

**UNIVERSITÀ DEGLI STUDI DI MODENA E
REGGIO EMILIA**

Facoltà di Ingegneria – Sede di Reggio Emilia

Corso di Laurea in Ingegneria Gestionale

**INTEGRAZIONE DI STANDARD
DI CLASSIFICAZIONE PER L'E-COMMERCE:
UNA NUOVA METODOLOGIA BASATA
SUL SISTEMA MOMIS**

Relatore
Prof. Ing. Domenico Beneventano

Tesi di laurea di
Viviana Bonini

Correlatore
Ing. Francesco Guerra

Parole chiave:

E-commerce

Ontologia

Integrazione delle Informazioni

Content standard

Metodologia di integrazione

*“La scienza è sempre imperfetta. Ogni volta che
risolve un problema, ne crea almeno dieci nuovi.”*

George Bernard Shaw

Alla mia famiglia

Indice

Sommario.....	1
Introduzione.....	3
Capitolo 1.....	7
Il commercio elettronico.....	7
1.1 Che cos'è il commercio elettronico.....	8
1.2 Due strategie per un'iniziativa di e-commerce.....	8
1.3 Che cosa sono gli eMarketplace.....	10
1.3.1 Il punto di vista delle imprese.....	12
1.3.2 Il futuro degli eMarketplace.....	13
1.4 Commercio sul Web: l'eProcurement.....	17
1.5 Prospettive dell'e-commerce.....	18
Capitolo 2.....	21
Ontologia e Web Semantico.....	21
2.1 Il Web Semantico.....	22
2.2 Che cos'è l'ontologia.....	23
2.3 Ontologia dei prodotti.....	25
2.3.1 Obiettivi dei marketplace per la gestione delle ontologie.....	26
2.4 Interoperabilità tra sistemi informativi.....	26
2.5 Tecnologie del Web semantico.....	27
2.5.1 XML: eXtensible Markup Language.....	28
2.5.2 RDF e RDF SCHEMA.....	28
2.6 Il database lessicale WordNet.....	29
2.6.1 La Matrice Lessicale.....	32
2.6.2 Relazioni lessicale e sintattiche.....	33
2.6.3 Categorie sintattiche.....	37
Capitolo 3.....	41
Sistemi standard di classificazione dei prodotti per l'E-Commerce.....	41
3.1 Perché classificare prodotti e servizi.....	42
3.1.1 Per trovare e acquistare.....	43
3.1.2 Per il marketing e la distribuzione.....	44
3.2 Gli effetti degli standard di classificazione sulla catena logistica.....	47

3.3 Schema di classificazione	48
3.4 Le differenze tra classificazione e identificazione.....	49
3.5 Convenzioni per la denominazione dei prodotti	50
3.6 Ecl@ss	53
3.6.1 L'estensione degli obiettivi del progetto ecl@ss.....	54
3.6.2 Funzioni e applicazioni.....	55
3.6.3 La struttura.....	56
3.6.4 Le possibilità di ricerca con ecl@ss	59
3.6.5 Aggiornamento e mantenimento	60
3.6.6 Prospettive future di ecl@ss	61
3.6.7 Cooperazioni.....	61
3.6.8 Qualità e manutenzione dei codici.....	62
3.7 UNSPSC (UNiversal Standard Products and Services Classification).....	63
3.7.1 Classificazione gerarchica	64
3.7.2 ECCMA: organizzazione distributrice di UNSPSC	66
3.7.3 Distribuzione di UNSPSC	69
3.7.4 Rivisitazione e aggiornamento	70
3.8 NAICS: dati nuovi per un'economia nuova.....	71
3.8.1 Classificazione delle nuove industrie	72
3.8.2 Diversi settori economici coinvolti	73
3.8.3 Concetti rivisti della classificazione	74
3.8.4 I nuovi sistemi di numerazione.....	75
3.8.5 Le prospettive di NAICS	77
Capitolo 4	79
Integrazione delle Informazioni	79
4.1 Tre livelli di eterogeneità dei dati	80
4.2 Cataloghi di prodotti: descrizione della struttura.....	81
4.3 Descrizione del contenuto dei prodotti	82
4.4 Descrizione di prodotto a seguito di riclassificazione	83
4.5 Allineamento dei content standard.....	84
4.5.1 Allineamento di standard orizzontali.....	85
4.5.2 Allineamento tra standard orizzontali e verticali.....	86
4.5.3 Allineamento di standard verticali.....	86
4.6 Integrazione dei documenti	86
4.6.1 Esempio di integrazione di documenti	87
4.6.2 Integrazione a un solo livello	92
4.6.3 Integrazione multilivello (layered approach)	92
4.7 Integrazione dei modelli di dati	94
Capitolo 5	97
MOMIS: un sistema intelligente di integrazione	97
5.1 Integrazione Intelligente di Informazioni	98
5.1.1 Il programma I^3	98

5.1.2 Architettura di riferimento per sistemi I^3	99
5.1.3 Il mediatore	102
5.2 Il sistema MOMIS	103
5.2.1 Il linguaggio ODL_{I3}	105
5.2.2 Gli strumenti di MOMIS	105
5.3 Wrapping degli schemi sorgente	107
5.3.1 I dati semistrutturati	107
5.3.2 Sorgenti XML	108
5.3.3 Wrapping di sorgenti XML	109
5.4 Il processo di integrazione	109
5.5 Generazione del Thesaurus comune	110
5.5.1 Estrazione delle relazioni intra-schema	111
5.5.2 Estrazione delle relazioni inter-schema	111
5.5.3 Arricchimento dell'insieme delle relazioni	111
5.5.4 Validazione delle relazioni	112
5.5.5 Inferenza di nuove relazioni	112
5.6 Il calcolo delle affinità	113
5.7 La generazione dei cluster	113
5.8 La costruzione delle classi globali	114
5.9 Fusione degli attributi	114
5.9.1 Fusione degli attributi di relazioni validate	115
5.9.2 Fusione di attributi di relazioni non validate	115
5.10 La generazione delle mapping table	116
5.11 SI-Designer: il tool per la creazione dello schema globale	117
5.11.1 L'automa di utilizzo di SI-Designer	118
5.11.2 Source	119
5.11.3 SIM A	120
5.11.4 SLIM	121
5.11.5 SIM B	122
5.11.6 Cluster	123
5.11.7 TUNIM	125
Capitolo 6	127
Integrazione di standard di classificazione con il sistema MOMIS	127
6.1 Il Challenge di OntoWeb	128
6.1.1 Ipotesi metodologica	128
6.1.2 Mapping tra schemi	130
6.1.3 Come arrivare a un'ontologia comune: <i>case study</i> sul dominio "Writing Paper"	132
6.2 Un framework di integrazione dati per la classificazione prodotti	135
6.2.1 L'esempio "Writing Paper" con MOMIS	136
6.3 Estensione del processo di integrazione	141
6.3.1 Studio delle codifiche	141

6.3.2 Due approcci per la navigazione tra le codifiche	144
6.3.3 Dalla selezione dei frammenti al file XML	148
6.3.4 Inizializzazione dei Wrapper XML	151
6.3.5 Fasi dell'integrazione su SI-Designer.....	152
6.3.6 Mapping risultato tra i frammenti di codice sorgente	160
Capitolo 7	163
Individuazione di fattori critici e proposte metodologiche	163
7.1 Fattori critici nell'integrazione con MOMIS	164
7.2 Fattori riconducibili al sistema.....	165
7.2.1 Contributo del progettista	165
7.2.2 WordNet	169
7.3 Fattori riconducibili agli standard	171
7.3.1 Frammenti di codice non corrisposti	172
7.3.2 Gestione della gerarchia nei codici.....	177
7.4 Opportunità di sviluppo nell'integrazione tra codici con MOMIS	182
7.5 Metodologia di integrazione per l'interazione del progettista con MOMIS.....	183
7.6 Aggiornamento di sorgenti	186
7.7 WordNet: orizzonti di miglioramento.....	189
7.8 Raccomandazioni per la costruzione di un nuovo standard globale	193
Conclusioni.....	197
Appendice A.....	199
Software di supporto alla fase di testing	199
Appendice B.....	201
Test di riferimento in XML.....	201
Glossario.....	217
Bibliografia.....	221
Siti Internet di riferimento.....	227

Indice delle figure

Figura 1 Modelli B2B, Fonte Associazione Impresa Politecnico, 2002.....	9
Figura 2 eMarketplace B2B per categorie, 2000 & 2003. Fonte: Deloitte Consulting ..	12
Figura 3 Numero di eMarketplace orizzontali e verticali a livello mondiale (2001)	14
Figura 4 Stima del numero degli eMarketplace che sopravviveranno nel 2003.....	15
Figura 5 Ricavi da E-Commerce B2B in Europa, 2000-2004 (in miliardi di dollari)	15
Figura 6 Percentuale di B2B sull'E-Commerce totale	16
Figura 7 Overview di WordNet	30
Figura 8 Relazione di iponimia per la definizione <i>machine</i>	31
Figura 9 Lista dei beginner	31
Figura 10 Relazione tra lemmi e significati.....	32
Figura 11 Matrice Lessicale.....	33
Figura 12 Estrazione di informazioni da un catalogo prodotti	81
Figura 13 Relazione tra standard verticali e orizzontali	85
Figura 14 Integrazione con n*m mapping	90
Figura 15 Integrazione con n+m mapping.....	91
Figura 16 Mapping relativo all'esempio <i>Address</i>	91
Figura 17 Approccio di integrazione multi-livello	93
Figura 18 Il modello per la trasformazione dei dati.....	94
Figura 19 Diagramma dei servizi I^3	99
Figura 20 Architettura del sistema MOMIS	104
Figura 21 Struttura di ODB tools.....	106
Figura 22 Prima fase del processo di integrazione	110
Figura 23 Esempio di Mapping Table	117
Figura 24 Architettura di SI-Designer	117
Figura 25 Pannello Source	119
Figura 26 Pannello Thes.Rel: run SIM A	120
Figura 27 Modulo SLIM.....	121
Figura 28 Modulo SIM B.....	123
Figura 29 Pannello Cluster	124
Figura 30 Modulo TUNIM	125
Figura 31 Architettura di un generico mapping.....	131

Figura 32 Esempio di mapping di OntoWeb (il caso “writing paper”).....	134
Figura 33 Frammenti di codice UNSPSC e ecl@ss integrati con MOMIS	139
Figura 34 Risultato del mapping sul dominio “Writing Paper”	139
Figura 35 Approccio Bottom-Up	145
Figura 36 Approccio Top-Down	145
Figura 37 Frammenti di codice su cui applicare l’approccio Bottom-Up.....	146
Figura 38 Segmento di NAICS sul dominio Farming.....	147
Figura 39 Segmento di UNSPSC sul dominio Farming.....	148
Figura 40 Approccio Bottom-Up sul dominio RadioTV	149
Figura 41 Acquisizione delle sorgenti nell’esempio RadioTV	153
Figura 42 SIM A nell’esempio RadioTV	154
Figura 43 Salvataggio dello stato XML	154
Figura 44 Modulo SLIM: apertura menù a scomparsa	155
Figura 45 Assegnazione Forma Base	156
Figura 46 Selezione del significato su WordNet.....	156
Figura 47 Costruzione delle relazioni inter-schema.....	156
Figura 48 Creazione manuale delle relazioni	157
Figura 49 Esempio di visualizzazione del Global Schema	158
Figura 50 Mapping finale dell’esempio RadioTV	161
Figura 51 Fattori critici del processo di integrazione.....	164
Figura 52 Frammenti dei codici UNSPSC e ecl@ss sul dominio RadioTV	167
Figura 53 Mapping tra commodity “specializzate” – Esempio RadioTV	167
Figura 54 Frammenti di codice di UNSPSC e ecl@ss relativi al dominio Bottle.....	168
Figura 55 Mapping finale tra UNSPSC e ecl@ss sul dominio Bottle.....	168
Figura 56 Frammento di codice UNSPSC sul dominio Cooker.....	170
Figura 57 Frammento di codice ecl@ss sul dominio Cooker	170
Figura 58 Test di tipo bottom-up sul dominio Book.....	173
Figura 59 Schema del dominio Household in UNSPSC fino al terzo livello gerarchico	174
Figura 60 Schema del dominio Household in ecl@ss fino al terzo livello gerarchico .	175
Figura 61 Ricerca on-line su ecl@ss dei codici prodotto.....	178
Figura 62 Ricerca on-line su UNSPSC dei codici prodotto	178
Figura 63 Frammenti di codice relativi al dominio Hardware	180
Figura 64 Frammenti di codice relativi al dominio Software	181
Figura 65 Mapping tra classi di prodotto: esempio RadioTV	185
Figura 66 Introduzione di relazioni di sinonimia nel pannello <i>Thes.Rel.</i>	185
Figura 67 Risultato della nuova metodologia di mapping sul dominio RadioTV	186
Figura 68 Aggiornamento delle sorgenti nell’esempio Household.....	188

Indice delle tabelle

Tabella 1 Classificare prodotti o servizi per trovare o acquistare.....	43
Tabella 2 Classificare prodotti e servizi per supportare le vendite e il marketing.....	45
Tabella 3 Codici comuni di identificazione.....	49
Tabella 4 Differenze tra codici di identificazione e di classificazione.....	50
Tabella 5 Gerarchia di ecl@ss.....	56
Tabella 6 Proprietà delle Classi di materiali in ecl@ss.....	57
Tabella 7 Proprietà delle Parole chiave in ecl@ss.....	58
Tabella 8 Descrizione livelli gerarchici in UNSPSC.....	65
Tabella 9 Conversione EGCC - EGCI.....	65
Tabella 10 Le nuove aziende di NAICS.....	72
Tabella 11 Settori interessati dalla classificazione NAICS.....	74
Tabella 12 Struttura gerarchica di NAICS e codici numerici per un settore.....	76
Tabella 13 NAICS (1997) VS SIC (1987).....	78
Tabella 14 Segmenti di ecl@ss.....	142
Tabella 15 Segmenti di NAICS.....	142
Tabella 16 Segmenti di UNSPSC.....	144

Sommario

Il commercio elettronico ha rivoluzionato il modo di scambiare informazioni tra gli attori economici. Un requisito fondamentale necessario per le transazioni on-line è la codifica dei prodotti nei documenti commerciali in formato elettronico (come cataloghi, siti Web, ordini d'acquisto, fatture) che permette alle applicazioni informatiche, attraverso la catena logistica, di processare i dati delle transazioni in modo automatico, migliorando la gestione e l'analisi per il supporto alle decisioni in tempi critici e con metodi efficienti. Molte organizzazioni hanno sfruttato le opportunità offerte dalle soluzioni Internet e altre si apprestano a farlo, condividendo i loro database e cataloghi con numerosi altri utenti. Di conseguenza non solo è cresciuto il numero di informazioni disponibili, ma è aumentata la "confusione" generata dalla sovrabbondanza di sorgenti di dati.

In questo contesto si è sviluppato il ruolo di intermediari degli eMarketplace, mercati virtuali che mettono in collegamento la domanda di beni o servizi di molti compratori e l'offerta di differenti venditori che partecipano a un processo di business in rete. Ma il confronto tra offerte commerciali sul Web è limitato dalla mancanza di standard per la codifica dei prodotti. L'esigenza quindi degli eMarketplace è la ri-classificazione di beni e servizi in accordo ai diversi modelli di standardizzazione.

Questa tesi ha l'obiettivo di affrontare il problema della ri-classificazione di beni e servizi nell'e-commerce proponendo una metodologia semi-automatica, supportata dal sistema a mediatore MOMIS, per l'integrazione dei diversi standard di classificazione dei prodotti.

Introduzione

Lo sviluppo di Internet ha drasticamente aumentato la disponibilità e l'accessibilità delle informazioni, ma l'ammontare dei dati ha reso sempre più difficile trovare, accedere, presentare e mantenere le informazioni ricercate da una grande varietà di utilizzatori. L'utente si trova a prendere in considerazione tutti i documenti trovati per estrarre le informazioni desiderate.

Molte organizzazioni, nel contesto del commercio elettronico, hanno sfruttato le opportunità offerte dalle soluzioni Internet e altre si apprestano a farlo, condividendo i loro database e cataloghi con altri utenti sul Web. Di conseguenza non solo è cresciuto il numero di informazioni disponibili, ma è aumentata la "confusione" generata dalla sovrabbondanza di sorgenti di dati.

L'esplosione delle comunicazioni in rete ha favorito un fenomeno che sarebbe stato difficile immaginare: gli aspetti ontologici dell'informazione hanno acquisito un valore strategico. Questi aspetti sono indipendenti dalle forme di codifica dell'informazione stessa, che può essere quindi isolata, recuperata, organizzata, integrata in base a ciò che più conta: il suo contenuto. La standardizzazione di tali contenuti risulta oggi cruciale nella prospettiva delle aziende integrate e del commercio elettronico ed è indispensabile per semplificare i processi di comunicazione.

In questo ambito si sta affermando con forza il ruolo di intermediari assunto dagli eMarketplace, mercati virtuali dove la domanda di un compratore può teoricamente scegliere tra i prodotti distribuiti e offerti da numerosi venditori; ma il futuro di queste arene imparziali è condizionato dalla mancanza di standard che descrivano e classifichino beni e servizi in modo completo. Perciò gli eMarketplace dovrebbero

mettere a disposizione un ambiente che sappia mediare tra i differenti standard utilizzati da chi partecipa alla transazione, in modo tale che ogni attore possa utilizzare il proprio formato. In definitiva, il problema dell'eterogeneità degli standard di descrizione per prodotti, cataloghi e documenti si pone come ostacolo allo sviluppo dei mercati elettronici e può essere risolto studiando opportuni sistemi per l'integrazione delle sorgenti di dati; in particolare per il successo e la crescita delle iniziative di commercio elettronico diventa strategica la definizione di un modello universale di standard per la classificazione dei prodotti (content standard).

