di questo progetto, iniziato nel 2010, avviene riportando integralmente il *paper* “*Adriamuse.com: a new communication platform*”<sup>64</sup> redatto alla fine del 2012 nel quale viene descritto il funzionamento della piattaforma *Adriamuse.com* (consultabile in [www.adriamuse.org](http://www.adriamuse.org)) realizzata da Iuav nell'ambito del *Work Package 5 “Enlarging the network – IT tools for promoting museums and events”*.

Oltre a questo *paper* viene allegato il rapporto “*Guidelines for implementing IT tools in the tourism and cultural sector*” presentato nel gennaio 2014 nell'ambito del *Work Package 4 “Harmonizing the strategy, developing common methodology and tools”*.

## ADRIAMUSE.COM: A NEW COMMUNICATION PLATFORM

### INTRODUCCION

The AdriaMuse project evolved from the IPA Adriatic 2007-2013 Cross-Border Cooperation Programme. Co-financed by the European Union, it comprises 11 project partners, from five bordering countries along the Adriatic Sea shores. The participants include, on the western coast of Italy, the project leader - the Province of Rimini, the Artistic, Cultural and Natural Heritage Institute of the Emilia-Romagna Region; the IUAV University of Venice, the Veneto Region; the Province of Pesaro and Urbino; the Province of Campobasso; Skupa Ltd, C from the Molise Region. On the eastern Adriatic coast participants include the National Museum of Montenegro, the Albanian Municipality of Shkodra, the Government Service of Zenica-Doboj in Bosnia Herzegovina, and the Region of Istria in Croatia.

The strategy behind the AdriaMuse project is to strengthen and consolidate the partnerships as well as to promote sustainable development within the Adriatic area.

---

<sup>64</sup> autori Vincenzo Giannotti e Ivano Boscolo Nale

These goals can be achieved by harmonizing any cultural tourism-related activities among the various partners, adopting two main types of actions. Ones that:

- encourage museums within the Adriatic area to be part of a network - Euromuse.net, an international showcase dedicated to gathering information related to European museums and major events, reaching a wider range of potential visitors;
- promote museum activities "beyond their walls", offering museum proposals even to those who do not go regularly to museums.

To attain both these goals, the network must not only be strengthened, but it must also be integrated. This implies adopting a set of information tools and communication services that help to create a common ground where knowledge, experiences and ideas can be shared among the partners.

IUAV's efforts have focused in this domain, to build the activities whose profile is described herein.

## **IUAV'S ROLE**

IUAV has devoted its efforts to implementing Work Package 5, "Enlarging the network – IT tools for promoting museums and events".

The main goal of WP5 is to make it easier to access and to consult information related to the events being organized by the Museums of the Adriatic Area. Another equally important goal is to be able to disseminate this type of information via web, using state-of-the-art tools and technology, including semantic and collaborative devices.

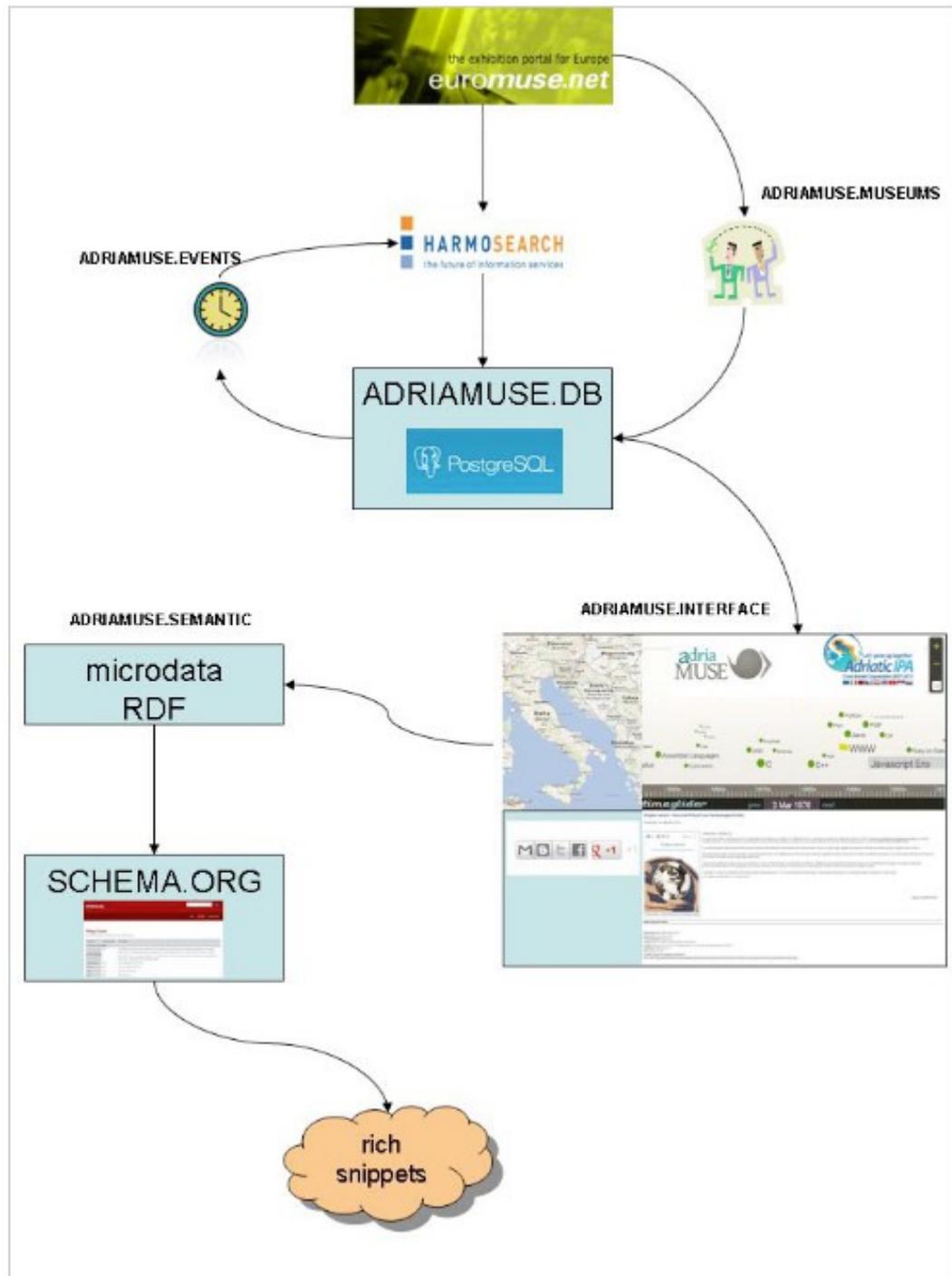
Data is accessed through the project's main portal that integrates a specific communication and collaboration platform called [Adriamuse.com](http://Adriamuse.com).

The following chapters describe the [Adriamuse.com](http://Adriamuse.com) platform and the technical requirements needed for its implementation.

## THE ADRIAMUSE.COM PLATFORM

### THE CONCEPT MODEL

The diagram below illustrates the architecture of Adriamuse.com:



The museums participating in the Adriamuse project are required to join the Euromuse.net and to load all information related to the events (Exhibitions in Euromuse) they are organizing and wish to promote via the network.

Two application modules transfer the information from Euromuse to Adriamuse:

- **Adriamuse.events** is an application module that uses the web services provided by the Harmosearch project <sup>65</sup>. It allows periodic acquisition of data from Euromuse.net. Data collection is done according to the Harmonise ontology model <sup>66</sup> and then stored in a local database called **Adriamuse.db**.
- **Adriamuse.museums** is a semi-automatic procedure that allows users to feed the Adriamuse.db database with the Museum Identity Records related to the events. However, Identity Records cannot be captured using the Harmosearch web services, since the project does not provide this feature. The Adriamuse.museum procedure stems from the collaboration between the IUAV development team and the Euromuse management team.

Users can access the data, arranged in the above mentioned way, through two additional modules:

1. **Adriamuse.interface**, a module that allows users to query the database asking information on events or on the Museum Identity Records, using two tools: a map and a timeline. Specific social buttons enable users to share events, submit their comments, and read the comments of others.
2. **Adriamuse.semantic** - the representation of events, accessed through the Adriamuse interface, is made using HTML5 microdata<sup>67</sup>, according to Sche-

<sup>65</sup> The aim of the HarmoSearch project (Harmonised Semantic Meta-Search in Distributed Heterogeneous Databases) is to analyse and develop a platform-independent search mechanism for tourism services with heterogeneous data bases.

<sup>66</sup> The Harmonise project has defined a reference ontology for actors operating in the tourism industry which allows them to exchange information according to conceptual domains.

<sup>67</sup> Microdata is a WHATWG HTML specification used to nest semantics within existing content on web pages.

ma.org<sup>68</sup>. Semantic enrichment, understood by specific software, can enhance the data related to events.

This solution is very useful since most of the common search engines, in the query results, manage to highlight even the information represented in microdata (Rich snippets<sup>69</sup>).

### Thing > Event

An event happening at a certain time at a certain location.

Property	Expected Type	Description
<b>Properties from Thing</b>		
additionalType	URL	An additional type for the item, typically used for adding more specific types from external vocabularies in microdata syntax. This is a relationship between something and a class that the thing is in. In RDFa syntax, it is better to use the native RDFa syntax – the 'typeof' attribute – for multiple types. Schema.org tools may have only weaker understanding of extra types, in particular those defined externally.
description	Text	A short description of the item.
image	URL	URL of an image of the item.
name	Text	The name of the item.
url	URL	URL of the item.
<b>Properties from Event</b>		
attendee	Person or Organization	A person or organization attending the event.
attendees	Person or Organization	A person attending the event (legacy spelling; see singular form, attendee).
duration	Duration	The duration of the item (movie, audio recording, event, etc.) in <a href="#">ISO 8601 date format</a> .
endDate	Date	The end date and time of the event (in <a href="#">ISO 8601 date format</a> ).
location	Place or PostalAddress	The location of the event or organization.
offers	Offer	An offer to sell this item—for example, an offer to sell a product, the DVD of a movie, or tickets to an event.
performer	Person or Organization	A performer at the event—for example, a presenter, musician, musical group or actor.