Questa tesi propone una metodologia semiautomatica di integrazione dei principali standard di classificazione esistenti, attraverso un sistema a mediatore, denominato MOMIS (Mediator EnvirOnment for Multiple Information Sources), progettato per integrare sorgenti di dati eterogenee e distribuite nel contesto dell'Integrazione Intelligente delle Informazioni (*I³*).

In base alle analisi effettuate, l'importanza di una metodologia per l'integrazione è risultata evidente soprattutto considerando che il sistema MOMIS è un sistema di integrazione semiautomatico, ovvero effettua parte delle integrazioni in modo automatico (in base agli schemi delle sorgenti) e in parte richiede l'interazione con il progettista per effettuare delle inevitabili scelte durante l'integrazione. Un'altra ragione che rafforza lo studio e la formulazione di una metodologia è quella di considerare il contesto specifico dell'integrazione di standard di classificazione e quindi di adattare il sistema MOMIS a problemi di integrazione riconducibili a questi standard.

Il Capitolo 1 presenta l'impatto che il commercio elettronico, in particolare di tipo B2B, ha avuto fino ad oggi sulle attività di business; definisce inoltre il concetto di eMarketplace, supporto strategico agli scambi commerciali on-line.

Nel Capitolo 2 si studia il problema della standardizzazione del contenuto delle informazioni, riconducibile alle ontologie e si descrivono alcune tecnologie per il Web (in particolare il linguaggio XML) e un esempio di ontologia lessicale, WordNet, utile nel seguito della trattazione.

Il Capitolo 3 propone una serie di definizioni e specifiche sul vasto panorama degli standard di classificazione, per illustrare requisiti e benefici; quindi si sofferma sull'analisi più approfondita di alcuni standard: UNSPSC, ecl@ss, NAICS.

Successivamente, nel Capitolo 4, si riporta lo stato dell'arte della ricerca sull'Integrazione delle Informazioni e nel Capitolo 5 viene presentato il mediatore

MOMIS, utilizzato nella fase sperimentale della tesi per l'integrazione semi-automatica di standard di classificazione.

Nell'ultima parte (Capitoli 6-7), sulla scorta delle conoscenze raccolte in letteratura e dell'esperienza acquisita nella fase di *testing*, si formula una metodologia generale per l'integrazione di sorgenti di dati in forma semi-strutturata, quali gli standard di classificazione dei prodotti, con il sistema MOMIS.

Capitolo 1

Il commercio elettronico

Lo sviluppo del commercio elettronico negli anni '90 ha radicalmente trasformato gli scenari economici internazionali e la frenata della New Economy, avvenuta nell'ultimo anno, non può far sottovalutare i grandi cambiamenti che le Information and Communication Technology (ICT) hanno portato e continueranno a portare, incrementando le abilità delle imprese a creare valore e a competere nei mercati globalizzati.

La macchina a vapore, le ferrovie, il motore a combustione interna e le applicazioni industriali dell'elettricità generarono nuove industrie e servizi e, più importante, permisero alle imprese di lavorare in modo diverso e più efficiente. Tali innovazioni ebbero come risultato un aumento di produttività, che tuttavia impiegò decenni ad affermarsi dai settori interessati dal cambiamento tecnologico al resto dell'economia. In questa ottica i computer e Internet rappresentano un elemento di discontinuità per la rapidità della loro diffusione.

Le Information Technology (IT) sono diventate in breve tempo certezze comuni della vita economica e sociale moderna, aprendo opportunità e nuove strade per tutti i soggetti demografici in molte aree geografiche del mondo; la rivoluzione digitale infatti è un fenomeno globale che rappresenta per i Paesi in via di sviluppo una possibilità di crescita senza precedenti, tale da farli competere con quei Paesi che non sapranno impiegare le potenzialità tecnologiche [UN Report, 01].

1.1 Che cos'è il commercio elettronico

Il Patricia Seybold Group¹, grazie alle conoscenze acquisite sul campo in anni di lavoro della propria organizzazione, dà una definizione di e-commerce basata sull'esperienza.

Commercio elettronico, e-commerce (conosciuto anche come e-business), significa svolgere un'attività economica via computer - il che interessa tutti gli aspetti di questa attività, comprendendo l'intero processo di gestione del ciclo vitale del prodotto e del cliente - dalla pubblicità al marketing, passando per le vendite, gli ordini, la fabbricazione, la distribuzione, l'assistenza ai clienti prima e dopo l'acquisto, la ricostituzione delle scorte [Seybold, 98].

Nel prendere parte a iniziative di e-commerce gli attori economici utilizzano tutte le tecnologie disponibili per conferire la massima efficienza alle relazioni di affari. Tali tecnologie includono Internet, ma anche sistemi telefonici avanzati, apparecchi digitali delle dimensioni di una mano, televisione interattiva, smart card e un'infinità di tecnologie emergenti. Tutte queste tecnologie a diretto contatto con la clientela sono sorrette, dietro le quinte, da database integrati della stessa, call center, flussi operativi razionalizzati e sistemi per le operazioni commerciali che ne garantiscono la riservatezza; richiedono sistemi per comunicare tra loro senza problemi, in maniera affidabile e sicura al di là dei confini aziendali, di quelli geografici, dei fusi orari.

1.2 Due strategie per un'iniziativa di e-commerce

L'E-Commerce, che può comprendere tutto il ciclo di vita del cliente e del prodotto, si applica con due strategie:

1. B2B: Business to Business

Il B2B, inteso in senso lato come universo delle applicazioni tra imprese (o comunque organizzazioni) rese possibili dalle tecnologie dell'informazione e della

¹ Tra le principali società di consulenza americane per soluzioni e-business di successo

comunicazione basate su Internet, è un mondo estremamente complesso, ben più complesso di come in un primo momento anche gli addetti ai lavori hanno cercato di modellizzare: più passa il tempo e la sua conoscenza aumenta, più ci si rende conto che alcune schematizzazioni “del passato” risultano eccessivamente semplicistiche e non consentono di cogliere efficacemente la complessità della realtà.

A questo fine si possono utilizzare due differenti assi di lettura del B2B [Bertelè, 02]: il primo si riferisce al soggetto cui fa capo l'applicazione (se un'impresa specifica, un intermediario, oppure un altro operatore); il secondo alle funzionalità offerte dall'applicazione (se transazionali di commercio elettronico oppure di natura diversa). Sulla base di tali assi, possiamo distinguere tra (Figura 1):

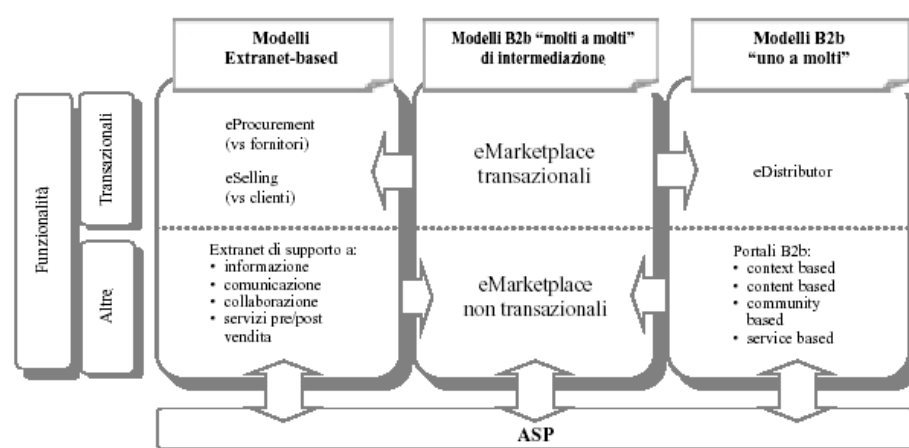


Figura 1 Modelli B2B, Fonte Associazione Impresa Politecnico, 2002

- modelli Extranet-based, intesi come quelle applicazioni Internet realizzate dalla singola impresa industriale o di servizi per interagire con i propri partner commerciali. Tali applicazioni possono essere finalizzate specificatamente a portare on-line i processi di acquisto o di vendita (in questo caso, si parla rispettivamente di eProcurement o di eSelling), oppure ad offrire ai propri partner di filiera funzionalità diverse (ad esempio, possibilità di accedere ad informazioni commerciali e tecniche sempre aggiornate, verificare le giacenze a magazzino, interagire in modalità sincrona o asincrona con il personale dell'impresa, ecc.);
- modelli di intermediazione “molti a molti”, o eMarketplace, intesi in senso lato come intermediari B2B finalizzati a supportare qualsiasi processo connesso alla relazione cliente/fornitore, anche se di natura non strettamente transazionale (ad

esempio se si limita a far incontrare domanda e offerta senza però consentire la chiusura della transazione on-line).

- altri modelli B2B “uno a molti”, che includono, a livello di funzionalità transazionali, gli eDistributor che rappresentano l’equivalente on-line dei distributori tradizionali e, negli altri casi, siti che possono mettere a disposizione delle imprese funzionalità di varia natura: contenuti informativi, strumenti per ricercare le risorse in rete utili (ad esempio directory B2B), strumenti interattivi (forum, chat, mailing list, ecc.) finalizzati alla creazione di comunità virtuali di business, ecc. (denominate genericamente col termine di Portali B2B).

La Figura 1 evidenzia nella parte inferiore un’altra categoria di operatori che stanno iniziando a svolgere un certo ruolo nello scenario del B2B, quella degli Application Service Provider (ASP), che è trasversale rispetto alle altre tre tipologie di attori. Gli ASP sono fornitori di applicazioni software che vengono rese accessibili agli utenti in remoto via Internet e pagate a consumo, con una formula analoga al *pay per use*.

2. B2C: Business to Consumer

Il Business to Consumer (B2C) è quella forma di commercio in cui beni e servizi vengono venduti all’utente finale che, rispetto al commercio tradizionale, si aspetta un’offerta più ampia e comparabile e una semplificazione del procedimento di acquisto per trarre vantaggio in termini di tempo, di diffusione e quindi di risparmio rispetto alla forma classica.

A fronte di queste richieste la dinamica del commercio elettronico B2C prevede un maggiore lavoro a monte del processo di vendita per implementare e per sfruttare le tecnologie proprie di questa attività.

1.3 Che cosa sono gli eMarketplace

Internet può essere considerato un valido strumento per reperire e diffondere informazioni commerciali. Molte imprese hanno avviato la loro attività di e-business utilizzando siti Web per la diffusione di informazioni fondamentali riguardanti la loro azienda e i loro prodotti. Trasformare quei siti Web da catalogo on-line in una soluzione e-Commerce integrata è un obiettivo che richiede notevoli risorse. Oltretutto i

potenziali clienti possono avere grandi difficoltà nel trovare un sito Web tra milioni di altri siti utilizzando esclusivamente i tradizionali canali di ricerca.

Per questo motivo gli eMarketplace business-to-business (B2B eMarket) stanno diventando importanti luoghi di incontro e di scambio per molte aziende.

Un eMarketplace rappresenta un luogo di incontro tra molti compratori e molti venditori. Costituisce un'arena su Internet dove un intermediario affidabile offre nuove opportunità commerciali e funzioni aziendali innovative alle aziende registrate. L'intermediario non interferisce sulla definizione dei prezzi dei beni e dei servizi offerti sull'eMarketplace [Prevete, 01].

I B2B eMarketplace possono essere definiti, in modo più esplicito, come una soluzione basata su Internet che mette in collegamento la domanda di beni o servizi di un compratore e l'offerta di un venditore che partecipano a un processo di business.

Perché un'impresa dovrebbe utilizzare gli eMarketplace?

- Per raggiungere nuovi clienti e partner;
- Per offrire ai clienti un mezzo conveniente per acquistare prodotti;
- Per ridurre i costi utilizzando i servizi offerti dall'eMarketplace;
- Per reperire informazioni industriali riguardanti la concorrenza.

Esistono tre tipi fondamentali di eMarketplace:

- indipendenti;
- sales focused - orientati alle vendite;
- procurement focused - orientati agli acquisti.

L'eMarketplace indipendente è aperto a qualsiasi compratore o venditore affidabile appartenente ad una specifica industria o regione. Il provider dell'eMarketplace non è tenuto a selezionare i prodotti o le aziende, ma dovrebbe permettere al maggior numero di operatori dell'industria di ogni settore di parteciparvi.

L'eMarketplace è spesso gestito da terzi e, al fine di mantenersi indipendente, deve garantire la partecipazione di qualsiasi compratore o venditore affidabile. In realtà, molto spesso l'indipendenza non è assicurata (soprattutto nel caso in cui gli eMarketplace siano composti da soli venditori o da soli compratori, oppure quando tra gli azionisti siano presenti aziende del settore).

Gli altri due tipi di eMarketplace sono meno comuni, ma il loro numero sta

aumentando. Gli eMarket orientati alle vendite possono essere paragonati a grossisti o dettaglianti che vendono prodotti di diverse aziende. Ad essi però non possono aderire indistintamente tutti i venditori. Gli eMarket orientati agli acquisti costituiscono comunità commerciali aventi lo scopo di creare efficienti canali di acquisto.

Nel Business-to-Business la maggior parte degli eMarketplace si concentrano su settori industriali specifici. Altri si distinguono per le funzioni che offrono o per la regione o area geografica che rappresentano.

Nella maggior parte degli eMarketplace, nei quali i compratori possono effettuare il loro ordine, i contatti tra cliente e fornitore avvengono direttamente via e-mail. Le condizioni di pagamento e di consegna sono spesso definite direttamente tra il compratore e il venditore. Il rapido sviluppo dell'e-commerce attraverso gli eMarketplace sta incentivando le aziende ad attivare sistemi di integrazione ERP (come IFS PeopleSoft, SAP, Scala...). In questo modo gli ordini possono essere immessi direttamente nel sistema aziendale subito dopo la consultazione del catalogo prodotti messo a disposizione dall'eMarketplace.

1.3.1 Il punto di vista delle imprese

In accordo con gli obiettivi delle imprese, è possibile distinguere tra due tipi di eMarketplace:

- **eMarketplace Verticali**, si indirizzano ad un settore ben individuato e vi si scambiano beni e servizi relativi all'industria o alla categoria merceologica a cui fanno riferimento.
- **eMarketplace Orizzontali**, si rivolgono verso più settori e vi si scambiano beni e servizi appartenenti a diverse categorie merceologiche.

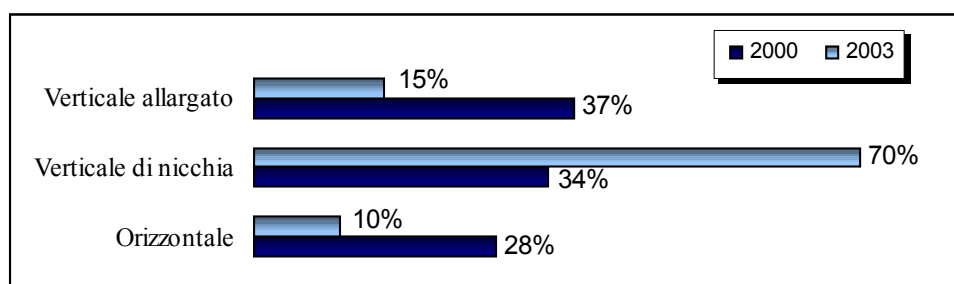


Figura 2 eMarketplace B2B per categorie, 2000 & 2003. Fonte: Deloitte Consulting

Gli eMarketplace verticali sono tipici dei settori tradizionali, come l'automotive e l'approvvigionamento di materie prime. I processi tradizionali hanno molti punti di inefficienza che contribuiscono all'aumento dei costi e provocano ritardi nelle transazioni economiche. Per questo i mercati verticali rappresentano un'opportunità per creare valore aggiunto. Fra le industrie più attrattive per soluzioni di eMarketplace vi sono quelle informatiche, elettroniche, chimiche, di veicoli a motori, di forniture per ufficio e altre. I prodotti scambiati nei Marketplace verticali sono complessi e hanno brevi cicli di vita.

I mercati orizzontali, in opposizione a quelli verticali, non sono orientati ai bisogni di una particolare impresa. Essi distribuiscono materiali indirettamente per varie industrie verticali e offrono loro diversi servizi. Nati per commercializzare materie prime, gli eMarketplace orizzontali sono oggi sempre più presenti nel settore dei servizi.

1.3.2 Il futuro degli eMarketplace

Alla fine del 2000 Deloitte Consulting ha stimato che il numero degli eMarketplace B2B a livello mondiale aveva raggiunto quota 1500, 1300 dei quali realizzati negli Stati Uniti [Deloitte, 01]. La stessa società prevede tuttavia che circa 1000 eMarketplace assumeranno connotazioni diverse rispetto alle attuali, mentre i restanti 400-500 si evolveranno in eMarketplace di nicchia che forniranno servizi on-line e/o piattaforme tecnologiche.

A metà Ottobre 2001 eMarketServices² ha invece contato, recensito e classificato nella propria directory 1125 eMarketplace B2B tra orizzontali e verticali.

Al contrario, Keenan Vision in [Keenan, 01] prevede che nel 2004 ci saranno 4.000 eMarketplace soltanto negli Stati Uniti, su un totale mondiale di 8.000. Questa previsione fa riferimento a 230 settori industriali individuati negli Stati Uniti. Keenan Vision sostiene che si svilupperanno tre tipi di eMarketplace, secondo i flussi commerciali che si hanno a livello globale, regionale/nazionale, locale. Si prevede che i mercati locali si inseriscano in snodi commerciali più ampi.

² eMarketServices è un osservatorio sull'e-Business Internazionale nato dalla collaborazione tra l'ICE (Istituto nazionale per il Commercio Estero) e le principali Agenzie europee per la promozione del commercio estero con lo scopo di favorire l'e-Business presso le PMI dei paesi membri a supporto dei processi di internazionalizzazione.

Secondo una previsione decisamente più restrittiva, Forrester Research ritiene invece che entro il 2003 gli eMarketplace saranno soltanto 181, e di questi 128 saranno verticali [Lief, 99].

Tra i principali istituti di ricerca si sta affermando la convinzione che gli eMarketplace verticali diverranno il modello di maggior successo. Oltre agli eMarketplaces Verticali, Forrester prevede che ci saranno soltanto quattro eMarketplaces Orizzontali che venderanno beni indiretti e MRO (Maintenance, Repair and Office, ovvero beni indiretti e per ufficio).

eMarketplace Verticali			
Agricoltura	72	Macchinari e veicoli	19
Alimentare	86	Metallurgia e mining	43
Ambiente	16	Nautica	19
Automobilistico	32	Pasta di legno e carta	17
Aviazione	25	Prodotti e servizi IT	53
Chimica e plastica	47	Pubblicità e media	22
Edilizia	63	Telecomunicazioni	39
Elettronica e componenti	68	Tessile e pelletteria	30
Finanza	44	Trasporti e logistica	108
Packaging	11	Sanità e farmaceutico	52
Legname	24	Altro	83
eMarketplace Orizzontali			
Ospitalità	16		
Servizi	58		
Ufficio attrezzature	24		
Altro	54		
			<i>Totale 1125</i>

Figura 3 Numero di eMarketplace orizzontali e verticali a livello mondiale (2001)

Fonte eMarketservices, 2001

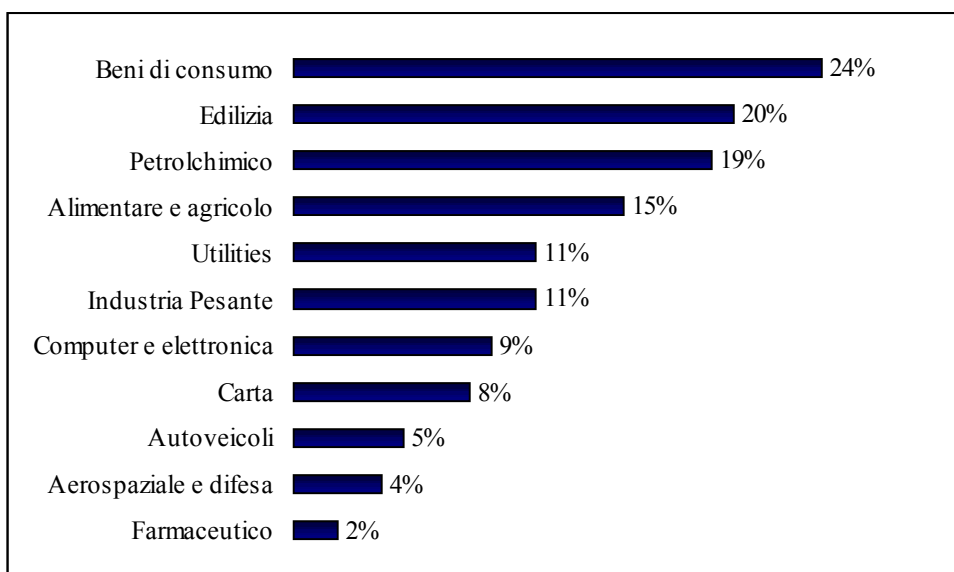


Figura 4 Stima del numero degli eMarketplace che sopravviveranno nel 2003

1.3.2.1 Il quadro europeo

Secondo stime dell'Istituto Nazionale per il Commercio Estero è sensato prevedere una forte crescita del settore B2B in Europa nei prossimi anni: si passerà dai 26 miliardi di USD del 2000 ai 797 miliardi di USD del 2004. Tuttavia, includendo un calcolo ponderato che tiene conto della maturazione degli eMarketplaces verticali, i ricavi stimati nel 2000 per i prossimi anni vanno rivisti al ribasso come risulta dall'“eEurope Report 2001”.

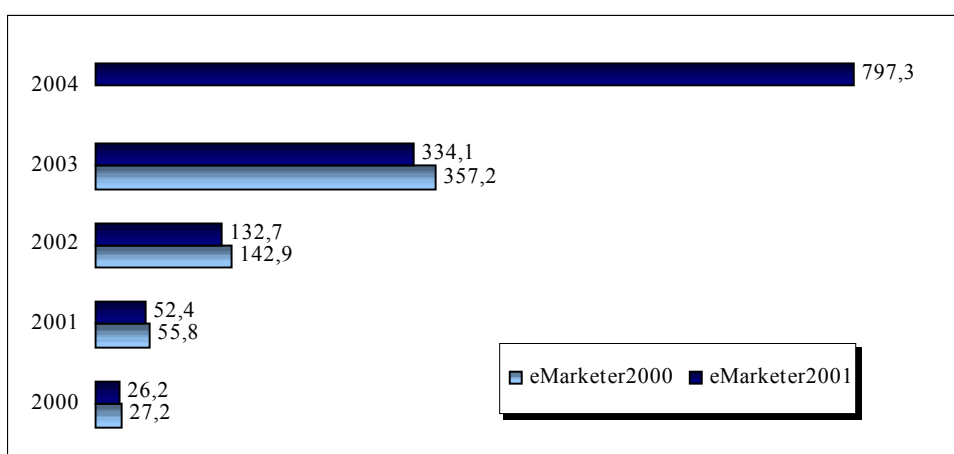


Figura 5 Ricavi da E-Commerce B2B in Europa, 2000-2004 (in miliardi di dollari)

La quota maggiore dei ricavi provenienti dell'e-commerce a livello globale proviene da transazioni B2B. Nel tempo, l'incidenza del B2B sul totale delle attività di e-commerce aumenterà ulteriormente in Europa, passando dal 77% del 2000 all'81% nel 2004.

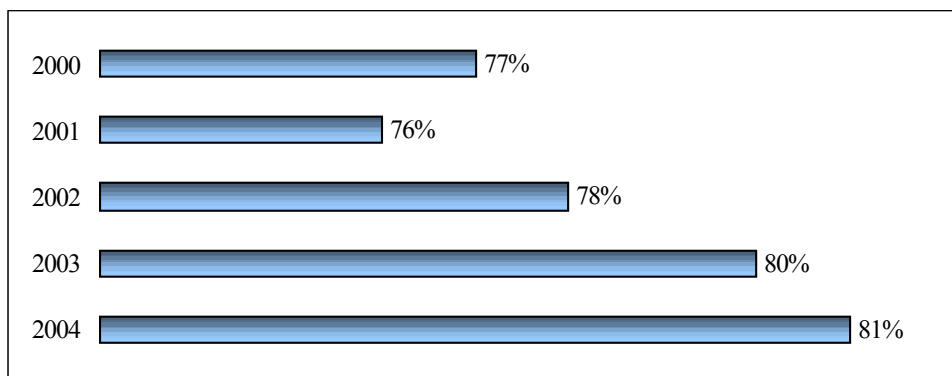


Figura 6 Percentuale di B2B sull'E-Commerce totale

In Italia negli ultimi anni il comparto degli eMarketplace è stato caratterizzato da uno slancio imprenditoriale notevole, risultato di un clima che, dopo le prime delusioni delle start-up Internet del B2C, riponeva maggiori speranze nel B2B. Lo testimonia il numero elevato di Marketplace nati negli ultimi anni: l'Associazione Impresa Politecnico in [Bertelè, 02] ne ha censiti ben 124. Ma anche in questo comparto è già iniziato un profondo processo di riassetto: alcune iniziative, a causa delle difficoltà incontrate ben superiori a quelle previste, hanno abbandonato il modello dell'eMarketplace per adottare modelli di business differenti (ad esempio quello del portale B2B o del fornitore di servizi alle imprese); altre, non poche, hanno già volontariamente o forzatamente chiuso i battenti.

La maggior parte dei Marketplace italiani sono "indipendenti", cioè guidati da gruppi imprenditoriali che non hanno diretti interessi nella filiera. Ed è proprio tra questi che si è già verificata la mortalità maggiore, in particolare tra quelli che si erano posti obiettivi finanziari di breve termine eccessivamente ambiziosi o non avevano particolari asset su cui fondare il proprio modello di business.

In questa fase embrionale e pionieristica, i modelli di business sono molteplici ed eterogenei: si passa da eMarketplace informativi pubblici, che si pongono come unico obiettivo quello di far incontrare on-line più efficientemente domanda e offerta, a eMarketplace che mirano specificatamente a consentire a clienti/fornitori di gestire, in modalità anche personalizzabili, processi collaborativi lungo la supply chain.

1.4 Commercio sul Web: l'eProcurement

Col termine eProcurement si intende l'acquisto e la vendita di beni e servizi tramite Internet. Tipicamente i siti Web predisposti per l'eProcurement permettono ad utenti qualificati e registrati di cercare acquirenti o venditori di beni e servizi. A seconda dell'approccio, l'acquirente o il venditore possono indicare il prezzo o richiedere delle offerte. Le transazioni sono poste in essere e concluse. Il miglior offerente può qualificarsi per sconti sul volume, offerte speciali o una migliore tempestività nella consegna, a seconda dell'importanza data a ciascun parametro.

Un software per l'eProcurement può rendere possibile l'automazione di alcuni acquisti e vendite. Le aziende che partecipano si aspettano di acquisire un maggiore controllo del magazzino, ottimizzare il numero di agenti ed incrementare i cicli produttivi. Il processo dell'eProcurement va integrato attraverso la gestione della catena distributiva (supply chain) computerizzata.

Una soluzione di eProcurement offre diversi vantaggi, tra cui:

- Costi minori: aumento di transazioni incrociate tra imprese, che implica un più basso prezzo a fronte di volumi di acquisto maggiori
- Efficienza dei processi: l'eProcurement rende più semplice per l'azienda incrementare il controllo sugli acquisti e lavorare con i fornitori in grado di offrire le migliori condizioni.
- Aumento del controllo: l'eProcurement minimizza il rischio di acquisti non conformi agli standard fornendo un metodo efficace per controllare tutte le esigenze di acquisto.
- Accesso globale: l'eProcurement fornisce un accesso ad una base fornitori globale, che permette di trovare facilmente un partner con cui lavorare in grado di rispondere alle proprie esigenze e ridurre il costo dei materiali grazie ad un rapporto di fornitura più stretto.

L'automazione dell'acquisto indiretto di beni e servizi attraverso la quale le aziende cercano di risparmiare tempo e denaro non può essere definita una semplice "webbizzazione" del tradizionale processo di acquisizione ordini. Per questo motivo richiede la scelta di un software adeguato e un approccio graduale.

Le aziende che si sono affidate ad un software per supply chain, con lo scopo di fare

leva sulle potenzialità di Internet per ridurre i costi di approvvigionamento, stanno avendo più problemi del previsto nella transizione al business elettronico.

In base a uno studio di eMarketServices fatto in Italia [ICE, 02], circa i due terzi di queste si aspettano di processare meno del 25% delle loro transazioni nei due anni successivi all'introduzione del software per l'eProcurement. Il 20% circa fa riferimento ad una percentuale fra il 26 e il 50% nel medesimo arco di tempo, mentre solo il 10% stima di gestire via supply chain più del 50%.

In sostanza appare come le aziende abbiano sottostimato il tempo richiesto per l'implementazione dell'applicazione, inclusa la formazione, l'integrazione con i sistemi informativi aziendali e la modifica di alcuni processi interni.

Mettere in piedi un sito con 3 o 4 partner arreca qualche beneficio, ma fino a quando la maggior parte, se non la totalità, delle aziende fornitrici non è on-line, con un training ed un livello di integrazione appropriati, l'introduzione della supply chain non potrà realizzare pienamente il suo scopo.

Nondimeno, con un potenziale ed immediato 300% di ritorno dell'investimento con solo qualche fornitore on-line, la domanda di applicazioni per la supply chain è destinata a crescere, incrementando di conseguenza la richiesta di servizi di integrazione.

1.5 Prospettive dell'e-commerce

Secondo Andy Blackburn, Vice Presidente del BCG (Boston Consulting Group), il commercio elettronico B2B sta procedendo velocemente in termini di penetrazione del mercato, ma allo stesso tempo rimane immaturo nel suo sviluppo. Dal 2004 gli acquisti on-line rappresenteranno il 40% del totale, di cui solo l'11% sarà preceduto da un processo di negoziazione in rete [BCG, 02]. Per molte società la negoziazione dei prezzi su Internet è il primo step con il quale instaurare una collaborazione on-line tra venditori e compratori.

In base a queste considerazioni, nonostante l'alto livello di penetrazione degli eMarketplace, si può dire che la maggior parte delle transazioni sul Web riguarda ordini e aggiornamenti, non negoziazione dei prezzi.

A differenza di quanto si profetizzasse all'inizio, il commercio elettronico B2B non sarà distruttivo per le relazioni tradizionali tra acquirente e fornitore; anzi la maggioranza delle società afferma che la comunicazione off-line resta necessaria anche per completare le transazioni su Internet. Le relazioni personali tra clienti e fornitori rimarranno fondamentali per un business di successo.

Il ritardo nella migrazione del processo di negoziazione in rete può aprire un'opportunità ai fornitori di competere sul prezzo. I venditori possono utilizzare la tecnologia per offrire forme più sofisticate di collaborazione on-line, quali progettazione in linea, project management, coordinamento della programmazione degli approvvigionamenti. In questo modo i venditori potranno differenziare l'offerta creando relazioni più forti con il cliente e resistendo alla pressione dei prezzi.

Dalla ricerca del BCG emerge l'evoluzione dei marketplace: data la complessità e la differenziazione dell'e-commerce, compratori e venditori si aspettano di lavorare con solo pochi eMarketplace per categoria. Allo stesso tempo gli stessi eMarketplace si aspettano di avere solo pochi grandi concorrenti. Ma lo sviluppo di standard riconosciuti e di strumenti adeguati per supportare queste ambizioni è ancora insufficiente.

Pertanto diventa importante, se non addirittura indispensabile, il contributo della ricerca scientifica e industriale per mettere a punto i tool e gli standard necessari.

Capitolo 2

Ontologia e Web Semantico

L'ontologia – “lo studio dell'essere in quanto essere”, diceva Aristotele – è usualmente concepita come una disciplina strettamente filosofica, lontana dal mondo della tecnologia.

Eppure, negli ultimi anni l'esplosione delle comunicazioni in rete ha favorito un fenomeno che sarebbe stato difficile immaginare: gli aspetti ontologici dell'informazione hanno acquistato un valore strategico. Tali aspetti sono intrinsecamente indipendenti dalle forme di codifica dell'informazione stessa, che può essere quindi isolata, recuperata, organizzata, integrata in base a ciò che più conta: il suo contenuto. La standardizzazione di tali contenuti risulta oggi cruciale nella prospettiva delle aziende integrate e del commercio elettronico ed è indispensabile per semplificare i processi di comunicazione.

Salvo alcune eccezioni, gli aspetti di natura ontologica hanno avuto in passato un ruolo molto limitato nel mondo informatico. Il World Wide Web ha drasticamente aumentato la disponibilità e l'accessibilità delle informazioni; già nel 1998 una ricerca di Bharat & Broder stimava in 300 milioni il numero di documenti statici visitabili on-line [Bharat, 98]. Allo stesso tempo l'ammontare di questi dati ha reso sempre più difficile trovare, accedere, presentare e mantenere le informazioni cercate da una grande varietà di utilizzatori. Questo perché le informazioni sono presentate nel loro linguaggio naturale.

Si è creato un profondo gap tra le informazioni disponibili per i vari programmi, mirati a gestire le applicazioni citate e le informazioni mantenute in forma leggibile dall'utente.

Ricerca e presentare le informazioni nel modo desiderato è molto difficile. I risultati spesso sono imprecisi e troppo ampi. L'utente si trova a dover prendere in considerazione tutti i documenti trovati per poter estrarre le informazioni cercate.

Lo sviluppo del commercio elettronico è stato profondamente ostacolato dagli attuali metodi usati per presentare le informazioni. Gli strumenti disponibili sono insufficienti per estrarre informazioni relative ai prodotti da informazioni di testo debolmente strutturate. Inoltre i costi di sviluppo e mantenimento sono elevati e i servizi forniti sono limitati.

In risposta a questi problemi, molte nuove iniziative sono state intraprese per arricchire le informazioni disponibili con una semantica processabile dalla macchina.

Un supporto di questo tipo è essenziale per portare il Web a piene potenzialità. La rappresentazione esplicita della semantica sottostante a determinati programmi e pagine abilita un Web basato sulle conoscenze che fornisce un nuovo livello qualitativo di servizio. I servizi automatizzati miglioreranno la loro capacità di servire gli utenti nel raggiungere i loro obiettivi, comprendendo meglio i contenuti presenti nel Web e, quindi, provvedendo a filtrare e a ricercare più accuratamente le informazioni.

Per arrivare a questa forma evoluta del Web, il cosiddetto Web semantico, un punto chiave è dato dal supporto ontologico on-line allo scambio e all'integrazione di dati e informazioni. L'ontologia, utilizzata per descrivere la struttura e la semantica delle informazioni scambiate, sta diventando un elemento fondamentale in aree come la gestione della conoscenza e il B2B e-commerce [Ding, 02].