### Tacoma Dome Calendar - Tacoma Dome Events | Eventful

View **Tacoma Dome's** upcoming **event** schedule and profile - Tacoma, WA. City/neighborhood: Tacoma Disabled access: No obstacles to access.

Wed, Jun 29 **Britney Spears - Jessie and ...** - Tacoma Dome, Tacoma, WA, 98421

Fri, Jul 8 **Matthew Morrison** - Tacoma Dome, Tacoma, WA, 98421

Fri, Jul 8 **NKOTBSB** - Tacoma Dome, Tacoma, WA, 98421

[eventful.com/tacoma/venues/tacoma-dome-/V0-001.../events](http://eventful.com/tacoma/venues/tacoma-dome-/V0-001.../events) - Cached - Similar

*A query was made to the calendar of events. Using Schema.org and HTML5 microdata, the search engine, even at this level, can provide some information about the contents: where and when the event takes place, who will be featured in the upcoming events.*

<sup>68</sup> Schema.org provides a collection of shared vocabularies that can be used to mark up web pages in ways that can be understood by the major search engines: Google, Microsoft, Yandex and Yahoo!

<sup>69</sup> Snippets are the few lines of text that appear in the search results and give the user a sense of the content of the page and its relevance to the query made.

The data represented in Microdata and also in RDF<sup>70</sup> uses an *ad hoc* procedure. RDF is a standard model for data interchange on the Web, specified by the World Wide Web Consortium (W3C) with the purpose of making the Semantic Web<sup>71</sup> and Linked Data<sup>72</sup>.

## COMPONENTS

### Harmosearch services

The HarmoSearch technology enables all users who have joined the service to exchange data using specifically created web services. In our particular case, the data exchanged as part of the Adriamuse objectives, relates to the theme of cultural tourism.

Registration to this new beta-version service is currently free of charge.

HarmoSearch web services allow three types of actions:

- PushData: users can exchange data with other users;
- SimpleSearch: through simplified searching criteria users can search the data of other users;
- AdvancedSearch: advanced search criteria allows a user to search the data of another user.

At present, the users of HarmoSearch services also include EuroMuse.net and IUAV University, although their use is limited to the Event concept, as codified in the Harmonize ontology (which the users call "Exhibition").

Information related to each single Event includes the following:

1. id: name and identifier of the data provider (in this case, Euromuse)
2. eventTitle: title and sub title of the event

<sup>70</sup> RDF is a model for data interchange standards on the Web. <http://www.w3.org/RDF/>

<sup>71</sup> <http://www.w3.org/standards/semanticweb/data>

<sup>72</sup> <http://www.w3.org/standards/semanticweb/data>

- mainTitle (en + original)
  - subTitle (en + original)
3. description: description of the event
    - shortDescription
    - longDescription
  4. category: list of categories the event belongs to
  5. location: event location
    - name
    - address
    - country
    - city
  6. timeline: event calendar and timetable
  7. organizer: who organizes the event
    - name
  8. price: the price of tickets
  9. documentation: links to online resources such as photos, web pages ...

### **Adriamuse.db**

The local database Adriamuse.db was developed combining PostgreSQL + PostGIS. PostgreSQL is an open source object-relational database system, while the PostGIS project adds support for geographic objects to PostgreSQL, creating a spatial database for geographic information systems.

The local database, adriamuse.com, can help to be less conditioned by any potential problem that may arise to Euromuse.net or to the Harmosearch web services and in

addition, its use can improve the system's overall performance. Finally, the use of a local database makes it easier to construct interfaces and to organize the data according to the middle ontology schema.org.

The screenshot shows the phpPgAdmin interface. On the left is a tree view of the database structure, including 'public' and 'museum' tables. The main window displays a table with the following columns: 'Azioni', 'name\_local\_lang', 'local\_lang', 'short\_description\_local\_lang', and 'name\_eng\_lang'. The table contains 20 rows of museum data, including entries like 'Museo della bonifica', 'Museo Archeologico Oliveriano', 'Museo Civico', 'Centro culturale Mercato', 'Museo Internazionale del Vetro d'Arte e delle Ter...', 'Museo degli Sguardi', 'Musei Civici di Pesaro', 'MET Museo degli Usi e Costumi della Gente di Roma...', 'MUSAS Museo Storico Archeologico di Santarcangelo', 'Museo di Storia Naturale e Archeologia', 'Mateureka - Museo del Calcolo', 'Museo del PRECINEMA - Collezione Minici Zotti', 'Museo Sanitario di Campobasso', 'Museo della Città', 'Museo delle Valli e Oasi naturalistica di Campotto', 'Palazzo Moenigo', 'Ca' Pesaro - Galleria Internazionale di Arte Mod...', 'Torre dell'Orologio', 'Museo d'arte della città di Ravenna', 'Museo della Centuriazione Romana', 'Ca' Rezzonico - Museo del Settecento veneziano', 'Museo Corner', 'Palazzo Ducale', 'Museo Civico Archeologico Verucchio', 'Museo dei Bronzi Dorati e della Città di Pergo...', 'Casa di Carlo Goldoni', 'Museo del Merletto', 'Museo del Vetro', 'Museo di Storia Naturale', and 'Palazzo Fortuny'.

*the local Data Base Adriamuse.db*

## Capture of events

Event data is retrieved from Euromuse.net using the adriamuse.events software procedure. This procedure uses the Harmosearch web services to perform periodically the database updating.

To create Adriamuse.events, the following applications was developed:

- a client application, in Java, for the Harmosearch web services, used to retrieve event data from euromuse.net;
- a Php procedure for moving data into the local database.

## Acquisition of the Identity Record

Since Museum Identity Records cannot be acquired using the Harmosearch web services, data acquisition is based on a semi-automatic procedure, as set out by the Euromuse management team. The data is then transferred to the local database.

In addition, to properly represent the interface illustrated below, a PHP procedure was conceived. It connects the Identity data of museums with those related to the event.

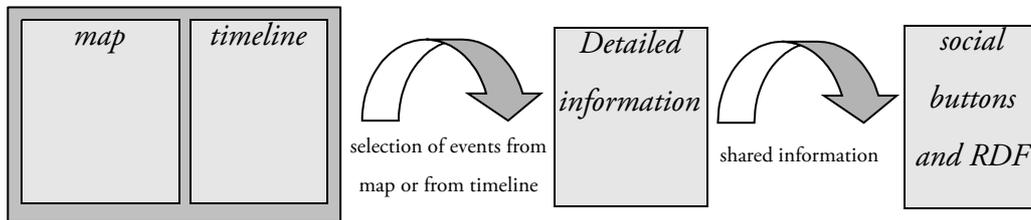
## Adriamuse.interface

The data is published on an interactive web browser interface which has a map and a timeline. Users will find information on the Museums (defined as a physical place where permanent or temporary exhibitions are organized) and events (exhibitions), both in English and in the local language. Museums are always associated to a place and will therefore show a pair of spatial coordinates represented with a symbol on a map. Events are always defined by temporal coordinates (a start and end date of the event) which are represented on a timeline. Events may have a spatial location, but not always (as in the case of web events).



To share information related to events, create reviews and RDF contents, some "2.0" tools can also be adopted.

The Adriamuse.interface is organized in the following manner:



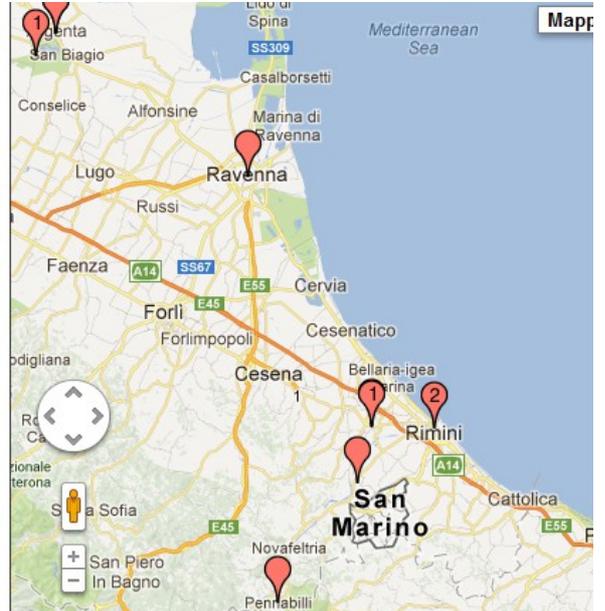
- icons related to museums and events appear on the map;
- icons show the number of events associated with a location;
- a timeline where users can set a time period (a month, a year ..) to see the events scheduled in that particular period, as represented along the time axis;
- because the map and timeline are related, when an icon is selected from the map, the corresponding event (the first, in case of multiple events) is centred in the timeline. In the same way, if an event is selected in the timeline, it is centred on the map;
- when an icon is selected from the map, a window balloon, with a short description of the event, opens, allowing users to access a detailed page.



**Data Presentation: map**

Map is an interactive tool which allows access to geographical information. The Map illustrates:

- the museums that have joined both Adriamuse and Euromuse;
- the icon indicating the museum shows whether the museum is associated with events;
- events organized outside the museum walls, not in the organizer’s location.



An event can be associated to a museum using the "location" concept, defined by the Harmonize ontology. Currently, although there have not been events outside the museum, the application can nonetheless fulfil this sort of requirement, as requested by the Adriamuse project, which strives to be able to organize exhibitions located somewhere other than the organizer’s museum. Map is provided by Google Maps, version 3.

**Data Presentation: Timeline**

Timeline is an interactive tool that allows users to browse information with a time reference.

Timeline shows all events, regardless of their location.

Even a "footloose" event (i.e. an event on the web) can be represented. When the duration of an event is undefined (i.e. a permanent



exhibition), it will simply be represented by an icon on the map.

Timeline was created using the widget software TimeGlider, a non-commercial license.

### **Linking GeoData to Time**

A geodatabase allows the direct mapping of events in parallel with a timeline as displayed on the map. The process operates as follows:

- accessing the system:
  - the map is automatically centred on "my city", where the user is;
  - the timeline is automatically centred on "today", the time the user is connected;
- all the events that include "today" are shown on the map;
- the timeline displays all the events visible in the map, around my city;
- panning or zooming the map, both the events shown on the map as well as the events shown in the timeline change;
- this feature can be disabled by using a specific button:



### **Museum event details**

By selecting an event/museum from the map, and the timeline, a new section opens up. It provides detailed information on the event and on the organizing body: title/name, address, long description, prices, photos etc...



## Mostra della Collezione

01 Jan 1990 - 01 Jan 2020

LE LUCERNE: Nonostante si tratti di una tipologia ormai ben identificata in Romagna e altrove, sono state rinvenute a Santarcangelo un gruppo consistente di lucerne che si caratterizzano per la loro specificità tecnica e decorativa. I tipi sono riconducibili a forme di imitazione delle lucerne africane che, sin dalla tarda età imperiale romana e poi fino al VI-VII secolo d.C. si diffusero con straordinario successo per tutto il Mediterraneo. Gli esemplari santarcangiolesi erano realizzati a matrice a doppia valva, una per la parte superiore e l'altra per il fondo, con decorazione sulla spalla e sul disco, molto stilizzata e ispirata alle coeve produzioni africane. I DIPINTI: La Madonna con il Bambino fra i santi Francesco e Giorgio di Luca Longhi, proviene dalla Chiesa di San Francesco, fondata nel XIII secolo fuori dalle mura, sull'antica via Emilia e sul limitare dell'area adibita ai mercati e alle fiere, e soppressa definitivamente nel 1862. La tavola è del 1531, firmata e datata in basso a sinistra, mentre a destra è raffigurato in completa armatura e inginocchiato, il committente Antonello Zampeschi, che ebbe in feudo Santarcangelo dal 1530 al 1534. Sempre in basso ma al centro, vi è lo stemma dei Zampeschi. Più recente è invece la Madonna con il Bambino di Guido Cagnacci, dipinto nel 1625. E' un'opera di particolare rilievo dato che si tratta di un raro dipinto giovanile che riflette il forte rinnovamento delle immagini sacre richiesto dal Concilio tridentino e dovuto all'impulso delle confraternite laicali e dei nuovi ordini religiosi.

### Opening Times

monday	00:00	00:00
tuesday	16:30	23:30
wednesday	16:30	19:30
thursday	16:30	19:30
friday	16:30	19:30
saturday	16:30	19:30
sunday	16:30	19:30

Organizer: IMC - Istituto dei Musei Comunali



### Gallery



### Social buttons and reviews

The page containing detailed information also has social buttons which can spread viral content on the web as: The Facebook Like button, the Twitter button, the Google +1 button, the RDF button (to export the events in RDF), the iCal button (to export the events in iCal) .



The above mentioned sharing tools help to disseminate the events published as part of the project.

In addition, Reviews express users opinion and can serve to assess the products and services. In our case, for instance, comments relate to an event in which the reviewer has participated in.

Your first-hand experiences really help other visitors.

★★★★☆ Your overall rating of this property

Very good the exposition and very good the place.

BBGVEJ  
Click to change

bbgvej

Send

Reviews

review 30 - by anonymous, 01 February, 2013  
★★★★★

Very interesting exposition. I suggest you visit it, better in weekday.

review 25 - by anonymous, 19 November, 2012  
★★★★★

Reviews help users to better identify pages with good content. Microdata and schema.org, enable Google to intercept information on reviews and comments within the content of a page. In a web search, this information leads to faster and more accurate access.

[Dragon Age: Origins for PC - Dragon Age: Origins PC Game - Dragon ...](#)  
★★★★★ Review by GameSpot - Nov 3, 2009  
Wii, **Dragon Age**, and Tiger Woods in this GameSpot news update for ... I had a great hope for **Dragon Age** Origins, but, it didn't turn out quite what I think ...  
[www.gamespot.com/pc/.../dragonage/index.html](http://www.gamespot.com/pc/.../dragonage/index.html) - 21 hours ago - [Cached](#) - [Similar](#)

### Adriamuse.semantic

The page containing the details of an event is written in the HTML5 standard. Semantics are provided by tagging the page with microdata, according to the schema.org ontology model. In addition, the same data can be produced in RDF

(RDF button).

The content of the web pages can therefore be easily intercepted by the search engines, improving visibility and, in the long run, even rankings.

### **Schema.org: Microdata and RDF**

Schema.org is a middle ontology model which provides a collection of shared vocabularies that webmasters use to mark up their pages in order to be grasped by most search engines as: Google, Microsoft, Yandex and Yahoo!

This model, using microdata, can provide semantic features to the web pages.

Microdata is used to tag information on web pages; Schema.org, in general terms, defines the way a specific resource, such as an event, must be structured.

## **CONCLUSIONS**

The first development phase of Adriamuse.com has been accomplished, currently delivered in the beta version. The communication platform will follow shortly and will be available from the main Adriamuse portal. The goal of the platform is the dissemination of information on the events that are organized by cultural institutions (Museums) operating within the Adriatic area who in turn become part of Adriamuse and Euromuse.net.

Given the project's overall architecture and its close ties to Adriamuse.com, Euromuse.net and Harmosearch, Adriamuse.com may very well improve the visibility of all the events which are currently being disseminated using the Euromuse.net channel. Although a personal viewpoint, the three project partners, in such case, would have to come up with an agreement. Adriamuse and euromuse can nonetheless be considered two complementary projects which, through a third project, harmosearch, could be a good example of rationalization applicable to different European project contexts as well.

An updated Adriamuse.com version is foreseen in the upcoming future, one that takes into account the growing number of events presented through this platform. A good presentation method should envisage not only the provider's needs (the

museum) but also the user's expectations (the tourist). The recommended techniques and technology systems should meet both their needs. In such a way, the cultural offer presented through [adriamuse.com](http://adriamuse.com) can be appreciated based on the user's expectations and on their personal profile.

This new and important line of research should be followed throughout this project and even implemented in future ones.

## **GUIDELINES FOR IMPLEMENTING IT TOOLS IN THE TOURISM AND CULTURAL SECTOR**

### **INTRODUCTION**

This document describes some of the recent standards, technologies and tools used in the fields of tourism and cultural marketing. The purpose of these tools is the enrichment of web content that any institution, company, person is able to produce, in order to ensure accessibility for automatic processing. In addition to this we introduce some application solutions useful for the promotion of events, always in the field of tourism and culture.

There is only one way nowadays to exploit the full potential of the Web: use the techniques and technologies of the Semantic Web. There are a lot of projects and communities, that are studying how to do the transition from the traditional web-based documents to the new web-based data and there are many institutions that also in tourism and culture, are actively participating in the implementation of the "linking open data" probably the most important project for the realization of the Semantic Web.

The international organization that promotes the development of linked data is the World Wide Web Consortium (W3C). It is also involved in development of standards and technologies, for the Semantic Web, which is defined as follow :

“In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.”

In these guidelines, we can't expect to fully investigate this very complex issue, constantly evolving in a very fast way: we only can give some suggestions on the most important technologies and applications and initiatives, to take into account. These are described in the next chapters.

## **BASICS TECHNOLOGIES**

### **RDF**

RDF (Resource Description Framework) - is a family of specifications published by W3C, and it was originally design as a metadata data model. Nowadays it is used as a general method for conceptual description or modeling of information that is implemented in web resources, in a way that computer applications can use and process it in a scalable manner.

The RDF data model is similar to classic conceptual modeling approaches such as entity–relationship or class diagrams, as it is based upon the idea of making statements about resources (in particular web resources) in the form of subject-predicate-object expressions.

These expressions are known as triples in RDF terminology. The subject is the resource; the predicate expresses a relationship between the subject and the object; the object is another resource or a value. For example, the statement "Titian is a painter" in RDF is as the triple: a subject denoting "Titian" is the subject, "is a" is the predicate, "painter" is the object.

RDF represents the basics for the semantic web implementation.

### **OWL**

OWL (Ontology Web Language) - is intended to be used when the information contained in documents needs to be processed by applications, as opposed to situations where the content only needs to be presented to humans. OWL can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms. This representation of terms and their interrelationships is called an ontology.

You can find a lot of predefined ontologies on the web: some of these are quickly described in the follow. OWL is a recommendation of the World Wide Consortium.

## **SKOS**

SKOS (Simple Knowledge Organization System) - is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web.

SKOS provides a standard way to represent knowledge organization systems using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be passed between computer applications in an interoperable way.

Using RDF also allows knowledge organization systems to be used in distributed, decentralised metadata applications. Decentralised metadata is becoming a typical scenario, where service providers want to add value to metadata harvested from multiple sources.

Essentially SKOS provide a bridge between the world of knowledge organization systems - including thesauri, classifications, subject headings, taxonomies, and folksonomies - and the linked data community, bringing benefits to both. Libraries, museums, newspapers, government portals, enterprises, social networking applications, and other communities that manage large collections of books, historical artifacts, news reports, business glossaries, blog entries, and other items can now use SKOS to leverage the power of linked data.

## **SPARQL**

SPARQL (SPARQL Protocol and Rdf Query Language) is an RDF query language, that is a query language for databases, able to retrieve and manipulate data stored in RDF format.

GeoSPARQL defines filter functions for geographic\_information\_system.

You can find a lot of SPARQL implementation. An example you can test is the Dbpedia “Virtuoso SPARQL Query Editor” by OpenLink Software

(<http://dbpedia.org/sparql>).

## Microdata

Microdata is a WHATWG<sup>73</sup> HTML specification used to nest metadata within existing content on web pages. Search engines can benefit from direct access to this structured data because it allows them to understand the information on web pages and provide more relevant results to users. Microdata uses a supporting vocabulary to describe an item and name-value pairs to assign values to its properties. Microdata is an attempt to provide a simpler way of annotating HTML elements with machine-readable tags than the similar approaches of using RDF.

Web developers can design a custom vocabulary or use vocabularies available on the web. A collection of commonly used markup vocabularies are provided by Schema.org (described in the follow) schemas which include: *Person, Event, Organization, Product, Review, Review-aggregate, Breadcrumb, Offer, Offer-aggregate*.

## ONTOLOGIES AND VOCABULARIES

### MACE

MACE (Metadata for Architectural Contents in Europe) <http://www.mace-project.eu/index.php> is a pan-European initiative for the interconnection and the dissemination of digital information in the domain of architecture. They say:

*“We connect various repositories of architectural knowledge and enrich their contents with metadata. The result are unique services for searching and browsing architectural contents, for instance, by conceptual connection, geography, language”.*

IUAV University participated to the Mace project as coordinator of the taxonomy development, on which the metadata system is based on, in collaboration with the Getty Research Institute in Los Angeles which has provided the *“repertoire of Art & Architecture Thesaurus (AAT)”*.