2.1 Il Web Semantico

L'aumento delle informazioni disponibili on-line necessita di un potere espressivo sempre maggiore del sistema. Obiettivo del Web Semantico è quello di avere un linguaggio capace di esprimere sia i dati che la sintassi dei dati, per esportare sul Web le regole dei sistemi di rappresentazione della conoscenza già esistenti [Berners-Lee, 01].

Per dare una struttura al contenuto di significati delle pagine Web è stato sviluppato il concetto di Web Semantico (Semantic Web), inteso come ambiente capace di far comprendere alle macchine i dati e i documenti semantici, non i testi e i discorsi umani. Il Semantic Web non è una rete distinta, ma un'estensione del World Wide Web, nella quale le informazioni sono date con significati ben definiti, capaci di far lavorare in cooperazione computer e persone.

La proprietà essenziale del World Wide Web è l'universalità. Il potere di un link di ipertesto è che "ogni cosa può linkare a ogni altra". Le tecnologie Web non devono quindi discriminare tra informazioni commerciali o accademiche, tra culture, linguaggi, media e così via. Le informazioni variano su diversi assi; uno di questi è la differenza tra informazioni prodotte principalmente per uso umano e quelle prodotte per le macchine. Da un lato abbiamo di tutto, dalle televisioni alla poesia; dall'altro database e programmi. Il Web Semantico ha lo scopo di mediare tra i due punti di vista.

Pertanto la sfida è quella di fornire un linguaggio che esprima i dati e le regole che stanno alla loro base e allo stesso tempo permetta di esportare sul Web le regole a partire da qualunque sistema di rappresentazione della conoscenza esistente.

2.2 Che cos'è l'ontologia

Le ontologie sono tecnologie essenziali, in quanto intrecciano una semantica formale comprensibile per un computer, con una semantica del mondo reale comprensibile per l'uomo.

Fin dall'inizio degli anni '90 le ontologie sono diventate un argomento di studio per molte comunità di ricerca sull'Intelligenza Artificiale. Più recentemente il concetto di ontologia si è esteso in campi come l'Integrazione Intelligente delle Informazioni, i sistemi cooperativi di informazione, il commercio elettronico, la gestione delle conoscenze. Il motivo della popolarità delle ontologie è in quello che promettono: una comprensione condivisa e comune di alcuni domini che possono essere comunicati tra persone e sistemi di applicazioni. Dato che le ontologie mirano a una conoscenza condivisa dei domini, il loro sviluppo è spesso un processo cooperativo che coinvolge diversi soggetti con esperienza e background differenti.

E' possibile definire le ontologie come "specifiche formali (machine-readable) ed esplicite di concettualizzazioni condivise"[Gruber, 93]. Se confrontiamo questa definizione con una data a proposito del Web Semantico da Berners-Lee ("strutturazione concettuale del Web in un modo esplicitamente machine-readable") [Berners-Lee, 01, si capisce che l'ontologia giocherà un ruolo chiave e pertanto sarà applicata in aree chiave del Semantic Web: tra questi l'e-commerce.

Per capire la definizione occorre spiegare i termini salienti: "Concettualizzazione" si riferisce al modello astratto di alcuni fenomeni della realtà che identifica i concetti rilevanti del fenomeno; "Esplicita" riflette il fatto che i tipi di concetti usati e i vincoli di utilizzo sono dichiarati; "Formale" si riferisce al fatto che l'ontologia deve essere leggibile dalla macchina. Ci sono vari gradi di formalità. Ampie ontologie come quella di WordNet forniscono un thesaurus di oltre 100000 termini che sono spiegati in un linguaggio naturale.

"Condivisa" indica che l'ontologia riflette una conoscenza condivisa, che non è ristretta a pochi individui, ma è allargata al gruppo.

Di base, il ruolo delle ontologie nei processi di ingegnerizzazione delle conoscenze è quello di facilitare la costruzione di un modello dei domini; l'ontologia fornisce un vocabolario di termini e relazioni con i quali modellizzare il dominio.

A seconda del livello di generalità necessario è possibile identificare vari tipi di ontologia che soddisfano diverse esigenze nei processi di condivisione e scambio della conoscenza [Guarino, 1998], [Van Heijst, 97]:

- Ontologie di dominio: catturano la conoscenza, valida per un particolare tipo di dominio (ad esempio valida per il dominio meccanico, elettronico, ...);
- Ontologie di metadati³: forniscono un vocabolario per descrivere i contenuti delle sorgenti di informazioni on-line [Weibel, 95];
- Ontologie di "senso comune": mirano a rappresentare conoscenze di senso comune, dando nozioni e concetti di base, come il tempo, lo spazio, ... [Fridman-Noy & Hafner, 97];

³ Metadata: informazione descrittiva relativa ai dati di una risorsa, compresi il dominio, la proprietà, le restrizioni, il modello di dati.

- Altri tipi di ontologie, cosiddette ontologie “method and task” [Fensel & Groenboom, 97], [Studer, 96]: forniscono un punto di vista ragionevole sulla conoscenza di dominio.

2.3 Ontologia dei prodotti

Al fine di scambiare informazioni su beni e servizi coinvolti nelle transazioni del commercio elettronico B2B, hanno acquisito un'importanza rilevante le ontologie di prodotto (content standard o standard di classificazione), che rappresentano la frontiera dei possibili sviluppi della ricerca nel tentativo di rendere semiautomatici i processi di trasmissione delle conoscenze necessarie a venditori e compratori che operano all'interno degli eMarketplace.

I content standard forniscono gerarchie di descrizione dei prodotti e definiscono sottoclassi di relazioni tra categorie di prodotti utilizzate per la ricerca intelligente di beni o servizi, per la composizione dei prodotti e per l'allocazione dei costi. Ogni prodotto di un catalogo ha un suo link a una certa categoria di prodotti, che attualmente descrive il prodotto.

Il più noto standard di classificazione prodotti è UNSPSC che presenta una gerarchia di codifica a 4 livelli con circa 15000 categorie. Non è descrittivo, in altre parole non contiene gli attributi per i prodotti, ma solo la gerarchia dei nomi dei prodotti. La gerarchia supporta la prospettiva dell'approvvigionamento per la classificazione.

Un altro content standard è ecl@ss che supporta il flusso di informazioni di prodotti lungo la catena logistica di un'impresa. È usato principalmente in Germania. Fornisce quasi 13000 categorie con attributi studiati in base alle esigenze di clienti e fornitori. La gerarchia è strutturata in accordo con le necessità del compratore. Ad esempio in ecl@ss “writing paper” appartiene agli approvvigionamenti per l'ufficio insieme a “pen”; in UNSPSC “pen” fa parte dell'approvvigionamento per l'ufficio, mentre “paper” fa parte dei prodotti per stampare. Essenzialmente i content standard sono ontologie di prodotti che danno una gerarchia di classi di prodotti insieme con gli attributi.

Nel prossimo capitolo le ontologie di prodotto saranno trattate in modo approfondito.

2.3.1 Obiettivi dei marketplace per la gestione delle ontologie

I marketplace elettronici devono portare avanti varie attività di gestione dell'ontologia per favorire lo scambio di informazioni tra soggetti economici [Fensel, 01a]:

- **Classificazione dei prodotti:** significa assegnare una certa classe dell'ontologia di prodotto al prodotto stesso, avendo come input la descrizione testuale del prodotto;
- **Ri-classificazione dei prodotti:** significa assegnare una nuova classe a un prodotto in accordo con una nuova ontologia di prodotto, noto che la classificazione del prodotto è data in qualche altra ontologia. Ciò richiede l'uso di complicate tecniche di allineamento/integrazione delle ontologie;
- **Personalizzazione** di gerarchie di prodotto e viste, generate su diverse ontologie di prodotto;
- **Mantenimento** di differenti versioni dell'ontologia, richiesto dalle frequenti variazioni nei content standard (ad esempio UNSPSC viene aggiornato ogni due mesi).

Da queste considerazioni si deduce che la gestione delle ontologie di prodotto (anche conosciuta come Content Management) riveste un ruolo prominente per le tecnologie del Semantic Web, in quanto fornisce numerose esperienze pratiche e requisiti essenziali.

I prossimi capitoli approfondiranno pertanto la conoscenza dei content standard, intesi come colli di bottiglia per la crescita futura degli eMarketplace.

2.4 Interoperabilità tra sistemi informativi

La popolarità di Internet e il grande sviluppo di nuove tecnologie ha portato alla creazione di tantissime applicazioni per l'e-commerce. Comunque la tecnologia non è l'unico fattore chiave per lo sviluppo di applicazioni elettroniche. Il contesto dell'e-commerce e specialmente il contesto del B2B, richiede che sia possibile un'effettiva comunicazione tra macchine. In altre parole, nella comunicazione, si rivela cruciale

l'interoperabilità semantica tra sistemi informativi.

Due fattori contribuiscono alla comunicazione non umana:

- **Dimensione sintattica:** il linguaggio comune nel quale le risorse implicate nella comunicazione possono essere espresse;
- **Dimensione semantica:** un modello informativo condiviso e un vocabolario tra i diversi sistemi presenti nell'intero processo.

La prima dimensione ha portato alla nascita di vari linguaggi (XOL [Karp, 99], SHOE [Luke, 97], OML [Kent, 99], RDF [Miller, 98], RDF Schema, OIL [Fensel, 01c], DAML+OIL). La dimensione semantica è invece in relazione con il modello di conoscenza e il vocabolario usato dai sistemi sviluppati nella comunicazione.

In tale senso l'uso di vocabolari e modelli condivisi aumenta l'interoperabilità tra sistemi informativi esistenti e futuri.

Di seguito si illustreranno alcune tecnologie utilizzate nello sviluppo della tesi, in parte riferite alla dimensione sintattica, in parte alla dimensione semantica

2.5 Tecnologie del Web semantico

In questa sezione si presentano vari linguaggi e strumenti per fornire una struttura di base al Web Semantico.

Un requisito importante per le informazioni che devono essere processate dalla macchina è la struttura dei dati. I database, per esempio, utilizzano tabelle con attributi per strutturare i dati. La principale struttura utilizzata sul Web è il document markup.

Ciò significa che certe sequenze di caratteri in un documento contengono informazioni che indicano il contenuto del documento. Il markup descrive il layout del documento e la struttura logica rendendo il documento auto descrittivo. Il markup prende spesso la forma di parole racchiuse tra parentesi angolari chiamate tags (<....>).

Un linguaggio di markup specifica quali markup tags possono essere utilizzati, come devono essere combinati e qual è il loro significato. HTML è il linguaggio più conosciuto e a più ampia copertura sul Web. Le uniche informazioni che HTML fornisce riguardano la struttura di contenuti, ma spesso ciò non è sufficiente per i servizi

automatizzati che sono obiettivo del Web semantico. Le applicazioni richiedono dei markup specializzati che specifichino il ruolo delle parti di programma rilevanti rispetto al contenuto.

2.5.1 XML: eXtensible Markup Language

XML è un meta linguaggio che incontra la necessità di definire markup tag specifici per l'applicazione. XML è un meccanismo per la rappresentazione standardizzata di altri linguaggi. In altre parole XML fornisce un formato di dati per documenti strutturati senza specificare il vocabolario. Per definire un vocabolario o le combinazioni di tag accettate si possono usare le DTD, Document Type Definition o un XML schema. Un linguaggio definito in XML è chiamato applicazione XML.

XML non implica una specifica interpretazione dei dati. L'unica interpretazione legittima del codice XML è che contiene il nome di entità con sottoentità e valori; ogni documento XML forma un albero ordinato con le sue etichette.

Questa generalità è sia un punto di forza sia una debolezza di XML. XML non specifica l'uso e la semantica dei dati. Chi usa XML per lo scambio di dati deve prima concordare un dizionario, il suo uso e il suo significato.

2.5.2 RDF e RDF SCHEMA

XML è un linguaggio standard per strutturare i dati, mentre il Resource Description Framework (RDF) è un meccanismo per dare informazioni sui dati, ad esempio per dargli significato. Come si intuisce dal nome non è un linguaggio, ma un modello per rappresentare i dati presenti sul Web, che vengono chiamati metadati.

Gli oggetti sono risorse nel vocabolario RDF. Le risorse sono tutti gli oggetti che hanno un identificatore nel Web, a partire dalle parti del documento specifiche e identificabili, fino ad arrivare ad oggetti come le persone o le imprese.

Il modello di dati RDF è molto semplice. Si compone di statement relativi alle risorse, codificati come triple nella forma oggetto - attributo - valore. Gli oggetti sono risorse e i valori sono risorse o stringhe.

Lo schema RDF fornisce un modo per definire specifiche proprietà del dominio e le classi di risorse cui queste proprietà possono essere applicate. La primitiva base di modellazione nello schema RDF consiste di:

1. Definizione di classi e sottoclassi che insieme permettono la definizione di una gerarchia di classi;
2. Definizione di proprietà e di sotto proprietà degli statement;
3. Domini e range statement, per restringere le possibili combinazioni di classi e proprietà;
4. Il tipo degli statement.

Queste primitive possono essere usate per costruire lo schema di uno specifico dominio. RDF e lo schema RDF forniscono un semplice meccanismo di rappresentazione della conoscenza per le risorse Web. Per una spiegazione più chiara del contenuto dei dati sono richiesti linguaggi più ricchi.

2.6 Il database lessicale WordNet

WordNet è un database lessicale elettronico, la cui struttura è ispirata alle teorie psicolinguistiche legate alla memoria lessicale umana. È stato sviluppato presso l'università di Princeton dal Cognitive Science Laboratory sotto la direzione del professor A. Gorge Miller⁴ ed esiste anche in una versione multilingue europea⁵.

Il successo di WordNet si basa sul fatto che è disponibile on-line, in versione freeware, costruito sui concetti e quindi più ricco rispetto a una lista di parole in ordine alfabetico come i dizionari tradizionali.

WordNet organizza nomi, verbi, aggettivi ed avverbi in insiemi di sinonimi (synset), dove ognuno di loro rappresenta un concetto. I synset inoltre sono legati tra loro attraverso diverse relazioni, come la sinonimia, antonimia, iponimia, iperonimia, ecc.

La differenza dalle classiche procedure per organizzare informazioni lessicali, come ad esempio i dizionari, consiste nel fatto che le parole non sono organizzate in una lunga lista ordinata, ma sono legate in modo gerarchico.

⁴ <http://www.cogsci.princeton.edu/~wn/>

⁵ <http://www.let.uva.nl/~ewn>

In questo modo WordNet conserva tutte le caratteristiche dei vecchi dizionari aggiungendo però degli strumenti in più; infatti se, ad esempio, si inserisce la parola *machine* si ottiene una overview (Figura 7) della stessa, completa di tutti i significati, con i sinonimi e la categoria (nome, verbo) di appartenenza; quindi per ogni categoria, se si specifica l'esatto significato, si può scegliere il tipo di relazione da visualizzare; infine vengono visualizzati tutti i synset legati in modo gerarchico alla relazione scelta (Figura 8).

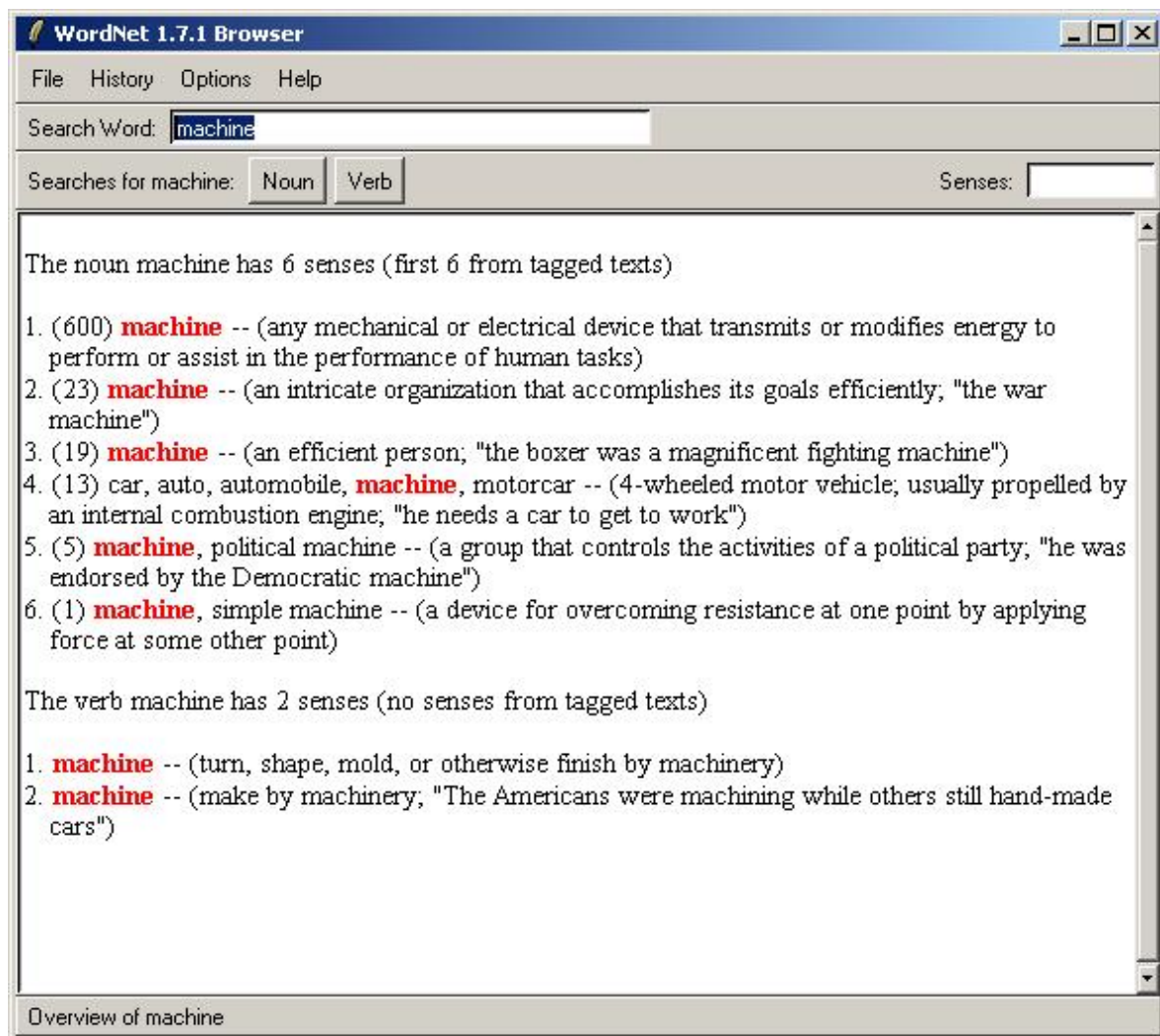


Figura 7 Overview di WordNet

Ogni gerarchia, generata dai diversi significati di *machine*, termina in un nome che viene chiamato beginner. In WordNet sono presenti 9 diversi beginner (Figura 9), dai quali si generano tutti i nomi presenti nel database.

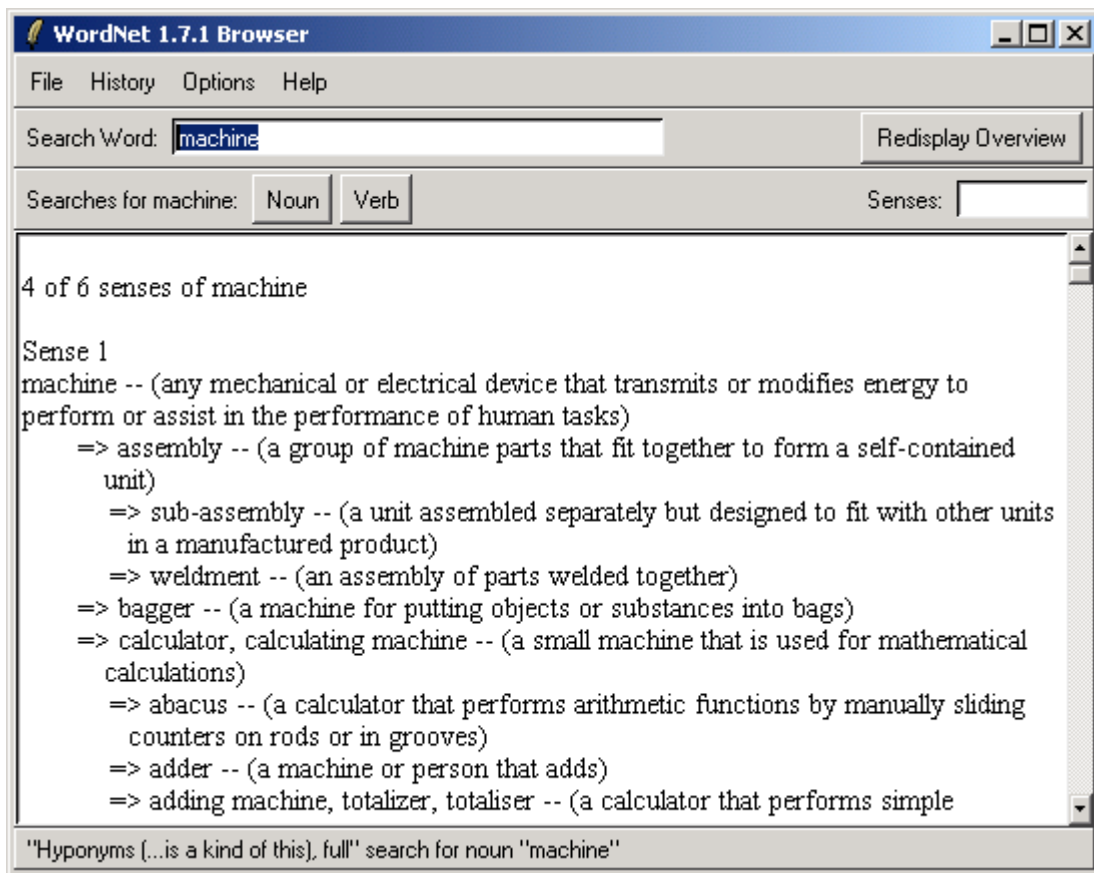


Figura 8 Relazione di iponimia per la definizione *machine*

{abstraction}
{act, human action, human activity}
{entity, something}
{event}
{group, grouping}
{phenomenon}
{possession}
{psychological feature}
{state}

Figura 9 Lista dei beginner

WordNet è considerato la più importante risorsa disponibile per i ricercatori nei campi della linguistica computazionale, dell'analisi testuale, e di altre aree associate. Attualmente è arrivato alla versione 1.7.1 e contiene 129625 lemmi organizzati in 99759 synset divisi in 4 categorie: nomi, verbi, aggettivi e avverbi. Questa suddivisione in categorie, pur comportando una ridondanza rispetto ai dizionari classici (ad esempio la

parola *back* che è presente in tutte le categorie), ha il vantaggio di visualizzare, in modo chiaro, le diverse categorie e di sfruttarle al meglio in base alle loro diverse strutture

Ogni categoria è organizzata in modo diverso:

- I nomi in 9 gerarchie individuate da altrettanti beginners, generate da relazioni di iponimia/iperonimia;
- I verbi in 617 gerarchie, generate da varie relazioni d'implicazione (*entailment relation*);
- Gli aggettivi e gli avverbi in iperspazi N-dimensionali (cluster).

Tali diverse strutture sono necessarie, al fine di meglio rappresentare la complessità psicologica della conoscenza lessicale.

2.6.1 La Matrice Lessicale

Alla base del progetto del WordNet c'è la Matrice Lessicale. La necessità di creare una matrice lessicale nasce dall'osservazione che una parola è un'associazione convenzionale tra il suo significato e il modo in cui viene letta o scritta. Per ridurre l'ambiguità derivata dal termine parola useremo:

- **Word meaning** o significato, se ci riferiamo al concetto lessicale o significato;
- **Word form** (Forma Base) o lemma, se ci riferiamo al modo in cui viene letta o scritta.

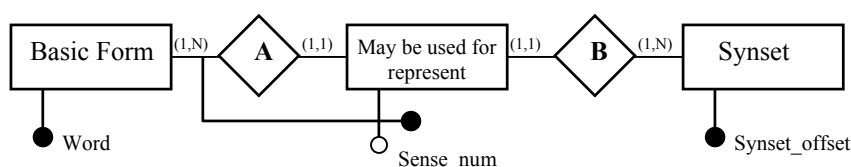


Figura 10 Relazione tra lemmi e significati

Quest'associazione è di molti a molti (Figura 10), e dà luogo alle seguenti proprietà:

- **Sinonimia**: proprietà di un significato di avere due o più parole in grado di esprimerlo.
- **Polisemia**: proprietà di una parola di esprimere due o più significati.

Questa relazione esistente tra le word-form e le word-meaning può essere sintetizzato attraverso la matrice lessicale (Figura 11), dove nelle righe sono rappresentate le word meaning (significati) e nelle colonne le word form (forma/lemma base).

Ogni cella della matrice ci indica che il lemma, in quella colonna, può essere rappresentato per esprimere il significato in quella riga; quindi, la cella $E_{1,1}$ implica che il lemma F_1 può essere usato per esprimere il significato M_1 ; nel caso ci fossero due lemmi nella stessa colonna, significherebbe che il lemma è polisemo (ha più significati), mentre nel caso di due lemmi nella stessa riga significherebbe che i due lemmi sono sinonimi.

Word	Word Forms				
Meaning	F_1	F_2	F_3	-----	F_n
M_1	$E_{1,1}$	$E_{1,2}$			
M_2		$E_{2,2}$			
M_3			$E_{3,3}$		
-					
-					
-					
M_m					$E_{m,n}$

Figura 11 Matrice Lessicale

Dalla Figura 11 possiamo affermare che F_1 e F_2 sono sinonimi e che F_2 è polisemo. Ogni cella “ e ” è una definizione (*entry*), $e=(f, m)$, dove f individua la word-form e m la word-meaning.

2.6.2 Relazioni lessicale e sintattiche

Inizialmente WordNet è stato ideato come un insieme di relazioni semantiche tra concetti lessicali, ma con il proseguire del progetto si è osservato che c’era un altro tipo di relazione che non poteva essere trascurato e sono le relazioni lessicali.

Attualmente, anche se ad essere enfatizzate sono le sole relazioni semantiche, occorre ricordare che WordNet organizza le relazioni in due principali categorie:

- **Relazioni semantiche:** le quali esprimono un legame tra i significati come la specializzazione e la generalizzazione (iponimia e iperonimia);

- **Relazioni lessicali:** le quali esprimono un legame tra i lemmi, come la sinonimia, la polisemia e l'antonimia.

Abbiamo visto come WordNet è organizzato in relazioni semantiche e ognuna di esse è una relazione tra significati; si è anche visto che i significati sono organizzati in synset, quindi è possibile immaginare una relazione semantica come un puntatore tra synset. Caratteristica delle relazioni semantiche è la reciprocità: se esiste una relazione R tra il significato $\{x, x', \dots\}$ e il $\{y, y', \dots\}$, allora esiste anche una relazione R' tra $\{y, y', \dots\}$ e $\{x, x', \dots\}$.

2.6.2.1 Sinonimia

Da quanto argomentato, possiamo affermare che la sinonimia è la relazione più importante del WordNet, infatti, un buon conoscitore della lingua inglese, dalla sola relazione di sinonimia tra i vari lemmi, può risalire al significato.

Per distinguere la relazioni di sinonimia dalle altre relazioni, i termini sinonimi sono racchiusi fra parentesi graffe {}, mentre gli altri insiemi prodotti dalle altre relazioni lessicali sono racchiusi fra parentesi quadre [].

La definizione di sinonimia, è generalmente attribuita a Leibnitz, secondo il quale:

“Due espressioni sono sinonime se la sostituzione di una per l'altra non cambia il vero significato della frase nella quale è fatta la sostituzione.”

Questa definizione impone una condizione estrema e rende i sinonimi molto rari o inesistenti. Una seconda definizione invece fa cadere l'ipotesi sul contesto:

“Due espressioni sono sinonime in un contesto linguistico C se la sostituzione di una per l'altra, nel contesto C , non altera il vero valore della frase.”

Per meglio comprendere, prendiamo come esempio (*blank*) che, in un contesto di falegnameria, può quasi sempre essere sostituito con (*plank*) senza alterare il significato, mentre in altri contesti questa sostituzione risulterebbe inappropriata.

La definizione di sinonimia in termini di sostituibilità divide necessariamente il WordNet in nomi, verbi, aggettivi e avverbi, infatti, parole che appartengono a diverse categorie sintattiche non possono essere interscambiate.

2.6.2.2 Antinomia

L'antinomia è una relazione lessicale tra lemmi, la quale indica che uno è il contrario dell'altro. Solitamente l'antinomia di una parola "x" è "non x". Pur apparendo come una semplice relazione di simmetria l'antinomia è una relazione abbastanza complessa. Occorre tenere presente che l'antinomia è una relazione lessicale tra word form, e non una relazione semantica tra word meaning; i significati {*rise, ascend*} e {*fall, descend*} possono essere concettualmente opposti, ma non sono antinomi; mentre sono esempi di antinomie [*rise/fall*] così come [*ascend/descend*].

2.6.2.3 Iponimia/Iperonimia

Diversamente dalle relazioni di sinonimia ed antinomia, le quali erano relazioni lessicali tra word forms, *iponimia/iperonimia* è una relazione semantica tra word meanings: per esempio, {*maple*} (acero) è un iponimo di {*tree*}, e {*tree*} è un iponimo di {*plant*}. Un concetto rappresentato dal synset {*x, x', ...*} è detto iponimo del concetto rappresentato da {*y, y', ...*} se un madrelingua inglese convalida la frase costruita come: *x is a (kind of) y*.

La relazione di iponimia è transitiva e simmetrica ed è equivalente alla relazione di specializzazione. Ogni iponimo eredita tutte le caratteristiche dei concetti più generici e aggiunge almeno uno per distinguerlo dal sovraordinato e dagli altri iponimi del sovraordinato. La corrispondente relazione inversa, equivalente alla generalizzazione, è l'iperonimia, la quale lega un sovraordinato, ai subordinati, ereditando dagli iponimi le caratteristiche generali.

La relazione di iponimia/iperonimia stabilisce la regola centrale per l'organizzazione dei nomi in WordNet, essa è definita solo per i nomi e i verbi, ed è la relazione più frequente in WordNet. Nel caso dei verbi la relazione di iponimia è chiamata troponimia: un verbo che esprime una maniera specifica di operare di un altro verbo: *X* è un troponimo di *Y* se fare *X* è fare *Y* in qualche maniera.

2.6.2.4 Meronimia

La meronimia è una relazione semantica definita dal punto di vista lessicale come:

Un concetto rappresentato dal synset {x,x',...} è un meronimo di un concetto

rappresentato dal synset $\{y,y',...\}$ se un madrelingua inglese convalida la frase costruita come: *An y has an x (as a part) or An x is a part of y.*

Sono definiti tre tipi di aggregazione:

1. **HAS MEMBER** – esempio *school* inteso come istituzione educativa ha due aggregazioni:
 - *school HAS MEMBER: staff, faculty*
 - *school HAS MEMBER: schoolteacher, school teacher;*
2. **HAS PART** – esempio *school* inteso come luogo dove si riceve l'educazione ha una sola aggregazione:
 - *school HAS PART: classroom, schoolroom;*
3. **HAS SUBSTANCE** – esempio *quartz* inteso come minerale ha due aggregazioni:
 - *quartz HAS SUBSTANCE: silicon, Si, atomic number 14*
 - *quartz HAS SUBSTANCE: silica, silicon oxide, silicon dioxide.*

La meronimia è una relazione transitiva e asimmetrica, la sua relazione duale è la olonimia; anch'essa può essere usata per costruire gerarchie di concetti meronimi/olonimi, con la differenza che, in questo caso, uno stesso meronimo può avere più olonimi: in altri termini, un concetto può contemporaneamente far parte di differenti concetti composti.

I meronimi sono fattori distintivi che gli iponimi possono ereditare. Per esempio, se *beak* (becco) e *wing* (ala) sono meronimi di *bird* (uccello) e *canary* (canarino) è un iponimo di *bird*, allora, ereditando, *beak* e *wing* deve anche essere meronimo di *canary*.

Quando tutte e tre le relazioni, iponimia, meronimia, e antinomia si incrociano, il risultato è altamente interconnesso in una rete complessa; sapere dove una parola è situata in questa rete è un'importante informazione per la conoscenza del significato della parola.

2.6.2.5 Relazioni morfologiche

Un'altra importante classe di relazioni lessicali sono le relazioni morfologiche tra word form. Inizialmente questo tipo di relazione non era stato considerato ma, con l'avanzamento del progetto, cresceva la necessità di inserire questo nuovo tipo di

relazioni, infatti, bastava inserire un termine al plurale, ad esempio (*trees*), per far sì che WordNet dava come risultato che il termine non era incluso nel database.

La morfologia delle parole nella lingua inglese è abbastanza semplice e si manifesta come:

- Declinazioni per i sostantivi, ad esempio distinzione fra singolare e plurale;
- Coniugazione per i verbi.

Per far sì che WordNet riconoscesse queste forme è stato inserito un software che non modifica la struttura del database, nel senso che, all'interno le parole non sono replicate, infatti, le parole sono memorizzate nella forma canonica. Il software introdotto ha il compito di interfacciarsi fra l'utente e il database lessicale, in modo tale da tradurre ogni termine in input nella forma canonica e, successivamente, inviarlo al database.

Nonostante la semplicità della morfologia inglese rispetto le altre lingue, la realizzazione di questo componente non lo è stata a causa della massiccia presenza di verbi irregolari.

2.6.3 Categorie sintattiche

2.6.3.1 Nomi in WordNet

Nelle moderne teorie psicolinguistiche, per definire un nome si usa un termine sovraordinato insieme a fattori che lo distinguono; questo metodo è anche alla base dell'organizzazione dei nomi in WordNet, infatti, la relazione di iponimia genera una gerarchia semantica. La gerarchia generata è limitata in profondità e raramente raggiunge il dodicesimo livello. Un modo per costruire tale gerarchia è di far rientrare tutti i nomi in una singola gerarchia, in questo caso, il livello più alto dovrebbe essere semanticamente vuoto. Inizialmente si era indicato come radice *{entity}*, seguito dagli immediati iponimi: *{object/ thing}* e *{idea}*. Successivamente si è visto che, anche questi generici concetti astratti, portavano una piccola quantità di informazione semantica. L'alternativa è stata di partizionare i nomi con un insieme di unique beginner dai quali si generano altrettante gerarchie.

Approssimativamente WordNet contiene 57,000 nomi (word forms) organizzati in 48,000 word meaning (synset).

2.6.3.2 Aggettivi e avverbi in WordNet

Tutti i linguaggi assegnano agli aggettivi la proprietà di modificatori o elaboratori del significato dei nomi; quella di modificatori di nomi è la caratteristica principale associata alla categoria degli aggettivi.

I synset degli aggettivi in WordNet contengono principalmente aggettivi, anche se sono state anche aggiunti nomi e frasi preposizionali che hanno la funzione di modificatori.

WordNet contiene approssimativamente 19,500 aggettivi in word forms, organizzati in 10,000 word meanings (synset).