---

<sup>73</sup> Web Hypertext Application Technology Working Group

## **ISO 21127**

ISO 21127 is “*A reference ontology for the interchange of cultural heritage information*”.

This standard ISO 21127 derives from the CIDOC Conceptual Reference Model (CRM) which provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

The CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives.

## **Dublin Core Metadata Initiative (DCMI)**

This project started in 1995 during a workshop held in the city of Dublin, Ohio, which was intended to define a set of basic elements with which to describe digital resources. Following the launch of this initiative in 1996, was defined as a first set of 15 basic elements - the Dublin Core Metadata Element Set - which includes, among others: the Title, the Author, the Subject, Description, the Source identifier and so on.

Starting in 2000, the Dublin Core community focused on "application profiles" -- the idea that metadata records would use Dublin Core together with other specialized vocabularies to meet particular implementation requirements. During that time, the World Wide Web Consortium's work on a generic data model for metadata, the Resource Description Framework (RDF), was maturing. As part of an extended set of DCMI Metadata Terms, Dublin Core became one of most popular vocabularies for use with RDF, more recently in the context of the Linked Data movement.

## **WordNet**

WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the browser. WordNet is also freely and publicly available for download. WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms– strings of letters– but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity. These kind of relationships can be interpreted as specialization relations between conceptual categories. In other words, WordNet can be interpreted and used as a lexical ontology in the computer science sense.

WordNet was created at the Cognitive Science Laboratory of Princeton University.

## **Harmonize**

The European Project Harmonise aims at building a technological infrastructure based on a shared ontology, to enhance the cooperation of European SMEs in the tourism sector. They say:

*“Harmonise enables seamless data exchange by following a semantic mapping approach. Instead of a fixed standard, Harmonise provides a reference model, the Harmonise ontology, which local data formats can be mapped against and in this way mapped to other local data formats or standards. In the sense of such a reference model, the Harmonise ontology defines all relevant concepts and relationships between concepts of the problem domain. A specific mapping mechanism and reconciliation engine then enable*

*the translation of data from one local format to another and enable seamless data exchange and global interoperability.*<sup>74</sup>

Harmonize Ontology was used in the Adriamuse Project in order to link the Adriamuse.com platform to the Euromuse.net.

### **FOAF (Friend of a Friend)**

*“The Friend of a Friend (FOAF) project is creating a Web of machine-readable pages describing people, the links between them and the things they create and do; it is a contribution to the linked information system known as the Web. FOAF defines an open, decentralized technology for connecting social Web sites, and the people they describe.”*<sup>75</sup>

FOAF is a descriptive vocabulary expressed using the Resource Description Framework (RDF) and the Web Ontology Language (OWL). It provides the most used ontology to describe persons and their relationships.

## **PROJECTS AND IT TOOLS**

### **Harmosearch**

Electronic data exchange has been an important task within the field of travel and tourism already for decades, since it is a highly information-dependent industry. What has been a very technical job for a few large players in the past has become a challenging task for thousands of small companies today, due to the lack of a unique international data standard.

HarmoSearch will continue ongoing research in overcoming this interoperability problem by introducing tools to easily connect to an open data network and to find relevant information sources. An intelligent mapping tool, automatic translation of data queries and semantic registries about data providers will make it very easy – especially for SMEs – to exchange data and participate in the global online business

---

74 From “Harmonization, Interoperability and Standard” - Euromuse.net project

75 From <http://www.foaf-project.org/>

in a cost-efficient way.

HarmoSearch is based on the work of past projects and activities, like Harmonise, Harmo-TEN, a CEN Workshop Agreement, activities of the non-profit HarmoNET association, the euromuse.net project and portals like e.g. VisitEurope.com.

Harmosearch technology is going to be used in the Adriamuse Project in order to exchange data between Euromuse.net and Adriamuse platform.

### **Europeana**

Europeana.eu is an internet portal that acts as an interface to millions of books, paintings, films, museum objects and archival records that have been digitised throughout Europe. More than 2000 institutions across Europe have contributed to Europeana and more than 30 million digital object are collected, at the moment.

*“The Europeana Semantic Elements (ESE) format provides a basic set of elements for describing objects in the cultural heritage domain in a way that is usable for Europeana. The current version of the ESE is an updated version of the one that has been used from the start of the Europeana prototype in November 2008. It is a Dublin Core-based application profile providing a generic set of terms that can be applied to heterogeneous materials thereby providing a baseline to allow contributors to take advantage of their existing rich descriptions.”<sup>76</sup>*

### **Schema.org**

Schema.org is an initiative launched on 2011 by Bing, Google and Yahoo! with the aim of creating and supporting a common set of schemas for structured data markup on web pages. They propose using their ontology (Schema.org ontology) and Microdata in HTML5, to mark up website content with metadata about itself. Such markup can be recognized by search engine spiders and other parsers, thus gaining access to the meaning of the sites. They say:

*“Many sites are generated from structured data, which is often stored in databases. When this data is formatted into HTML, it becomes very difficult to recover the original*

<sup>76</sup> From “Europeana Semantic Elements Specification and Guidelines” 14/07/2013

*structured data. Many applications, especially search engines, can benefit greatly from direct access to this structured data.*“

*“A shared markup vocabulary makes it easier for webmasters to decide on a markup schema and get the maximum benefit for their efforts. So, in the spirit of sitemaps.org, search engines have come together to provide a shared collection of schemas that webmasters can use.”*

## **DBpedia**

*“DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. DBpedia allows you to ask sophisticated queries against Wikipedia, and to link the different data sets on the Web to Wikipedia data.”*<sup>77</sup>

In September 2013, version 3.9 of DBpedia was released. The dataset describes four million entities, out of which 3.22 million are classified in a consistent ontology, including 832,000 people, 639,000 places, 116,000 music albums, 78,000 films, 18,500 video games, 209,000 organizations, 226,000 species and 5,600 diseases. The DBpedia data set features labels and abstracts for these entities in up to 119 different languages; 24.6 million links to images and 27.6 million links to external web pages; 45 million external links into other RDF datasets, 67 million links to Wikipedia categories, and 41.2 million YAGO2 categories. The DBpedia project uses the Resource Description Framework (RDF) to represent the extracted information and consists of 2.46 billion RDF triples, 470 million extracted from the English edition of Wikipedia, 1.98 billion from other language editions, and about 45 million as links to external data sets.

If you visit the Dbpedia web site (<http://dbpedia.org>) you can find a wide set of IT tools and web application, useful to use Dbpedia and for semantic web.

DBpedia has been described by Tim Berners-Lee as one of the more famous parts of the decentralized Linked Data effort (see the next chapter).

---

<sup>77</sup> From Dbpedia <http://wiki.dbpedia.org/About>

## Linking Open Data Project

The concept of Linked Data was originally proposed by Tim Berners-Lee in his 2006 Web architecture note: *“The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data. Like the web of hypertext, the web of data is constructed with documents on the web. However, unlike the web of hypertext, where links are relationships anchors in hypertext documents written in HTML, for data they links between arbitrary things described by RDF. “*

Technically speaking, Linked Data refers to data published on the Web in such a way that it is machine readable, its meaning is explicitly defined, it is linked to other external datasets, and it can in turn be linked to from external datasets as well. Conceptually, Linked Data refers to a set of best practices for publishing and connecting structured data on the Web. A big graph illustrates the status of the dataset published in linked data format, both by the Linking Open Data community and other contributors. The last update was in September 2011, however, this diagram is greatly interesting since, by accessing the website that offers it (<http://lod-cloud.net/>), you can click on this to access the same image to a new image clickable from which you can reach any of the metadata of a dataset. It should be noted that the DBpedia dataset is thicker and more "connected" to those offered in the diagram. Often DBpedia is considered as an entry point to the network of linked data.

## Protégé

Protégé is a free, open-source platform that provides a growing user community with a suite of tools to construct domain models and knowledge-based applications with ontologies. At its core, Protégé implements a rich set of knowledge-modeling structures and actions that support the creation, visualization, and manipulation of ontologies in various representation formats. It can be customized to provide domain-friendly support for creating knowledge models and entering data. Further, Protégé can be extended by way of a plug-in architecture and a Java-based Application Programming Interface (API) for building knowledge-based tools and applications.

The Protégé-OWL editor enables users to build ontologies for the Semantic Web, in particular in the W3C's Web Ontology Language (OWL), and provides several tools for reasoning and research via SPARQL.

Protégé was developed by Stanford University, and it is released under the MPL (Mozilla Public License).

### **Apache Jena**

Apache Jena (or Jena in short) is a free and open source Java framework for building semantic web and Linked Data applications. The framework is composed of different APIs interacting together to process RDF data. It provides a programmatic environment for RDF, RDFS and OWL, SPARQL, GRDDL, and includes a rule-based inference engine.

### **9.3 IL PROGETTO: MONITORAGGIO E CONTROLLO PARTECIPATO DELLA MOBILITÀ**

Questo progetto viene descritto riportando un significativo estratto del *paper* "Il monitoraggio e la gestione, attraverso servizi LBS, di eventi diffusi sul territorio"<sup>78</sup> presentato in occasione della "Smart city exhibition" tenutasi a Bologna nel ottobre 2013.

#### **ABSTRACT**

Lo sviluppo di nuove tecnologie d'informazione digitale, smartphone e tablet, connessi e dotati di sistemi di posizionamento hanno avuto un forte impulso nel settore turistico creando la figura del "Turista Smart". Attraverso la realizzazione di due applicazioni per smartphone e di una piattaforma per la gestione del tracciamento degli utenti, è stata analizzata, a scopo sperimentale, la "geografia" di due iniziative culturali producendo analisi post-evento in cui il tracciamento dei partecipanti rappresenta una significativa opportunità per la definizione di servizi Location Based di gestione dei flussi turistici.