Gli aggettivi in WordNet sono suddivisi in due principali classi:

- **Descrittivi** - come *big*, *interesting*, *possibile* - che attribuiscono ai nomi valori di attributi bipolari e, conseguentemente, sono organizzati in termini di opposizioni binarie (antinomia) e similitudine di significato. L'organizzazione semantica degli aggettivi descrittivi è completamente differente da quella dei nomi; infatti, nella loro organizzazione non esiste nessuna relazione di iponimia poiché non avrebbe nessun senso dire che un aggettivo "*is a kind of*" un altro aggettivo.

La principale relazione per gli aggettivi è l'antinomia, ciò si evince dalla funzione degli aggettivi, che è quella di attribuire valore agli attributi, e dal fatto che quasi tutti gli attributi sono bipolari. Gli aggettivi che non sono bipolari sono collegati per similitudine semantica agli aggettivi bipolari. Possiamo immaginare l'organizzazione degli aggettivi come cluster che contengono un synset principale e dei synset satelliti. Le coppie di antinomi sono nei synset principali dei cluster e sono dette relazioni di antinomia dirette. I synset principali sono collegati ai synset satelliti da una relazione di similarità

- **Relazionali** – come *presidential*, *nuclear*, *musical* che sono ritenuti varianti stilistiche dei modificatori dei nomi e sono considerati come dei riferimenti incrociati ai nomi, inoltre non hanno antinomi e non sono organizzati in cluster.

Gli avverbi sono simili agli aggettivi, infatti, spesso derivano da essi. La loro proprietà è di modificare il significato dei verbi.

2.6.3.3 Verbi

I verbi sono la categoria più importante, infatti, ogni frase ne deve contenere almeno uno; ma, ciò nonostante, nel linguaggio ci sono molto meno verbi che nomi.

WordNet contiene 21,000 verbi in word forms (dei quali circa 13,000 sono stringhe uniche) e circa 8,400 word meaning (synset). Sono anche incluse espressioni come: *look up* e *fall back*.

I verbi, come i nomi, sono organizzati in una gerarchia di specializzazione, ma anziché avere 9 unique beginner ne hanno 617. Inoltre queste gerarchie sono più piatte rispetto a quelle dei nomi.

La relazione di iponimia è sostituita con quella di troponimia, che è un particolare tipo d'implicazione; ogni troponimo V_1 di uno o più verbi V_2 *implica anche* V_2 . Consideriamo la coppia *limp-walk* (zoppicare-camminare) se è esatta la frase: “*to limp* è anche *to walk* in una certa maniera”, questo mi dice che *limp* è troponimo di *walk*; in questo caso i verbi sono anche in una relazione d'implicazione (*entailment relation*), infatti “*He is limping*” implica “*He is walking*”.

Capitolo 3

Sistemi standard di classificazione dei prodotti per l'E-Commerce

Il gran numero di siti Internet, nati negli ultimi anni, ha incrementato la disponibilità di informazioni sul Web, ma, a causa della sua struttura, le informazioni che viaggiano in rete sono sempre meno “machine-readable” e “machine-understandable”⁶.

Molte organizzazioni, nel contesto del commercio elettronico, hanno sfruttato le opportunità offerte dalle soluzioni Internet e altre si apprestano a farlo, condividendo i loro database e cataloghi con numerosi altri utenti. Di conseguenza non solo è cresciuto il numero di informazioni disponibili, ma è aumentata la “confusione” generata dalla sovrabbondanza di sorgenti di dati. In questo contesto si è sviluppato il ruolo di intermediari assunto dagli eMarketplace, che assumeranno un valore strategico nel futuro del commercio elettronico, come arene imparziali in cui la domanda di un compratore potrà teoricamente scegliere tra i prodotti distribuiti e offerti sul mercato da numerosi venditori. Ma il futuro dei Marketplace elettronici è condizionato dalla mancanza (o dalla sovrabbondanza) di ontologie standard che descrivano e classifichino beni e servizi in modo completo.

Una transazione economica sul Web è resa difficile dal bisogno degli acquirenti di

⁶ Le informazioni scambiate su Internet si presentano spesso in un formato non leggibile o comprensibile per i calcolatori (no machine-readable o machine-understandable), quindi non possono essere processate.

utilizzare solo uno standard per classificare e descrivere i prodotti forniti dai venditori, così da assicurare una facile integrazione con i sistemi ERP aziendali. Perciò l'eMarketplace dovrebbe mettere a disposizione un ambiente che sappia mediare tra i differenti standard utilizzati da chi partecipa alla transazione, in modo tale che ogni attore possa utilizzare il proprio formato.

Per concludere, l'esigenza dei B2B eMarketplace è ri-classificare beni e prodotti in accordo ai diversi modelli di standardizzazione (vedi sezione [2.3.1]).

Il Granada Research in [Granada, 02] presenta una formulazione completa di definizioni e specifiche relative al vasto panorama degli standard di classificazione. Nella prima parte del capitolo si farà riferimento a questa ricerca per illustrare i requisiti e i benefici degli standard. Nella seconda si illustreranno gli standard di classificazione utilizzati nella fase sperimentale della tesi.

3.1 Perché classificare prodotti e servizi

L'adozione di standard di classificazione prodotti nella nuova era del commercio elettronico sta diventando un obbligo per facilitare il commercio tra venditori e acquirenti. Grandi compagnie stanno iniziando a identificare i loro prodotti con l'obiettivo di analizzare le voci di spesa comprese nell'acquisto di un bene.

La rapida ascesa del commercio elettronico, in particolare su Internet, e lo sviluppo di nuove tecnologie (quali linguaggio XML – sezione [2.5.1] – motori di ricerca, data mining tool⁷) rendono la codifica dei prodotti inevitabile. Le stesse organizzazioni governative cominciano a richiederla.

La maggior parte dei sistemi di codifica per le aziende sono stati molto costosi da sviluppare. Tipicamente occorre un anno per lo sviluppo, ogni oggetto preso in considerazione richiede circa un'ora e mezza per l'assegnazione di un codice; inoltre i fornitori di un'impresa non aderiscono abitualmente agli schemi di codifica dei loro clienti, ma con questi devono dialogare.

C'è stata una duplicazione di spese e di lavoro nel creare codici. Se si fosse adottata una singola, universale convenzione di codifica, anche nel caso in cui le imprese avessero

⁷ Data mining tool: strumenti per analizzare ingenti moli di dati allo scopo di individuare particolari tendenze e andamenti

desiderato adeguarla alle loro esigenze, ci sarebbe stato un grande risparmio, in termini di tempo e denaro.

3.1.1 Per trovare e acquistare

Le convenzioni di codifica di un prodotto o di un servizio apportano notevoli benefici alle funzioni di acquisto di una compagnia. Questi benefici sono elencati in Tabella 1.

Sistema di classificazione prodotti	
Pro	Contro
I compratori possono trovare i fornitori di una data categoria di prodotti	Richiede sforzo per applicare i codici
I manager degli acquisti possono analizzare le spese e svolgere approvvigionamenti strategici	
Un sistema di codifica consente, attraverso le divisioni della società e i sistemi informativi, di dare un'immagine uniforme delle spese dell'azienda	
Un unico sistema di numerazione integra l'intero flusso del processo.	
Facilita il controllo e il rispetto dei limiti di spesa...	
Integra le categorie delle procurement card con tutti i dati delle transazioni di acquisto.	
Un sistema di codifica standardizzato è facile da implementare e permette di risparmiare tempo e denaro rispetto allo sviluppo di un proprio sistema di codifica.	

Tabella 1 Classificare prodotti o servizi per trovare o acquistare

Ricerca dei prodotti. Una convenzione di denominazione comune permette ai sistemi informativi di elencare prodotti simili all'interno di una stessa categoria. Quando una persona cerca la categoria, trova precisamente le cose che sono mostrate e nient'altro. Per ogni raccolta di documenti, quando viene usato un vocabolario di termini per indicare i contenuti di ogni documento, la precisione della ricerca cresce enormemente. L'uso di semplici stringhe di testo su tali collezioni di dati porta ad un gran numero di risultati irrilevanti. Lo stesso principio si applica a record di informazioni di prodotti

dotate di struttura con un vocabolario di termini pre-impostato che è stato sviluppato dai partecipanti al business.

Analisi semplificata di spesa. Quando ogni transazione d'acquisto di un'impresa ha una struttura che individua gli identificatori di prodotto, i manager degli acquisti sono in grado di analizzare le spese dell'azienda. Con identificatori che fanno parte di una tassonomia gerarchica, le categorie individuali possono essere riunite in categorie più generali. Ciò permette all'Ufficio acquisti di:

- Consolidare i fornitori preferenziali per semplificare e costruire relazioni strategiche. Un limitato numero di fornitori, selezionati per soddisfare i bisogni dell'azienda, introduce miglioramenti nelle consegne e nei servizi.
- Risparmiare dal 15 al 20% sulla spesa totale sul volume d'acquisto. Garantendo un elevato volume di acquisti ad un ristretto numero di fornitori si ottiene uno sconto circa del 20% sui prezzi. Questo corrisponde ad un risparmio annuo del 20% sulle spese totali.

Controllo e standardizzazione tra le funzioni aziendali. I codici portano una singola, uniforme visione di tutte le spese di una società. Riuniscono insieme tutti i reparti e le divisioni, includendo le funzioni di business come gli acquisti e i pagamenti. I codici, utilizzati nei sistemi informativi, sono strumenti essenziali per un controllo rapido del catalogo prodotti e dei venditori autorizzati, dei flussi di lavoro approvati e del denaro allocato per le spese.

3.1.2 Per il marketing e la distribuzione

Quando una compagnia di vendita applica una struttura standard ai suoi prodotti e servizi, tale struttura può essere automaticamente processata da altre applicazioni informatiche, come i motori di ricerca, gli ERP, i software di aggregazione dei cataloghi, gli agenti software⁸ e simili. Con la codifica, il produttore può rapidamente diffondere nel mondo la sua offerta commerciale. In Tabella 2 si schematizzano pro e contro della standardizzazione per quanto riguarda il marketing e la distribuzione.

⁸ Sul Web gli agenti software (software agent) sono programmi che raccolgono informazioni o migliorano altri servizi, operando senza la costante presenza umana o una programmazione precisa delle attività.

Sistema di classificazione prodotti

Pro	Contro
I propri prodotti vengono automaticamente inseriti nelle Web search e negli aggregatori di cataloghi.	Classificare i prodotti in maniera uniforme ai concorrenti dà ai clienti la possibilità di comprare a prezzo inferiore.
Soddisfare i clienti inviando cataloghi di prodotti precodificati	
Minimizzare gli errori nelle ordinazioni e i resi, aumentare la qualità delle comunicazioni con il canale di distribuzione.	
Abilita al benchmark il mercato	
Acquisisce dati chiari e consistenti dal canale di vendita	
Resta aggiornato con le tendenze del mercato: i tag standardizzati sono la chiara evoluzione del commercio elettronico.	
Un sistema di codifica che soddisfa i consumatori.	

Tabella 2 Classificare prodotti e servizi per supportare le vendite e il marketing

Distribuzione e raccolta automatica di informazioni sui prodotti. Queste tecniche portano nuove opportunità, ad esempio raggiungere un'audience più ampia con costi inferiori, ma anche molti pericoli; c'è il rischio di essere considerati un disturbo o di essere visti solo come un ulteriore venditore di un bene. Nell'era del commercio elettronico le attività di marketing includono:

- Mandare cataloghi elettronici dei prodotti ai propri clienti, distributori e partner di canale;
- Registrare i propri prodotti nei motori di ricerca e nei marketplace, nelle directory di prodotto e nei giornali elettronici di commercio;
- Pubblicizzare i propri prodotti e la propria impresa utilizzando tag XML come metadati;
- Preparare il proprio sito Web alle interazioni con agenti intelligenti.

Queste applicazioni richiedono convenzioni comuni per la denominazione di prodotti e servizi.

Coordinamento del canale di distribuzione. In aggiunta ai benefici derivanti dalla propagazione automatica delle informazioni sui prodotti in tutto il mondo, le informazioni di prodotto codificate aiutano a coordinare un canale di distribuzione. Quando un venditore vende i suoi prodotti attraverso terze parti, come i distributori, i rappresentanti indipendenti, i commercianti, i rivenditori e gli altri partner di canale, coordinare le spedizioni e la gestione dei materiali diventa un aspetto critico per il successo delle vendite.

Se il produttore e i suoi partner di canale utilizzano cataloghi e documenti per il commercio dotati di tag, il numero di errori decresce; ognuno fa riferimento agli stessi prodotti, non c'è confusione. Il produttore può ottenere dati accurati sulle vendite e sui materiali dai suoi canali e di conseguenza può fare accurate previsioni di vendita. I partner di canale possono fare richieste precise oppure ordini di nuovi prodotti o di prodotti sostitutivi. Il coordinamento del canale aumenta.

I dati di canale migliorano le analisi. La classificazione delle transazioni sul canale digitale aiuta il produttore a determinare lo stato delle attività nel canale. Ciò permette di vedere quando nasce o diminuisce la domanda per uno specifico prodotto, per una categoria o per un'area geografica. Se vari produttori usano un sistema di classificazione, possono ottenere informazioni aggregate per codice di prodotto o per categoria. Questo aiuta il produttore a capire e ad agire strategicamente nel suo canale.

Tutti i prodotti sono beni indifferenziati? Descrivere i prodotti della propria impresa con lo stesso codice utilizzato dai concorrenti sembra equivalente a ridurre tali prodotti a beni indifferenziati (commodity). Questo non è l'obiettivo. Una convenzione comune di codifica è il primo passo per mettere le informazioni sui propri prodotti nelle mani di persone interessate. Dare un nome comune per un dato tipo di prodotti è equivalente ad organizzare la disposizione delle merci in un supermercato: tutti i cereali in una sezione; gli accessori di bellezza in un'altra. Una comune convenzione di denominazione è il futuro degli scaffali, è una vetrina virtuale per il commercio elettronico. Come nel mondo fisico il marchio gioca un ruolo di potere, così giocherà un ruolo di differenziazione anche nel mondo non fisico, insieme con altri aspetti, come il servizio ai clienti, l'imballaggio, lo sviluppo di nuovi prodotti.

3.2 Gli effetti degli standard di classificazione sulla catena logistica

Internet, EDI⁹ e gli standard d'identificazione dei prodotti (includendo anche i codici a barre) hanno portato ad una profonda automazione delle transazioni. Finora il contenuto delle transazioni, le descrizioni, i nomi dei prodotti e servizi che sono stati visti, comprati o spediti hanno fatto a meno della standardizzazione. Ma le informazioni descrittive relative ai prodotti sono importanti per il marketing automatico, per la presentazione e l'analisi degli acquisti di prodotti e tali attività stanno diventando pratica comune per le funzioni di compravendita delle compagnie.

Le informazioni di prodotto, contenute in file e Web page processabili, vengono originate dal fornitore del prodotto, e automaticamente circolano attraverso il sistema elettronico a rete. I tag di identificazione dei prodotti popolano i cataloghi, le directory, i sistemi di ricerca e le applicazioni aziendali che sono svolte da clienti, distributori e rivenditori.

Gli standard metadati per i prodotti giovano a tutti gli attori della catena logistica:

- **Ufficio acquisti e funzioni correlate.** Un identificatore comune permette a chi riceve le informazioni sui prodotti di compilare e aggregare i dati provenienti da varie sorgenti, mantenendo una struttura unificata per la ricerca e l'analisi. Nella fase di approvvigionamento vi sono difficoltà a gestire efficientemente numerose transazioni di piccolo valore per le scorte di prodotti non in produzione e non in rivendita e per servizi esterni. Molte aziende non hanno molta consapevolezza dei loro volumi di spesa per gli approvvigionamenti e beni non di produzione. Gli standard di classificazione dei prodotti permettono all'ufficio acquisti di identificare e selezionare fornitori strategici e, utilizzando motori di ricerca e cataloghi controllabili, indirizzare gli acquisti dei dipendenti automaticamente verso tali fornitori;
- **Produttori.** I produttori, utilizzando codici standardizzati, possono portare la struttura elettronica globalmente interconnessa (includendo il Web, Internet e altre applicazioni a rete intercompany) nella vendita dei loro prodotti. La diffusione automatica d'informazioni sui prodotti raggiunge un gran numero di potenziali

⁹ EDI (Electronic Data Interchange) è un formato standard per lo scambio di dati nelle attività di business.

acquirenti con costi molto inferiori dei cataloghi cartacei. Inoltre, codici leggibili elettronicamente assicurano notevole accuratezza e produttività nelle comunicazioni con i partner (includendo distributori, venditori, commercianti). I codici minimizzano i resi, permettono una migliore gestione dei materiali e degli assortimenti, permettono la raccolta di informazioni standardizzate relative al mercato;

- **Distributori, rivenditori e altri partner di canale.** Chi fornisce servizi di assistenza tecnica ante e post vendita di supporto tecnico deve lottare con diverse forme di informazione sui prodotti collezionate da centinaia di produttori che non hanno una comune tassonomia. La carenza di standard di informazioni di prodotto rende la corrente aggregazione e diffusione di tali contenuti un'attività costosa e inefficiente, duplicata da ogni distributore presente nel canale. L'adozione di descrizioni di prodotto leggibili dalla macchina risolve questo problema.

Gli standard per descrivere i contenuti di transazioni commerciali sono la prossima frontiera del commercio elettronico. Così come standard quali HTML e TCP/IP¹⁰ hanno portato verso il grande successo di Internet, gli standard di classificazione dei prodotti porteranno ad un altro livello il commercio elettronico.

3.3 Schema di classificazione

Per codifica di un prodotto si intende l'associazione di un codice al prodotto. I codici servono a classificare o identificare un prodotto.

La codifica dei prodotti nei documenti commerciali in formato elettronico (come cataloghi, siti Web, ordini d'acquisto, fatture) permette alle applicazioni informatiche, attraverso la catena logistica (venditori, compratori, distributori, utenti), di processare i dati delle transazioni in modo automatico, migliorando la gestione e l'analisi per il supporto alle decisioni in tempi critici e con metodi efficienti; questo non sarebbe possibile senza le codifiche.

¹⁰ Nelle Information Technology un protocollo è un set di regole che definiscono dei riferimenti nella comunicazione in rete. TCP/IP (Transmission Control Protocol/Internet Protocol) è un protocollo per Internet o per reti private (Intranet o Extranet).

Un buono schema di classificazione prodotti deve essere gerarchico, così che ogni bene di consumo rappresenti un'unica istanza di classi più ampie e famiglie. L'organizzazione gerarchica consente di scendere al livello di specificità che meglio si adatta alle esigenze di una data impresa.

Inoltre, per mantenere una tassonomia gerarchica, uno schema di classificazione va costantemente aggiornato (per aggiungere nuovi prodotti e modificare le strutture esistenti al fine di adattarli alle novità del mercato) e deve essere reattivo alle sollecitazioni delle imprese; l'assegnazione dei codici ai prodotti e servizi non può risultare influenzabile per prevenire pubblicità indesiderate ai prodotti di un'organizzazione a scapito delle altre.

3.4 Le differenze tra classificazione e identificazione

I codici di identificazione sono diversi da quelli di classificazione e non servono ad individuare i prodotti o ad analizzare la spesa o la conoscenza del prodotto.

Codice	Oggetto identificato
Numero di previdenza sociale	Persona
UCC\EAN(codici a barre)	Prodotti
Duns number	Compagnia

Tabella 3 Codici comuni di identificazione

1. **I codici di identificazione** sono usati per rendere non ambigua l'identificazione di un oggetto. La corrispondenza uno a uno tra i codici e gli oggetti è molto utile per registrare e collegare record di oggetti e azioni fatte sull'oggetto.
2. **I codici di classificazione** sono usati per raggruppare cose simili in categorie comuni. Con la classificazione, beni simili sono membri di una classe. Classi simili sono membri di una classe ancora più generica o di una famiglia. La relazione tra i beni e la relazione tra un bene e la sua classe sono informazioni necessarie per l'individuazione degli oggetti, per l'analisi dei consumi e la consapevolezza del

prodotto. In altre parole, i codici di classificazione sono necessari per cercare e trovare appropriati prodotti e servizi, per identificare dove si sono fatte spese e per promuovere i prodotti con reali prospettive di vendita.

Differenza tra classificazione e identificazione. Per illustrare la differenza tra i codici di identificazione e i codici di classificazione, si considera un codice a barre che individua un particolare case di un computer. Questo codice identificativo non può essere riunito in una categoria più generica: come accessori per il computer, non vi sono legami logici con i codici che identificano le varie marche di case per il computer. I codici identificativi non permettono la categorizzazione degli acquisti in categorie più generiche e non permettono il confronto tra diversi produttori.

	Codice di classificazione	Codice di identificazione
Principio cognitivo	Indica la relazione tra oggetti, simili e non	Identifica gli oggetti senza ambiguità
Caratteristica chiave del codice	Gerarchia	Unicità
Cifre del codice	Mostrano le classi e le sottoclassi con i loro membri	Creano corrispondenza uno a uno tra i simboli e gli oggetti (i digit non hanno altro significato)
Business function	Trovare beni e servizi. Analisi delle attività per ulteriore miglioramento	Tracciare e conservare le registrazioni.

Tabella 4 Differenze tra codici di identificazione e di classificazione

Solo una tassonomia gerarchica può permettere l'aggregazione o la specializzazione e l'analisi di comparabilità e la valutazione. Una tassonomia organizza tutti i beni e i servizi in categorie logiche. Un aspetto critico della tassonomia è la presenza di molte categorie intermedie tra gli oggetti specifici e le classi più generiche.

3.5 Convenzioni per la denominazione dei prodotti

La maggior parte dei sistemi commerciali di classificazione dei prodotti si dividono in quelli che classificano i prodotti e servizi e quelli che classificano i processi che

producono i prodotti. Quando l'obiettivo è trovare prodotti, analizzare le spese o aumentare la visibilità del prodotto sul mercato, lo schema di classificazione deve aderire al prodotto. Un'orientazione al prodotto organizza i beni in accordo a come sono usati in input.

Uno schema di classificazione di beni e servizi orientato ai prodotti deve presentare caratteristiche che permettano all'utente di trarre vantaggio dal suo utilizzo. Le caratteristiche chiave vengono elencate in seguito.

La classificazione gerarchica permette di scendere lungo la scala gerarchica (**tecnica Drill Down**) attraverso una vasta gamma di prodotti fino ad arrivare esattamente a quello che si cerca, ma permette anche di salire nella gerarchia (**tecnica Roll Up**), riunendo un gran numero di record di acquisto in un unico contenitore, e di sviluppare analisi relative alla situazione aziendale. Dato che è gerarchica, una tassonomia universale di prodotti e servizi può essere utile per ogni tipo di organizzazione.

Per esempio, analizzando le spese, un compratore di un'azienda manifatturiera troverà la categoria "forniture industriali" troppo generica. Vorrà spezzare in categorie più specifiche: "detersivi" e "solventi". Per un compratore, un publisher o una società di servizio finanziario, la categoria "forniture industriali" non sarà ugualmente dettagliata. Il nome "forniture industriali" può essere specifico per un produttore, ma non esserlo abbastanza per il publisher o per la società di servizi finanziari. In altre parole il livello di specificità dipende dal tipo di impresa e dai suoi obiettivi. Una gerarchia di classificazione assicura che un'impresa possa trovare un significativo livello di analisi che le permetterà di agire nel modo migliore.

Dato che lo schema di classificazione è gerarchico, un'impresa può organizzare un intero gruppo di transazioni di acquisto per mezzo di categorie che le permettono di agire in modo strategico. Invece di concentrarsi sui particolari di una transazione per un singolo prodotto, può aggregare le transazioni ad un livello che identifica un gruppo significativo, come le forniture per l'ufficio.

Un gruppo è una famiglia di prodotti a cui l'azienda acquirente può avvicinarsi per trovare un singolo punto di approvvigionamento. Dato che con un solo contratto si può coprire la famiglia, l'impresa può ottenere un trattamento preferenziale.

Generalmente l'azienda vuole identificatori esattamente al livello di specificità che le permetterà di individuare un tipo di attività conveniente. L'azione può essere il consolidamento di un fornitore, la negoziazione di una fornitura per un determinato volume di acquisto, o anche la sostituzione o l'eliminazione di un oggetto.

In aggiunta alla gerarchia, ogni specifico nome della classificazione dovrebbe avere associato un **codice unico**. Ciò assicura che non ci siano ambiguità riguardo al prodotto cui si riferisce il nome. Questo è importante quando i sistemi di classificazione servono come codice di riferimento per traduzioni multi-linguaggio di informazioni sui prodotti e sui servizi. In questo modo è possibile tradurre il codice nelle tassonomie del cliente concepite per scopi personali.

Per esempio, una multinazionale che vuole analizzare le spese per area e per divisione, se usa nomi di prodotti propri di ogni stato, darà luogo a una proliferazione di categorie che non possono essere ricondotte facilmente a gruppi comuni; usando un singolo codice e traducendolo si risolve il problema.

Anche l'uso di un solo numero di identificazione per un bene permette all'utente di spostare i beni nella gerarchia senza la perdita del potere identificativo. Questo è utile quando una compagnia vuole modificare la tassonomia per adattarla agli schemi di classificazione interni.

Infine, i nomi e le organizzazioni di prodotti individuali e di servizi nei gruppi più alti e più generici cambieranno nel tempo. Ci sono tanti componenti elettronici che 40 anni fa non esistevano. Un numero unico permette alle industrie di spostare in avanti lo schema di classificazione nel tempo, mantenendo traccia del modo in cui oggetti si sono trasformati in nuovi oggetti. In effetti, un numero unico permette riferimenti incrociati che assicurano la consistenza.

Uno schema di classificazione deve essere uno **schema consistente**: la consistenza significa che un singolo oggetto viene identificato in un unico luogo; i vari prodotti possono essere riuniti in un gruppo logico di livello superiore. Una gerarchia consistente permette l'aggregazione o la disaggregazione fino al livello appropriato, senza perdite di precisione.

Un buono schema che identifica tutti i prodotti e i servizi in vendita e mette tutti questi oggetti in una gerarchia può essere considerato uno **schema completo**. Gli utilizzatori di una classificazione richiedono che tutte le categorie rilevanti e tutti i fornitori vengano rappresentati accuratamente dallo schema di classificazione; in questo modo il management può fare analisi e ottimizzare.

Nuovi prodotti e servizi nascono continuamente e altri diventano obsoleti. Il sistema di classificazione deve avere un corpo amministrativo che risponda tempestivamente e senza costi ai cambiamenti del mercato. La gestione dello schema deve essere sensibile agli eMarketplace per aggiornamenti continui.

Agenzie esterne sono i migliori candidati per gestire lo schema. Questo assicura accuratezza, consistenza e integrità. Una terza parte vede l'intera categoria di prodotti e la famiglia in modo uniforme in modo che, per i prodotti di una data società, i prodotti vengano assegnati in modo accurato e consistente.

Dato che una terza parte non ha interesse nel differenziare i prodotti, fornirà maggiore integrità nel classificare i prodotti di un produttore.

Nei paragrafi seguenti si tratteranno i tre standard di classificazione presi in considerazione nella fase sperimentale della tesi.

3.6 Ecl@ss

L'iniziativa di ecl@ss¹¹ è nata nel 1997 come sottogruppo del gruppo di lavoro VCI Material Management of Technical Goods [Ecl@ss, 00]. Il sottogruppo fu chiamato Material-Merkmalleisten. VCI ha scoperto che lo scambio di informazioni tra gli addetti alla gestione dello stabilimento e gli addetti alla gestione dei materiali era lento e pieno di errori a causa del diverso modo di descrivere gli stessi prodotti usato dai due reparti. Così si è deciso di creare uno standard per la descrizione dei componenti tipici degli stabilimenti chimici.

L'idea era definire set standard di attributi per certi tipi di attrezzature che sarebbero stati usati nelle documentazioni dello stabilimento come dati principali. Con tali standard sarebbe stato facile trovare parti di ricambio o valutare le loro condizioni d'acquisto.

Il sottogruppo ha notato che creare set di standard di attributi significava implicitamente creare classi di beni tecnici: tutti i beni che erano appropriatamente descritti da un set d'attributi formavano una classe. Tuttavia non sono stati approfonditi gli aspetti della classificazione, ma si sono coinvolti due imprenditori: uno di una società d'ingegneria chimica e uno di una società di ingegnerizzazione di controllo del processo, calcolando che il processo avrebbe richiesto due anni.

¹¹ www.eclass.de

3.6.1 L'estensione degli obiettivi del progetto ecl@ss

Nel 1998 gli obiettivi del progetto ecl@ss sono stati estesi. I reparti d'approvvigionamento delle società dei membri del VCI si sono avvicinati per due ragioni al progetto del set standard d'attributi. La prima era lo sviluppo di cataloghi elettronici, la seconda la necessità di avere gruppi di materiali unificati per le multinazionali.

Le richieste del catalogo erano simili al caso di gestione dei materiali. Un catalogo elettronico, infatti, è un'applicazione centrata attorno ad un'ampia lista di elementi e alle loro descrizioni. I cataloghi elettronici usano uno schema di categorizzazione gerarchico o un modello *noun-modifier*¹² per disporre i dati in modo gestibile. I più avanzati forniscono una funzione che permette all'utilizzatore di scegliere il prodotto in base alle sue caratteristiche. Entrambi gli aspetti di classificazione e del set d'attributi giocano un ruolo chiave.

Mentre le applicazioni dei cataloghi elettronici e della gestione dei materiali erano simili al progetto iniziale, la gamma dei prodotti nei cataloghi era molto più ampia di quanto non lo fosse lo scopo originale del progetto del set standard d'attributi.

Quest'ultimo era limitato a materiali per gli stabilimenti chimici, includendo approvvigionamenti per l'ufficio e i computer.

L'altro aspetto da prendere in considerazione erano i gruppi di materiali, usati dal reparto d'approvvigionamento per modellare la struttura del mercato. La principale applicazione dei gruppi di materiali era l'analisi del volume d'approvvigionamento della compagnia in base al segmento di mercato: per effettuare il benchmark con imprese multinazionali occorre avere un sistema di raggruppamento dei materiali accettato in campo internazionale.

Si è deciso quindi di ampliare gli obiettivi del progetto in due direzioni. La prima direzione era definire il range dei materiali considerati, che è stato esteso a tutta la gamma di materiali che un'industria chimica avrebbe potuto acquistare. La seconda direzione era la reperibilità dei risultati del progetto. Si è costituita una gerarchia di gruppi di materiali al top delle classi corrispondenti al set standard di attributi. Il gruppo VEBA, un gruppo molto differenziato e multinazionale, al quale appartenevano anche alcuni membri del VCI, disponeva già di un sistema simile che stava usando da un paio

¹² Un modello *noun-modifier* consente di aggiornare e modificare i nomi presenti su un documento.

d'anni internamente. Questo sistema è stato scelto come punto di partenza per gli ulteriori progetti.

Nell'Agosto 1999 è stato presentato un primo risultato al pubblico. Era stata preparata una gerarchia a tre livelli, che comprendeva beni tecnici e servizi. Il quarto livello che ospita le classi relative al set standard d'attributi era solo scarsamente popolato. L'obiettivo di tale pubblicazione era iniziare un'ampia discussione per guadagnare visibilità e ottenere correzioni, ampliamenti da varie fonti. La versione non era da considerarsi pronta all'uso.

L'industria chimica tedesca ha dato origine al processo di creazione di ecl@ss. Ben presto si è deciso di andare oltre la direzione geografica e di estendere il codice a tutti i segmenti. I membri del gruppo di lavoro appartenevano a imprese multinazionali, sapevano fin dall'inizio che uno standard locale non poteva funzionare, ma che i beni e i servizi che stavano descrivendo sarebbero stati venduti a tante altre industrie. L'industria chimica aveva fornitori comuni con queste altre industrie e questi fornitori non potevano gestire numerosi standard specifici di segmento.

L'internazionalizzazione diventa così un'attività centrale di sviluppo. I primi passi sono stati fatti in America e in Asia; l'unione Europea segue il progetto con interesse. La versione 2.0 è bilingue, l'originale è in tedesco ed è accompagnata da una traduzione in inglese. In futuro potranno servire altre traduzioni. Il gruppo di ecl@ss si è aperto ad altre industrie. Attualmente molti dei membri originali hanno attività in altri settori industriali; l'estensione a segmenti diversi diventa naturale ma non è ritenuta ancora sufficiente.

3.6.2 Funzioni e applicazioni

Ecl@ss supporta e integra le seguenti funzioni e applicazioni:

- Ingegnerizzazione (sistemi CAE)¹³
- Gestione dei magazzini
- Gestione dello stabilimento
- Approvvigionamento:
 - Richiesta di offerte;

¹³ CAE (Computer-Aided Engineering) indica la progettazione e analisi di prodotti e processi assistita dal calcolatore.

- Ordini;
- Cataloghi elettronici;
- Analisi.
- Vendite:
 - Informazioni sui prodotti;
 - Offerte;
 - Processamento degli ordini;
 - Cataloghi elettronici.
- Mercati virtuali.

3.6.3 La struttura

Ecl@ss si compone di tre elementi interconnessi: la gerarchia delle classi dei materiali, il set standard di attributi e il sistema di parole chiave.

La gerarchia delle classi di materiali

La gerarchia delle classi dei materiali è un albero a quattro livelli. È la struttura di supporto alla quale sono etichettate parole chiave e set di attributi. I nodi dell'albero sono collettivamente chiamati classi di materiali. Le classi di materiali a livelli diversi hanno nomi diversi (Tabella 5):

Livello	Nome
1	Segmento
2	Gruppo principale
3	Gruppo
4	Classe di beni

Tabella 5 Gerarchia di ecl@ss

A ogni classe di materiali sono associati i campi caratteristici in Tabella 6:

Campi associati a ogni classe di materiali	
Chiave identificatrice	È la primary key che identifica unicamente una classe di materiali
Codice della classe di materiali	È un codice di otto cifre che descrive la posizione della corrispondente classe di materiali nella gerarchia. I segmenti hanno codici di due cifre relativi alla radice, i gruppi principali hanno un codice a due cifre relativo al segmento dei loro genitori e un codice a quattro cifre relativo alla radice, e così via. I codici della classe di materiali sono riempiti con degli zero se necessario.

Nome identificativo	È un termine collettivo che descrive i beni e i servizi appartenenti alla corrispondente classe di materiali. Un nome che identifica in maniera univoca la classe di materiali in una data versione di ecl@ss. Ciò implica che sono richiesti ulteriori qualificatori se un termine è usato con differenti significati in diversi contesti.
Nome breve	È una variante del nome identificativo la cui lunghezza è limitata a 40 caratteri. Se il nome identificativo non supera i 40 caratteri sarà utilizzato anche come nome breve. Un nome breve identifica unicamente la classe dei materiali in una data versione di ecl@ss.
Definizione	Una precisa descrizione testuale dei beni e dei servizi che appartengono alla corrispondente classe dei materiali.
Versione di ecl@ss	È la versione di ecl@ss per cui questi valori e le caratteristiche dei vari campi sono valide. Le chiavi identificative non saranno mai cambiate nelle successive versioni. Tutti gli altri campi possono cambiare. Le chiavi identificative potrebbero scomparire nelle versioni successive se venisse rimossa la corrispondente chiave di identificazione dei materiali.