#### **INTRODUZIONE**

Nell'ambito di un POR FESR 2007-2013 attività 1.1.a2, finanziato dalla Regione Friuli Venezia Giulia, è stato realizzato il progetto "Monitoraggio e controllo partecipato della mobilità per l'erogazione di location based services", attraverso una collaborazione tra L'Università IUAV di Venezia, la società Movendo s.r.l. e Unisky s.r.l., spin-off Iuav. Scopo del progetto è stato lo studio di nuove applicazioni nell'impiego e nell'erogazione di Location Based Services (LBS). Tale attività ha avuto risvolti prettamente legati alla ricerca applicata ma non ha tralasciato di approfondire l'ampia gamma di servizi che l'impiego di questi sistemi può consentire di erogare sul tema del turismo e in particolare sulla figura del "Turista Smart" aprendo nuove possibilità nel campo dei viaggi organizzati, dei tour operator e delle grandi fiere distribuite sul territorio, sino ad interessare l'erogazione di servizi di gestione di flotte

---

<sup>78</sup> Autori Niccolò Iandelli, Vincenzo Giannotti, Luigi Di Prinzio

aziendali.

## **DEFINIZIONE E STRUTTURA DEI SERVIZI LOCATION BASED**

Due sono le definizioni più diffuse e condivise di Location Based Services:

“I Location Based Services sono i servizi di informazione accessibili da dispositivi mobili che usano la rete mobile e comunicano la posizione del dispositivo mobile”. (Virrantaus et al. [2001])

“Un servizio wireless-IP che utilizza informazioni geografiche per fornire servizi ad un utente mobile. Qualsiasi applicazione/servizio che sfrutta la posizione di un terminale mobile.”(OGC [2005])

Queste definizioni descrivono gli LBS come il risultato dell'intersezione di tre tecnologie: tecnologie dell'informazione e della comunicazione mobile (NTIC), Rete Internet e Sistemi Informativi Geografici (GIS) (Shiode et al. [2004]).

In generale l'erogazione di LBS presuppone l'interazione di due componenti: la prima componente è rappresentata dall'utente dotato di dispositivo mobile, in quanto strumento adatto a richiedere e a ricevere informazioni e servizi; la seconda componente è la rete mobile attraverso cui viene inoltrata la richiesta di informazioni/servizi al provider che a sua volta li fornisce di ritorno all'utente.

Ovviamente i servizi di cui stiamo parlando (location based) sono legati alla posizione dell'utente. Questa può essere ottenuta impiegando diverse tecnologie, sia locali, sia globali. I metodi di posizionamento locali a radio frequenza, soprattutto *bluetooth* e RFID, sono utilizzati per la navigazione interna ad ambienti, come ad esempio in un museo o in una fiera. Se la posizione non viene determinata automaticamente può anche essere specificata manualmente dall'utente. I metodi di posizionamento globali (il GPS è il più noto ma ne esistono altri) consentono di conoscere la posizione del dispositivo attraverso tecniche di triangolazione satellitare.

Conoscendo la posizione del dispositivo mobile e quindi dell'utente, il “Service e Application Provider” è in grado di offrire una serie di servizi e di informazioni. A titolo di puro esempio tali servizi possono includere: la ricerca di un percorso, la

ricerca di punti noti o informazioni rispetto alla posizione attuale, la guida dell'utente alla visita di un parco, il tracciamento dell'utente durante un trekking e la fornitura di statistiche sull'attività svolta, la raccomandazione di servizi ed eventi che si svolgono in zona.

## **LBS E PRIVACY**

Spesso, in tema di localizzazione e tracciamento, il timore degli utenti è che il terminale possa inviare anche informazioni personali e sui comportamenti, trasformandosi potenzialmente in uno strumento di controllo. In effetti, problemi legati alla privacy si possono avere quando un flusso dati viene collegato ad una persona e i dati raccolti vengono utilizzati senza autorizzazione per definirne un profilo. Il paradosso che si viene a creare, in questo caso, è dato dal fatto che gli stessi dati che possono contribuire a garantire il corretto funzionamento del servizio, rappresentano nel contempo un potenziale danno alla privacy.

Detto questo, la privacy può in ogni caso essere garantita, adottando alcuni dei principi già introdotti da Westin negli anni '70 quali l'*anonymity* (la localizzazione deve avvenire senza l'identificazione dell'utente, o meglio senza fornire informazioni reali sull'identità del target, impiegando per esempio codici cifrati o chiavi ID univoche), l'*intimacy* (l'utente deve sapere quali attori intervengono nelle operazioni di localizzazione e di analisi dei dati raccolti e deve inoltre poter decidere quali attori autorizzare) e la *reserve* (ogni tentativo di localizzazione deve essere segnalato – p.e. attraverso messaggi di alert – ed esplicitamente autorizzato dal soggetto tracciato). L'utente deve avere il pieno controllo sul trattamento dei propri dati e deve sapere in che modo e da chi sono utilizzate le informazioni che lo riguardano; vanno definite delle *policy* che indichino in maniera non ambigua a quali informazioni possono accedere le diverse entità e sotto quali vincoli.

## **LBS E TURISMO**

L'analisi dei pattern di movimento delle persone può avere importanti implicazioni nello sviluppo di infrastrutture, nella pianificazione dei trasporti, nella infomobilità e

anche nelle strategie di marketing del turismo; una dettagliata analisi dei flussi turistici e dei comportamenti del turista può fare la differenza in termini di gestione degli impatti sociali, ambientali e culturali che questi stessi possono comportare.

Nella maggior parte dei casi i problemi principali relativi alla localizzazione e alla stima del comportamento dei turisti sono legati al livello di accuratezza e / o alla validità dei dati raccolti. Storicamente il primo approccio nell'analisi dei flussi turistici è stato quello di monitorare il movimento delle persone tra due luoghi: dal punto di origine al punto di destinazione. L'obiettivo veniva perseguito analizzando i "diari dello spazio e del tempo" (Goodchild e Janelle, [1984]). Queste pratiche richiedevano che i soggetti fossero attivamente coinvolti nel processo di raccolta dei dati registrando, nel dettaglio, le loro attività durante l'intero esperimento. Spesso però a causa di errori nella registrazione delle posizioni i dati ottenuti erano di dubbia credibilità (Szalai, [1972]). Negli ultimi anni questi problemi sono in gran parte superati e il rapido sviluppo di dispositivi di tracciamento di piccole dimensioni, a basso costo ed alta affidabilità, ha portato ad un crescente volume di dati spaziali utilizzabili per la ricerca relativa ai flussi e al turismo. I dispositivi basati su GPS (Global Positioning System) offrono l'opportunità di avere dati ad alta risoluzione in termini di tempo (secondi) e di spazio (in metri), per lunghi periodi di tempo. Tuttavia, ad oggi, la maggior parte della ricerca condotta sulla base di materiale raccolto attraverso queste tecnologie, è stata legata per lo più a studi nel settore dei trasporti e della logistica. La raccolta di dati e lo studio delle attività delle persone e delle folle, per quanto interessanti, sono stati finora meno frequenti. La principale motivazione è data dal fatto che la raccolta di dati sui movimenti delle persone attraverso tecnologie *Location Based* (di seguito LB) è una pratica assai più complessa rispetto all'impiego delle stesse tecnologie su veicoli. Mentre per una vettura il sistema di tracking può essere considerato semplicemente "un accessorio in più" che è facilmente installabile e non influisce sui consumi o sulle caratteristiche del veicolo, nel caso dei pedoni il sistema di tracciamento deve essere necessariamente di piccole dimensioni e passivo, non intrusivo, in modo da garantire il normale comportamento del soggetto, e deve poter funzionare anche in ambienti complessi (strade strette, ambienti chiusi etc.): requisiti questi, spesso molto difficili da soddisfare.

Un passo avanti sostanziale è stato fatto con lo sviluppo della telefonia mobile che ormai ha raggiunto un grandissimo livello di pervasività che in Italia ha portato ad avere (dati 2009) 97 milioni di abbonamenti/carte ricaricabili di telefonia mobile a fronte di 60 milioni di abitanti. Inoltre, nell'ultimo biennio, si è avuto un vero e proprio boom dei telefoni cellulari di nuova generazione: smartphone dotati di dispositivo GPS e connessione web, che costituiscono ormai il 60% dei telefoni attualmente circolanti. Queste nuove tecnologie prospettano scenari interessanti per l'analisi di flussi di persone, dando per acquisita la tesi per cui ad un dispositivo corrisponde una persona.

## **SERVIZI DI RACCOMANDAZIONE PER IL TURISMO**

I Sistemi di Raccomandazione hanno lo scopo di fornire agli utenti consigli personalizzati e di loro probabile interesse riguardo ad un insieme di oggetti appartenenti ad uno specifico dominio (servizi per il turismo nel nostro caso), a partire da dati riguardanti le attese dell'utente in esame.

Da un punto di vista formale questo significa che un sistema di raccomandazione deve essere in grado di attribuire ad un dato utente un insieme di oggetti (servizi) su cui egli può o meno avere espresso una preferenza o una valutazione. Poiché non è detto che un tale utente abbia espresso una preferenza su un certo oggetto/servizio (il più delle volte è così), si tratterà di sviluppare degli algoritmi in grado di “predire” quella che sarà la scelta dell'utente, ovvero la valutazione che potrebbe dare di un certo servizio. L'ottenimento di tali valutazioni è uno dei problemi centrali della ricerca sui sistemi di raccomandazione in quanto nelle applicazioni tipiche un utente non esprime mai preferenze su ogni oggetto del catalogo a disposizione.

L'ottenimento delle valutazioni si realizza tipicamente implementando le seguenti soluzioni:

- *soluzioni content-based*: all'utente sono raccomandati oggetti simili a quelli che gli sono piaciuti nel passato;
- *soluzioni collaborative*: all'utente sono raccomandati oggetti che sono

piaciuti ad altre persone con gusti simili;

- *soluzioni ibride*: combinano entrambi gli approcci di cui ai punti precedenti.

E' chiaro come, per poter attuare le soluzioni di cui sopra, sia necessario partire da una base precostituita di valutazioni: queste possono essere ottenute attraverso un sistema di profilazione dell'utente che deve essere realizzato *ex ante* rispetto all'attivazione del sistema di raccomandazione e che dovrà essere messo a disposizione di quest'ultimo.

Nel caso del turismo e in particolare durante la fruizione *in loco* del bene turistico, assume grande importanza la possibilità di collocare l'utente in un determinato luogo in un determinato momento, come pure di valutare le modalità di utilizzazione del dispositivo mobile, con lo scopo di realizzare una raccomandazione di tipo contestuale, strettamente legata al luogo e ai comportamenti dell'utente, fornendo servizi avanzati anche in modalità *Augmented Reality*.

## **POSIZIONAMENTO IN AMBIENTI COMPLESSI**

Il posizionamento, la misura e la registrazione della posizione di un utente e/o turista, deve essere determinata in ambienti definiti "complessi": ambienti caratterizzati da scarsa visibilità dell'orizzonte, da luoghi senza visibilità del cielo (stanze chiuse), aree in cui ci sono interferenze di vario tipo. Tre sono le metodologie attualmente impiegate per raccogliere informazioni sul comportamento temporale e spaziale di persone o turisti: tecniche di osservazione diretta, osservazione a distanza e tecniche di osservazione indiretta. Tra le tecniche di osservazione diretta il "metodo dell'osservatore partecipativo", coinvolge l'osservatore che accompagna l'individuo sotto esame in prima persona. Questo approccio è ampiamente utilizzato nella ricerca antropologica, quando il ricercatore mira a portare la familiarità nella attività di ricerca. A questo si contrappone il "metodo dell'osservazione non partecipata". In questa tipologia l'osservatore può seguire il soggetto a una certa distanza, producendo molte informazioni, ma senza riuscire a svelare lo scopo e il significato che sta alla base delle decisioni e delle attività che i soggetti analizzati eseguono o

prendono. La seconda possibilità è legata all'osservazione a distanza, tecnica meno costosa dell'osservazione diretta che viene utilizzata per registrare e analizzare i flussi turistici già dagli anni '80. (Hartmann [1984]) è stato uno dei primi a usare questa tecnica a Monaco di Baviera, impiegando una telecamera posta in cima alla guglia del municipio a ottanta metri. Ha poi usato le immagini per stimare la percentuale di giovani turisti Nord americani tra il numero totale di persone che si radunano per vedere il Glockenspiel nel quadrato principale della Città Vecchia. Resta da chiedersi se l'identificazione dei turisti del Nord America, effettuata solo in base al loro aspetto, sia stata accurata. L'impiego di telecamere di video sorveglianza è sicuramente una tecnica di rilevazione a distanza efficace, però non sempre è facile ottenere i nastri di sorveglianza. Le immagini spesso appartengono all'autorità che posiziona le telecamere, di solito l'autorità comunale o di polizia, non sempre disponibile al rilascio delle immagini per scopi di ricerca ha notato che tutte le tecniche basate su rilevazione da telecamera, anche se efficaci per studiare il comportamento degli individui all'interno di una ristretta area, sono di scarsa utilità una volta che i pedoni/turisti escono dal campo visivo della ripresa. Questa critica vale per molti altri punti fissi di osservazione, sia che si utilizzi la fotografia "time-lapse", telecamere o registrazioni video (Hill [1984]; Helbing et al [2001]). L'impiego dei dati provenienti da reti di telefonia, rappresenta sicuramente una tecnica innovativa e con indubbi vantaggi, presenta però alcune debolezze. I telefoni cellulari sono in grado di fornire una fotografia oggettiva del comportamento dei soggetti ma non possono, per loro stessa natura, rivelare le motivazioni legate alle attività del soggetto. La terza metodologia si basa su metodi di osservazione indiretta tra i quali il "bilancio spazio-temporale" è il metodo più comune utilizzato per la raccolta dati sulla salute umana nelle scienze sociali in generale e negli studi del turismo in particolare. Il metodo si basa sulla registrazione sistematica dell'impiego del tempo di una persona su un dato periodo. Questo metodo non è osservazionale e si basa sull'impiego di diari o questionari spazio-temporali in cui si registra l'attività umana nel tempo e nello spazio, nel corso di un periodo di tempo limitato, di solito tra un giorno e una settimana. Il problema principale di questo metodo è la quantità e la qualità delle informazioni raccolte che dipende dalla capacità dei soggetti a ricordare eventi passati

e quale sia il grado di precisione e di dettaglio. Inoltre i questionari sono spesso molto brevi e contengono poche informazioni per timore che il soggetto perda la pazienza. Attualmente la miglior soluzione è rappresentata dall'integrazione di misure provenienti dalla rete cellulare e del sistema GPS. La tecnologia si basa sul fatto che le onde elettromagnetiche viaggiano a una velocità nota, la velocità della luce. Questo rende possibile il calcolo della distanza che il segnale ha percorso, attraverso il confronto tra l'ora esatta in cui è stato emesso il segnale e l'ora esatta in cui lo stesso segnale è stato ricevuto. Tuttavia, calcolare il tempo in modo molto accurato rende necessario l'uso di orologi atomici, estremamente costosi da inserire negli oggetti "riceventi". La tecnologia TOA non viene infatti impiegata nei sistemi di localizzazione radio terrestri dove si ha un elevato numero di stazioni che forniscono un servizio locale e non globale. Tale tecnologia viene impiegata nel sistema GPS (Global Positioning System) dove il numero di satelliti è più piccolo in proporzione alla zona servita. Una applicazione dei principi alla base della tecnica TOA è quella che si basa sullo scostamento dei tempi di arrivo. I sistemi che sfruttano la tecnologia dello scostamento del tempo di arrivo (TDOA) si basano su una serie di tre o più antenne a terra che raccolgono le trasmissioni dalle unità periferiche o portatili. Calcolando la differenza nel tempo che ha impiegato il segnale per viaggiare dall'unità portatile alle diverse stazioni, è possibile elaborare la differenza di distanza tra l'unità e le stazioni, definendo una curva iperbolica su cui l'unità può essere posizionata. Queste informazioni vengono poi trasmesse ad un sistema centrale che stabilisce la posizione dell'utente finale determinando il punto di convergenza dalle letture di almeno tre stazioni. La posizione può essere ottenuta anche da unità portatili e antenne fisse attive piuttosto che passive in attesa di ricevere i segnali dall'unità portatile. In questo caso, l'unità riceve il segnale da diverse antenne e può calcolare la sua posizione.

## **RETE CELLULARE**

L'uso commerciale delle comunicazioni cellulari, iniziato nel lontano 1983 negli Stati Uniti, era limitato, a causa del prezzo elevato sia del servizio sia dei dispositivi, principalmente all'ambito lavorativo. I costi della telefonia mobile hanno cominciato

a scendere drasticamente a metà degli anni '90, e oggi è ampiamente diffusa e a basso costo in tutto il mondo: anche nei paesi in via di sviluppo. Ormai possiamo considerare la società nel suo insieme come una società “Cellulare”.

Per la normale gestione di una rete telefonica cellulare è necessario che l'operatore di rete sia in grado di rilevare costantemente la vicinanza dell'abbonato ad una specifica antenna (Cella). Ciò consente all'operatore di trasmettere le chiamate in entrata e in uscita da e verso il cellulare dell'utente. Questa funzione consente anche il tracciamento del dispositivo, spesso in abbinamento con altre tecnologie in grado di migliorare l'accuratezza del rilevamento. L'esempio più ovvio è quello del GPS che ormai è quasi sempre integrato all'interno dei telefoni cellulari.

## PRINCIPALI TECNOLOGIE DI TRACCIAMENTO SATELLITARI

GPS, acronimo di *Global Positioning System*, è il sistema di posizionamento globale più diffuso che ha come scopo quello di determinare la posizione di un qualsiasi punto della superficie terrestre fornendo delle coordinate. Per fare questo il sistema si basa sul calcolo della distanza che intercorre tra l'antenna ricevente e un certo numero di satelliti utilizzando il tempo di percorrenza di un messaggio radio, che viene emesso da una costellazione di satelliti e captato dal ricevitore. La precisione complessiva dipende dalla configurazione geometrica dei satelliti, per migliorare la precisione si usano altre tecnologie definite dalle sigle: WAAS, EGNOS, MSAS e A-GPS. Tali sigle indicano dei sistemi di correzione che permettono di aumentare la precisione nel posizionamento. In particolare le prime tre si riferiscono ad un sistema satellitare, l'ultima invece si riferisce ad un sistema terrestre che utilizza altre fonti per migliorare il posizionamento. In questo modo è possibile ottenere precisioni di posizionamento che si avvicinano ai 2 metri, a seconda della qualità della configurazione dei satelliti nel momento della misura.

	GPS	TDOA, AOA	Cellular
<b>Precisione</b>	Alta	Media	Bassa
<b>Copertura</b>	Globale	Solo in zone con infrastrutture adeguate	Solo in zone con copertura cellulare

<b>Aree urbane/aree naturali</b>	zone rurali e aree aperte dove la vista del cielo è senza ostacoli.	Elevato costo delle infrastrutture, è conveniente solo in ambienti urbani, con un alto numero di potenziali utenti.	Più accurata nelle aree urbane dove la densità dei radiotrasmettitori è più alta
<b>Costi di utilizzo</b>	Nessuno. I segnali trasmessi dai satelliti sono gratuiti.	Il costo del servizio varia da luogo a luogo.	Il costo del servizio varia da luogo a luogo e seconda del fornitore ,del piano di fatturazione.
<b>Privacy</b>	La privacy non è un problema, l'utente controlla la propria visibilità.	La privacy degli utenti è discutibile: individuazione del utente in qualsiasi momento.	La privacy degli utenti è discutibile, individuazione dell'utente in qualsiasi momento.
<b>Real-Time</b>	Il sistema non può tracciare in tempo reale gli utenti. Gli utenti possono registrare e trasmettere la propria traccia,	Permette di tenere traccia dei soggetti in tempo reale.	Permette di tenere traccia dei soggetti in tempo reale.
<b>Frequenze radio</b>	Lavora ad alte frequenze assegnate appositamente per l'utilizzo globale.	Usa frequenze più basse, che possono, in linea di principio, essere utilizzate anche da altre parti.	Usa frequenze più basse, che possono, in linea di principio, essere utilizzate anche da altre parti.

Tabella 2 - Confronto tra le principali tecnologie di monitoraggio

## IL SISTEMA REALIZZATO

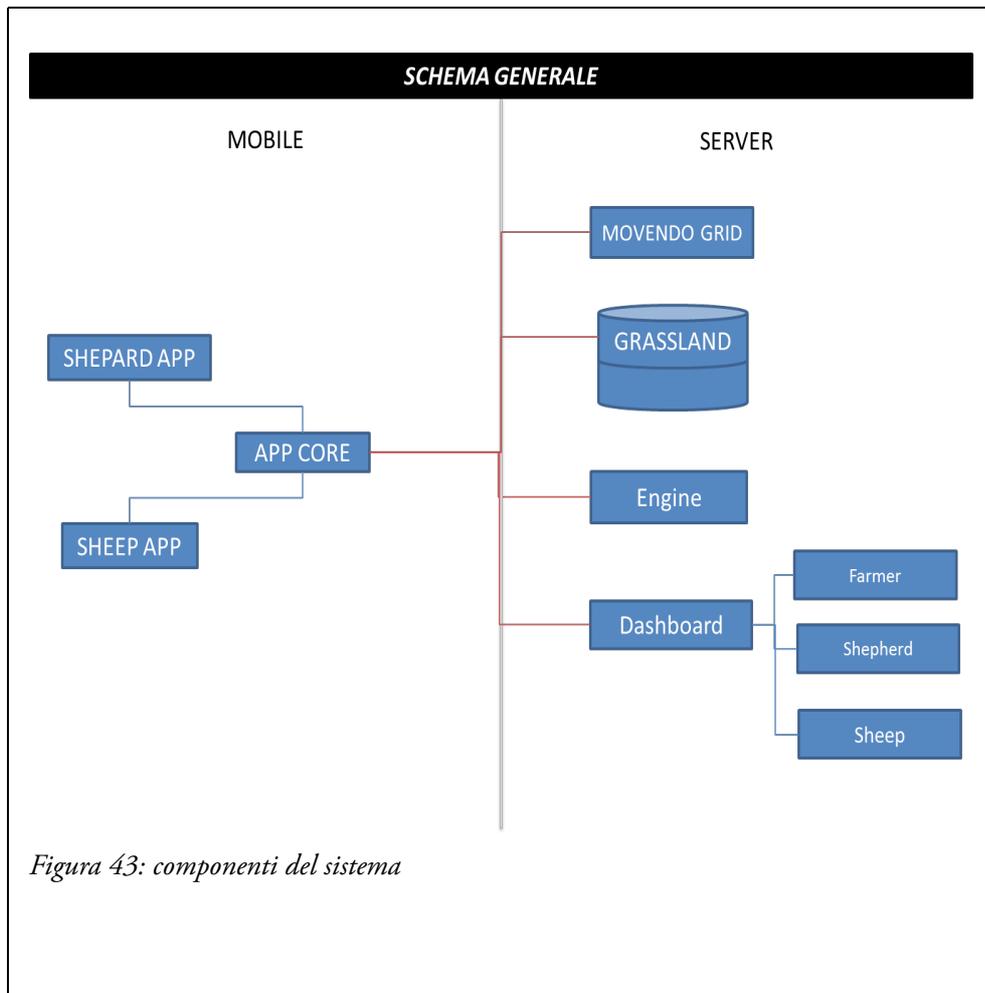
Sono stati realizzati due scenari di utilizzo (denominati *Free Tourism* ed *Expo*) con funzionalità specifiche sviluppate all'interno del *framework* per l'erogazione di *Location Based Services* (definibile anche *context awareness system*) sfruttando i parametri di posizione ricavati dai sensori installati a bordo degli *smartphones*.

La struttura del lavoro si basa sulla metafora di una fattoria, ogni nuova installazione del sistema rappresenta infatti l'apertura di una *Farm* in cui vengono definiti i ruoli degli utenti di sistema (la *Farm* è un'istanza di una particolare configurazione dei componenti del *framework*).

## UTENTI DEL SISTEMA

Sono identificati nel sistema tre tipologie di utenti, di seguito definite:

- *Farmer*: può essere identificato come il gestore di una specifica istanza della Farm (di seguito definita e identificabile ad esempio con una proloco, l'organizzatore dell'Expo/fiera ..etc.).
- *Sheperd*: è un utente che utilizza il sistema per monitorare un gruppo di utenti e comunicare con loro. La comunicazione può avvenire sia tra pari (*Shepherd to Shepherd*) sia verso utenti *Sheep*.
- *Sheep*: è l'utente che utilizza il sistema per un uso autonomo oppure per entrare a far parte di un "gregge" a cui fa capo uno *Shepherd*.



## COMPONENTE MOBILE

La componente mobile del sistema si compone di due APP con funzionalità distinte sviluppate su un APP *Core* comune. L'APP *Core* è il sistema di acquisizione delle coordinate, tecnologia di acquisizione e dati di contorno associati disponibili tra quelli nativi di Android. La *Shepherd\_App* è un applicativo Android con funzionalità dedicate all'utente *Shepherd*. Utilizzabile dalla figura di volontario/hostess (*Expo*), dalla figura del capogruppo (Turismo di gruppo) oppure dal gestore dei rappresentanti di azienda sul territorio. La *Sheep App* è un applicativo Android con funzionalità dedicate all'utente *Sheep*. Utilizzabile da un visitatore autonomo di un centro urbano (*Free tourism*), dall'avventore di una fiera (*Expo*) oppure da un rappresentante di azienda sul territorio.

## COMPONENTE SERVER

Lato server abbiamo diverse componenti *Engine*, *Dashboard*, *Movendo Grid* e *Grassland*. La componente *Engine* rappresenta il motore che interagisce con tutte le parti del sistema esegue le query e le elaborazioni richieste reindirizzandole alle varie componenti. Il *Engine* può essere acceduto da alcune oppure tutte le componenti dell'ecosistema (vedere le ipotesi alternative nel capitolo relativo ai flussi di informazioni) offrendo a queste dei *Web Services* (SOAP o REST) oppure delle apposite API riconoscendo le richieste tramite chiave oppure il più sofisticato OAuth (<http://hueniverse.com/oauth/>). Il vantaggio di utilizzare *Web Services* o API risiede nel fatto che la medesima funzionalità viene implementata una sola volta ed acceduta per esempio sia dalla componente *Mobile* sia dalla *Dashboard*. Le specifiche funzionalità saranno dettagliate negli appositi paragrafi in cui si descrivono le diverse tipologie di Farm disponibili. La *Dashboard* permette di accedere, da un'interfaccia web, in modalità differenziata in base alla tipologia di utente: *Farmer*, *Shepherd* e *Sheep*.

Il *Farmer* ha la piena gestione degli utenti fidelizzati e ne monitora i flussi tramite mappa e reports. Può definire un nuovo "gregge" e le figure *Shepherd* a loro guida e mettere in comunicazione tra loro gli *Shepherd*. Può aggiungere nuovi dati ai *dataset*

esistenti. Gestisce il profilo della *Farm* dove può pubblicizzare i servizi, il calendario degli eventi ed il download delle *Sheep* e *Shepherd* App (funzionalità CMS). Può associare sia agli utenti *Sheep* e *Shepherd* dei contenuti (es. coupon e buoni sconto) da utilizzare quando si trovano in determinati luoghi.

L'utente *Shepherd* ha una propria “area riservata” dove può monitorare il suo “gregge”, comunicare con ogni *Sheep* associata e comunicare con altri *Shepherd* a capo di un altro gregge. Può associare agli utenti *Sheep* appartenenti al suo gregge dei contenuti (es. coupon e buoni sconto) da utilizzare quando si trovano in determinati luoghi.

L'utente *Sheep* ha una propria area riservata dove può gestire il suo profilo, può monitorare gli spostamenti nel tempo e visualizzare delle statistiche.

La componente *MovendoGrid* è il sistema di stoccaggio della posizione degli utenti, già in uso da Movendo per la localizzazione dei mezzi tramite sistema GPS. Infine la *Grassland* che rappresenta l'insieme di tutti i *dataset* locali, *web services* e *API* di terze parti che servono al *Engine* per effettuare le elaborazioni.

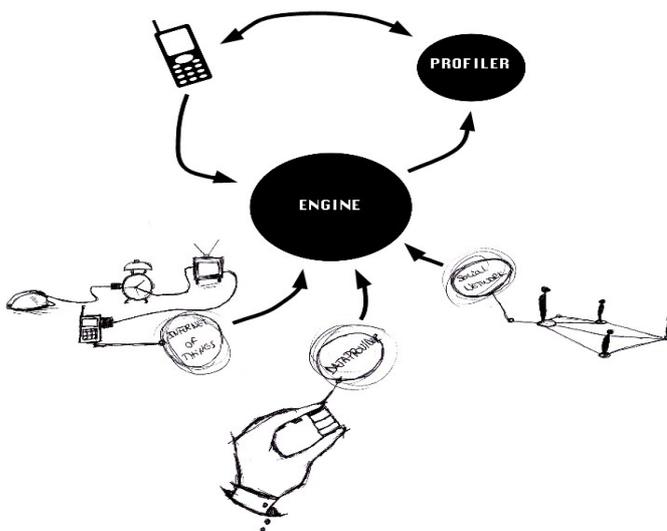
## 9.4 IL PROGETTO: RECOMMENDING TO PEOPLE

Come accennato nel cap. 3 della tesi, questo progetto è stato redatto con lo scopo di ottenere un finanziamento dal Ministero dell'Università e della Ricerca scientifica, in base all'articolo 11 del decreto ministeriale 8 agosto 2000 n. 593, per impiantare e avviare, attraverso uno spin-off universitario, un servizio di raccomandazione per il turismo culturale, da realizzarsi in prima battuta nella città di Venezia. Nel seguito viene riportato un breve estratto di descrizione generale del progetto, ricavato dal documento presentato per il bando. Poiché i risultati di questo progetto sono serviti come base per lo sviluppo della parte sperimentale della tesi, per maggiori approfondimenti si rimanda alla lettura di quest'ultima.

### DESCRIZIONE GENERALE DEL PROGETTO

Obiettivo generale del progetto è quello di realizzare un ecosistema digitale nel quale possano interagire diverse componenti (uomini, macchine, sensori, servizi...) con lo scopo di contestualizzare il turista culturale (ma più in generale qualunque turista) rispetto all'ambiente e allo scenario culturale in cui si trova e fornirgli suggerimenti e informazioni anche in base alle sue normali aspettative.

Il processo che sta alla base del servizio proposto, può essere così schematizzato:



Nel modello proposto i dati di contesto e le informazioni provenienti dalla *collective intelligence* e dal web sono raccolti, omogeneizzati ed organizzati utilizzando un *Engine* attraverso il quale essi vengono caratterizzati anche dal punto di vista semantico. L'output del *Engine* alimenta un modulo *Profiler* che ne caratterizza i contenuti in funzione del profilo e dei comportamenti dell'utilizzatore che ha richiesto il servizio. I suggerimenti e le raccomandazioni sono quindi inviate in maniera asincrona all'utilizzatore stesso che ne può usufruire attraverso una interfaccia multimodale ed eventualmente attraverso diversi dispositivi. Una caratteristica importante del servizio proposto è che le raccomandazioni saranno inviate nella maniera meno invasiva possibile, cercando di non modificare le abitudini e gli usi tipici del dispositivo da parte dell'utente.

Il dominio di applicazione del modello appena descritto rispetto alla ricerca proposta, riguarda il turismo culturale, limitatamente al contesto della città storica di Venezia.

In tale contesto i numeri dei potenziali utilizzatori del servizio sono, in prima battuta, quelli legati ai flussi turistici medi (si veda il rapporto "Turismo sostenibile a Venezia", redatto nel 2011 dall'Ufficio di Piano del Magistrato alle Acque, sui flussi turistici a Venezia: nel 2008 12 ML di escursionisti e 7 ML di pernottamenti nella Venezia insulare) e più in particolare a coloro che si appoggiano ai servizi recentemente attivati tramite la piattaforma di connettività e servizi wireless *Venice Connected* per la prenotazione e la vendita di servizi pubblici della città (oltre 190.000 ticket commercializzati nel 2011 a turisti provenienti da 188 paesi). Il sistema dovrà essere quindi opportunamente dimensionato per fare fronte al potenziale traffico generato dai numeri appena citati. In ogni caso esso sarà studiato affinché presenti caratteristiche di scalabilità tali da non comprometterne la funzionalità nel caso il numero degli utenti dovesse aumentare in maniera sostanziale rispetto a quanto inizialmente ipotizzato.

Le principali problematiche di Ricerca e Sviluppo che dovranno essere affrontate, possono essere elencate in maniera sommaria come segue:

**problematiche di natura tecnico-scientifica**

- precisione della localizzazione all'interno di una struttura urbana complessa, che risulta particolarmente sfidante e scientificamente rilevante nel contesto urbano di Venezia, con strette strade (calli) ed alti edifici, dove i segnali (sia satellitari che wireless) faticano a propagarsi correttamente
- acquisizione ed organizzazione di dati disomogenei strutturati/non strutturati provenienti da fonti diverse (*data fusion*)
- armonizzazione di dati strutturati (p.e. dati provenienti da sistemi diversi (trip advisor & google place)) ma in grado di fornire informazioni sovrapponibili (per esempio informazioni sugli hotel)
- *harvesting*<sup>79</sup> e analisi di dati non strutturati, raccolti dalla collective intelligence (in particolare commenti e opinioni dai Social Network)
- studio di uno strumento per la “cattura” dei comportamenti e delle azioni dell'utente, su base geografica e temporale
- caratterizzazione semantica<sup>80</sup> dei dati di contesto, studio di una ontologia<sup>81</sup> di dominio e impiego di thesauri e vocabolari

### **problematiche di natura tecnologica**

- impiego di architetture di tipo NoSQL per la memorizzazione di grandi moli di dati secondo modelli di scalabilità orizzontale
- impiego di tecniche statistiche di *data summarization*<sup>82</sup> e database

<sup>79</sup> raccolta di dati utilizzando procedure software

<sup>80</sup> processo di attribuzione di significato ad una risorsa, tramite l'associazione della stessa ad un'ontologia

<sup>81</sup> in ambito informatico una ontologia definisce, all'interno di uno specifico tema di interesse (dominio), i concetti utilizzabili e le regole che legano tali concetti. Lo scopo è quello di istruire un computer riguardo al contesto in cui i dati e i concetti vengono utilizzati.

<sup>82</sup> il processo di data summarization riguarda l'elaborazione di dati primitivi o derivati con lo scopo di ottenere ulteriori dati derivati

temporali probabilistici per gestire la enorme quantità di dati ridondanti generata

- studio e simulazione delle performance del sistema, con lo scopo di garantire l'erogazione di un servizio in grado di soddisfare i requisiti richiesti
- applicabilità di interfacce multimodali<sup>83</sup> per una migliore e più completa interazione tra uomo e macchina; tra le interfacce di sicuro interesse si studieranno anche metodi per non-vedenti che possano indirizzarli verso luoghi di interesse mediante comunicazioni sonore/audio

### **Luoghi di svolgimento del progetto**

Il progetto, pur presentando caratteristiche generalmente applicabili, sarà realizzato in prima battuta nella città di Venezia, con lo scopo di verificare i risultati della ricerca e di attivare il servizio a partire dalla scala locale. Questa scelta è motivata dal fatto che Venezia può rappresentare un banco di prova eccellente per l'impianto di un sistema come quello proposto sia per la sua spiccata vocazione turistica, sia per la disponibilità di un patrimonio culturale vastissimo e unico al mondo, sia per la presenza di una infrastruttura tecnologica di prim'ordine (rete in fibra, wi-fi pubblico, servizio *Venice Connected*) che consente di ipotizzare una effettiva buona riuscita del progetto e la fornitura di un servizio ad elevato valore aggiunto per il turista. Questa tesi è avvalorata dal fatto che questa iniziativa viene sostenuta da Venis spa, società per l'informatica e le telecomunicazioni del Comune di Venezia, che gestisce l'infrastruttura tecnologica della città.

---

<sup>83</sup> sono definite multimodali le interfacce che coinvolgono più di un input di comunicazione, per esempio tastiera e voce

## 9.5 IL PROGETTO: OMNITURIST

La versione *beta* del progetto OmniTurist (tuttora in corso) consente di esplorare e navigare le varie tipologie di contenuti (punti di interesse, alloggi, ristoranti, attività commerciali, servizi, itinerari) che vengono caricati nel sistema mediante una interfaccia di *back-office* opportunamente predisposta. Il prototipo dà ampio risalto alla visualizzazione su mappa, che rappresenta la principale area per la proposta dei prodotti turistici che vengono visualizzati nella posizione geografica corrispondente. Le diverse categorie di contenuti visualizzati vengono selezionate con delle voci di menu cliccabili. Ogni contenuto ha una visualizzazione di dettaglio, attivabile mediante clic sul contenuto visualizzato in modalità lista.

I contenuti sono modellati in modo piuttosto sofisticato, consentendo differenti funzioni di raccomandazione e suggerimento personalizzato basate sul contesto. Un contenuto di particolare interesse è quello dell'itinerario, che consente di associare (e di visualizzare poi sull'interfaccia utente) i vari punti di interesse lungo il percorso dell'itinerario selezionato.

L'utente gestisce il suo piano di viaggio (o itinerario personale), dove può aggiungere i punti di interesse, alloggi, ecc. preferiti. L'utente si registra e può accedere al sistema mediante il suo account *facebook*. Questo gli consente di memorizzare il suo piano di viaggio, al quale potrà successivamente accedere da *mobile*, mediante l'applicazione in corso di sviluppo. Va menzionato inoltre l'utilizzo e la customizzazione di un componente avanzato di gestione mappe (*leaflet*), che consente l'integrazione di diversi *layer* mappe, fra cui dei *layer* che possono essere appositamente predisposti dall'utente (in questo prototipo, per esempio, il *layer* con i numeri civici di Peschici). Il prototipo è strutturato nelle due componenti: *client* e *server*, nel seguito brevemente descritte anche dal punto di vista delle tecnologie impiegate.

### Tecnologie e componenti lato server

Il sistema è stato installato su una macchina *server*, accessibile *on line*. Le attività effettuate sono state:

- Modellazione dei contenuti prodotti turistici. Ogni tipo di contenuto (alloggio, ristorante, itinerario, evento, punto di interesse, attività, servizio) è stato rappresentato secondo un modello dati specifico, consentendo di definire opportunamente tutti i vari metadati necessari a rappresentarlo adeguatamente e abilitare gli algoritmi di raccomandazione basati sui contenuti.
- Implementazione delle componenti di raccomandazione e esposizione contenuti. Lato server, sono stati definiti e implementati i componenti per rendere disponibili i contenuti e le relative funzionalità di ricerca e raccomandazione alle applicazioni web e mobile (in corso di realizzazione). Le funzionalità sono disponibili mediante chiamate HTTP/JSON.
- Implementazione delle componenti lato server per la gestione del piano di viaggio ed esposizione via JSON delle relative funzionalità.

In fase iniziale del progetto stati stati inoltre predisposti l'ambiente di sviluppo su *repository* GitHub, come pure procedure specifiche per l'inizializzazione e aggiornamento della base dati delle macchine degli sviluppatori, per velocizzare le fasi di sviluppo e test.

### **Tecnologie e componenti lato WEBAPP**

Il prototipo WEBAPP è stato implementato secondo un modello innovativo, basato su un approccio *Rich Internet Application* (RIA), che è caratterizzato da un utilizzo avanzato di tecniche AJAX, Javascript e HTML 5. Questo consente la realizzazione di un'interfaccia utente estremamente interattiva ed efficace, come quella della versione nuova di Google Map a cui ci si è ispirati.

Le attività effettuate sono state:

1. Analisi tecnologica dei framework javascript da utilizzare lato WEBAPP. Sono state testate diverse opzioni e successivamente scelte quelle descritte al paragrafo *Tecnologie WEBAPP*.

2. Implementazione dello scheletro applicativo lato client, per consentire di inserire ed estendere successivamente le diverse funzionalità del sistema.
3. Implementazione di un primo set di funzionalità per la visualizzazione dei prodotti turistici geolocalizzati, con l'uso avanzato di componenti mappe geografiche.
4. Implementazione delle funzionalità di gestione del piano di viaggio (prima versione prototipale), autenticazione con *facebook* e rappresentazione e visualizzazione di itinerari.

### **Tecnologie e componenti lato mobile**

E' in corso lo sviluppo dell'applicazione *mobile*, che sarà caratterizzata da:

1. Accesso agli stessi contenuti e servizi esposti lato server alla WEBAPP.
2. Funzionamento *offline*, sia a livello di contenuti (prodotti turistici) che di mappe grafiche.
3. Accesso agli stessi dati di profilo, compreso il piano di viaggio, disponibile sul server.