Tabella 6 Proprietà delle Classi di materiali in ecl@ss

L'uso del set d'attributi

Ecl@ss fornisce set di attributi utili nello sviluppo e nel processamento di transazioni elettroniche. I set di attributi sono preparati per molti materiali e gruppi di materiali e sono archiviati e mantenuti in un database. L'obiettivo è legare un set d'attributi ad ogni punto finale della classificazione a otto cifre. Questa si può vedere come un'aggregazione di tutte le caratteristiche individuali che descrivono materiali o gruppi di materiali rilevanti.

Non appena un set di attributi è stato realizzato e controllato è disponibile nel sistema di classificazione di ecl@ss. Gli utilizzatori saranno forniti di un set completo di attributi collegati a un nome identificativo. Questo, definito come database di presentazione per i set d'attributi, sarà mantenuto invariato per un lungo periodo, probabilmente sei mesi. È ovvio che il set di attributi sarà sviluppato e dovrà essere costantemente ampliato, il Cologne Institute for Business Research mantiene un secondo database per le proposte di sviluppo, che dopo un certo intervallo di tempo viene riconciliato con il database di presentazione. Le parole chiave vengono aggiornate ogni mese e, se richieste, se ne aggiungono nuove. Per garantire una struttura descrittiva universalmente applicabile si devono usare regole accettate internazionalmente. Ecl@ss è basato su questi standard internazionali e quindi fornisce una piattaforma standard. Le regole sono riferite alla normativa ISO 13584, DIN EN

61360 e IEC 65B/349/CD. Questi standard sono interamente concentrati sulla descrizione di dati strutturati

Sistema di parole chiave

Nomi identificativi e nomi brevi sono parole chiave standard per la corrispondente classe di materiali. Una parola chiave addizionale può essere legata alle classi di materiali. Le parole chiave standard non hanno bisogno di essere definite.

A ogni parola chiave sono associati i campi caratteristici in Tabella 7:

Campi caratteristici associati alle parole chiave	
Chiave identificatrice	È la primary key che identifica unicamente una parola chiave.
Testo chiave	È il valore testuale della parola chiave. Per le parole chiave standard il testo chiave è identico al nome identificativo o al nome breve rispettivamente. Un testo chiave identifica univocamente una parola chiave in una data versione di ecl@ss.
Chiave della classe dei materiali	È la chiave identificativa della classe di materiali alla quale è attaccata la parola chiave. Una parola chiave è legata esattamente a una classe di materiali, ma una classe di materiali può essere legata a più parole chiave.
Codice della classe dei materiali	È il codice della classe di materiali alla quale la parola chiave è legata.
Versione di ecl@ss	È la versione di ecl@ss in cui questi valori dei campi caratteristici sono validi. Le chiavi identificative non saranno mai cambiate nelle successive versioni di ecl@ss. Tutti gli altri campi possono cambiare. Le chiavi identificanti possono scomparire nelle nuove versioni se la parola chiave corrispondente è rimossa

Tabella 7 Proprietà delle Parole chiave in ecl@ss

Le parole chiave sono specifiche di un certo linguaggio. Mentre può avere senso tradurre alcune parole chiave, come, per esempio, quelle che denotano un subset di una classe di materiali, è impossibile tradurre sinonimi, espressioni locali o dialettali.

Ecl@ss si compone di quattro livelli di classificazione gerarchica dei materiali con un indice di parole chiave contenente 13000 termini. I livelli gerarchici sono: segmento, gruppo principale, gruppo, classe di beni. Sono disponibili due posizioni per ognuno dei quattro livelli. In ogni livello possono essere incluse 99 classi. Questo sistema genera una struttura per classificare tutti i materiali e i servizi con un grado di dettaglio che soddisfa i requisiti delle aziende. Ecl@ss può essere applicato lungo l'intera supply

chain: la struttura del mercato di approvvigionamento è mappata con un elevato livello di aggregazione, facilitando l'aggregazione di dati relativi al potere di mercato sugli acquirenti. A basso livello la struttura è caratterizzata da legami tecnici di supporto nelle operazioni di mantenimento.

L'ampio indice di parole chiave permette di trovare le classi senza che l'utente conosca dettagliatamente la gerarchia: ciò assicura che ecl@ss possa essere usato come standard di comunicazione attraverso diversi settori e compagnie.

Le caratteristiche e i set di attributi di molti differenti gruppi di materiali sono preparati e conservati in database con riferimento al loro numero in ecl@ss.

L'obiettivo è allegare un set d'attributi ad ogni punto finale della classificazione. Il set d'attributi è un'aggregazione di caratteristiche individuali che descrivono i beni collegati. Questa procedura mira a ridurre al minimo i problemi di comunicazione.

La classificazione è strutturata in 22 segmenti. Ognuno di questi segmenti individuali può avere quattro livelli, ognuno dei quali può avere due posizioni, per mezzo delle quali si possono aggiungere delle posizioni addizionali. Il raffinamento della classificazione fino a otto posizioni fa sì che si possano cercare anche i singoli componenti del prodotto (come ad esempio i materiali). Molti codici dei livelli di classificazione hanno un set d'attributi collegato, che può essere ridefinito a vari livelli in relazione al prodotto o al servizio. Queste caratteristiche fanno sì che dopo che il prodotto è stato trovato, la descrizione elettronica della fornitura e della richiesta possano essere portate a termine con le stesse modalità. Ciò è necessario affinché l'e-commerce possa funzionare.

3.6.4 Le possibilità di ricerca con ecl@ss

Ricerca di parole chiave

È possibile iniziare una ricerca inserendo dei termini scelti dall'utente. Si può dire che le parole chiave dovrebbero essere "singoli nomi nel linguaggio locale". Quando s'immette una parola chiave, tutti i numeri di classificazione che sono legati a questa parola sono automaticamente individuati. È possibile che diversi numeri di classificazione siano legati alla stessa parola chiave.

Ricerca gerarchica

Quando si conduce una ricerca gerarchica, inizialmente si visualizza il primo livello.

Dopo il ventunesimo livello di attributi l'utente può continuare a cliccare per raggiungere il prossimo livello più raffinato.

Numero di classificazione

Quando un commerciante o un compratore entra a conoscenza di ecl@ss, utilizzerà progressivamente i numeri di classificazione che conosce. È infatti possibile immettere il numero direttamente per controllare le rettifiche o per controllare il numero stesso. In questo caso, i prodotti o i servizi allocati a questo numero di classificazione vengono visualizzati al rispettivo livello.

3.6.5 Aggiornamento e mantenimento

Nella sua versione esistente ecl@ss costituisce una cornice per ulteriori sviluppi ed espansioni, come le modifiche in corso del sistema di classificazione dei gruppi di materiali. Il consolidamento e gli ulteriori sviluppi di ecl@ss, come la definizione di nuovi set di attributi e l'allocazione di nuove caratteristiche dei materiali, sono stati portati avanti dalle stesse imprese che hanno provveduto allo sviluppo di ecl@ss. Un gruppo di consulenti specializzati offrirà alle compagnie interessate e alle associazioni la possibilità di cooperare nell'ampliamento di ecl@ss.

Un sistema di classificazione sopravvive in virtù del fatto che viene costantemente modificato per rispondere ai cambiamenti che avvengono nelle varie aree di prodotto. Chiunque usi ecl@ss ha il diritto di fare proposte di aggiunte ed estensioni del sistema di classificazione.

Quando arriva una proposta questa viene esaminata da un gruppo di persone che appartengono a differenti società coinvolte nel processo di sviluppo del sistema su base rotazionale e sono generalmente in grado di prendere decisioni in breve tempo. L'estensione o il nuovo numero viene integrato nel sistema ecl@ss e la persona che ha fatto la proposta viene informata del risultato via e-mail. Questo processo dinamico, dà la possibilità ad ogni compagnia che desideri influenzare il sistema di classificazione di farlo. È possibile che si formino alleanze di compagnie che daranno vita a nuovi main group, allocando loro parole chiave e rendendoli disponibili al Cologne Institute of Research. Questo può ad esempio accadere nell'industria dei servizi.

Ecl@ss non è orientato all'industria, ma è maggiormente progettato per mappare il mercato, l'opportunità di cooperazione nello sviluppo del codice è aperta alle industrie. In ogni caso si deve considerare che ha poco senso creare un sistema di classificazione orientato alle industrie. I risultati sarebbero ripetuti durante il processo di ricerca, con conseguenti codici sovrapposti.

L'utente può influenzare ecl@ss. In primo luogo lo può usare senza costi per svolgere le sue attività; inoltre, se lo ritiene utile, può impostare query via Internet, apportare modifiche o cambiamenti.

3.6.6 Prospettive future di ecl@ss

In cooperazione con varie imprese, il Cologne Institute for Business Research continua a supportare ecl@ss e cerca di introdurre una struttura di notazione per i codici di ecl@ss su Internet.

È stato pianificato che gli utilizzatori del sistema di qualità di ecl@ss che seguono le raccomandazioni del Cologne Institute for Business Research ricevano un sigillo di qualità che permetta loro di distinguersi dai concorrenti. Si assume che la classificazione venga usata nelle aziende e che la competenza in e-commerce sia implementata nella società, in particolare in aziende di piccole o medie dimensioni. A questo fine, il Cologne Institute for Business Research fornisce una lista delle raccomandazioni per le aziende di piccole dimensioni su come devono essere utilizzati i codici di ecl@ss.

3.6.7 Cooperazioni

Il progetto di ecl@ss prevede una serie di cooperazioni, che aumentano con lo sviluppo dello standard. Tra queste, due sono particolarmente autorevoli e vengono descritte di seguito.

BMEcat

Mentre ecl@ss aiuta compratori e venditori a trovarsi nel mercato elettronico, BMEcat

viene usato per standardizzare le transazioni elettroniche tra i partner. BMEcat è un nuovo standard per la digitalizzazione dei cataloghi di prodotto. Molte imprese hanno già cominciato a adottare gli standard e ora insistono affinché i loro fornitori rendano disponibili i cataloghi su Internet e la lista dei prodotti nel formato BMEcat.

SAP

SAP è stato coinvolto nello sviluppo del sistema ecl@ss. Nell'autunno 1999 alcuni gruppi di utilizzatori di SAP negli Stati Uniti e nel Giappone furono contattati e fu istituito un gruppo di progetto che mise gli utilizzatori di SAP nella condizione di incorporare ecl@ss come sistema di classificazione in SAP.

3.6.8 Qualità e manutenzione dei codici

Il mantenimento è di lungo periodo. In questa fase ecl@ss è ancora in costruzione. Ciò implica che cambierà abbastanza rapidamente fino a quando non verrà considerato in uno stato adatto all'utilizzo. Fino a quel momento sarà un sistema vivo. Saranno sviluppate nuove tecnologie e quelle vecchie perderanno di significato. Prodotti e servizi appariranno e spariranno. I mercati verranno modificati seguendo nuove strategie o modelli di business. Conseguentemente ecl@ss necessita di un'organizzazione che sia in grado di provvedere al suo mantenimento di lungo periodo. Tale organizzazione dovrebbe essere un'entità legale per proteggere copyright e brevetti. Dovrebbe essere accreditata universalmente. Si dovrebbe garantire il suo finanziamento per il lungo periodo. Il passaggio da struttura in fase di costruzione a struttura pronta all'uso non riguarderà l'intera struttura di ecl@ss. Certi sottoalberi della gerarchia insieme con il corrispondente set di attributi saranno considerati pronti all'uso. Quando una nuova industria si unisce alla comunità di ecl@ss si cominciano a considerare nuovi mercati di approvvigionamento; ci sarà, quindi, un nuovo sottoalbero in stato di costruzione.

Le imprese che adotteranno ecl@ss lo porranno tra i loro dati principali. Il mantenimento dei dati principali nelle organizzazioni complesse richiede un'appropriata gestione del sistema qualità. Alcuni settori industriali, come ad esempio l'industria farmaceutica, per motivi legali, richiedono la dichiarazione di qualità obbligatoria.

La gestione del mantenimento di ecl@ss deve essere documentata in un documento relativo alla qualità, che regola i seguenti temi:

- cambiamenti nel manuale della qualità;
- membri;
- persone e competenze;
- elezione di membri;
- presentazione di richieste di cambiamento;
- decisione di procedure e criteri;
- pubblicazione di nuove versioni;
- obiezioni;
- frequenza di cambiamento;
- criteri per la qualità tecnica di ecl@ss;
- linee guida per assicurare la qualità tecnica di ecl@ss.

3.7 UNSPSC (UNiversal Standard Products and Services Classification)

UNSPSC è nato inizialmente come uno standard *open source*¹⁴ progettato per l'analisi di spesa e per l'individuazione di risorse. I fornitori aggiungono UNSPSC nei loro cataloghi e fatture per descrivere i loro prodotti e servizi. L'inserimento di UNSPSC rende più facile per gli acquirenti trovare prodotti e analizzare cosa hanno comprato tra i numerosi venditori.

UNSPSC è progettato per includere tutti i prodotti e servizi che sono comprabili o vendibili. Un prodotto/servizio per essere incluso in UNSPSC deve fare parte di un gruppo di prodotti o servizi sostituibili. UNSPSC è uno schema di classificazione di commodity, non uno schema di classificazione di prodotti; include solo gli attributi primari usati per differenziare una commodity da un'altra. Attributi come marchi o

¹⁴ Per linguaggi di tipo "open source" ci si riferisce ai linguaggi i cui codici generatori sono messi a disposizione sul Web, per favorire la diffusione e coinvolgere ricercatori e altri utenti esperti nel miglioramento dei codici stessi.

caratteristiche fisiche sono normalmente esclusi, ma, tra gli obiettivi futuri, vi è quello di costruire un framework di standardizzazione che classifichi gli attributi, in modo che possano essere facilmente utilizzati per descrivere meglio un servizio o prodotto specifico offerto nel marketplace dell'e-commerce.

UNSPSC è stato gestito fino all'ottobre 2002 dalla Electronic Commerce Code Management Association (ECCMA)¹⁵; pertanto nei paragrafi successivi sarà presentata e discussa la struttura creata da ECCMA per l'aggiornamento, diffusione e controllo dello standard, poiché l'attuale codifica è frutto di questo schema di progettazione e lavoro; tuttavia è bene precisare che il nuovo gestore e distributore del codice è il consorzio UNDP¹⁶ (United Nations Development Programme).

3.7.1 Classificazione gerarchica

Il primo obiettivo della classificazione gerarchica UNSPSC è evitare le duplicazioni e fornire uno strumento che consenta, a chi è responsabile della classificazione dei prodotti o servizi, di identificare facilmente il codice appropriato senza la tabella UNSPSC.

La caratteristica più importante dello schema UNSPSC è che una commodity deve essere individuata da titolo e definizione unici e ogni titolo e ogni definizione devono essere facilmente differenziabili da tutti gli altri.

Nelle situazioni in cui ci sono molti utenti o molte sorgenti della stessa commodity la classificazione riflette l'utilizzo dominante della commodity all'interno del marketplace globale.

L'utilizzo principale varia nel tempo e da Paese o Regione e si compie ogni sforzo per riflettere l'uso internazionale di una commodity, al contrario delle applicazioni regionali o specifiche per un'industria. UNSPSC è una classificazione gerarchica a 4 livelli; la rappresentazione raccomandata del ECCMA prevede valori numerici a 2 digit separati in periodi: NN.NN.NN.NN.

¹⁵ www.eccma.org

¹⁶ www.un-spsc.net

Livello	Definizione
Segment	Aggregazione logica di famiglie per scopi analitici Es: 56.00.00.00 Furniture and Furnishings
Family	Gruppo di categorie di commodity correlate riconosciute comunemente Es: 56.10.00.00 Accommodation furniture
Class	Gruppo di commodity che corrispondono ad un uso o funzione comune Es: 56.10.17.00 Office furniture
Commodity	Gruppo di prodotti/servizi succedanei Es: 56.10.17.03 Desks

Tabella 8 Descrizione livelli gerarchici in UNSPSC

Il Global Commodity Identifier (EGCI) di ECCMA fornisce alla tabella UNSPSC una sequenza di identificazione che permette un controllo di versione e il collegamento tra tabelle incrociate. EGCI è collegato a titolo e definizione della classificazione; mentre lo schema di classificazione (EGCC) può cambiare se una commodity viene ri-classificata, EGCI non cambia mai. EGCI è un numero sequenziale, senza significato, assegnato a ogni record nella tabella UNSPSC.

Quando un valore EGCI è richiesto per rappresentare un livello EGCC pienamente espresso, la rappresentazione raccomandata di EGCI è NNNNNN.

La conversione degli esempi in Tabella 8 dovrebbe essere la seguente (Tabella 9):

Titolo	Definizione	EGCC	EGCI
Segment	Furniture and Furnishings	56.00.00.00	010745
Family	Accommodation furniture	56.10.00.00	009227
Class	Office furniture	56.10.17.00	010200
Commodity	Desks	56.10.17.03	008288

Tabella 9 Conversione EGCC - EGCI

Anche per UNSPSC è utile distinguere tra classificazione e identificazione:

Classificazione

- Progettata per assicurare consistenza al codice
- Non progettata come indice per il recupero dei dati
- Le classificazioni possono cambiare (*merge e split*)
- Le classificazioni si possono riutilizzare nel tempo

Identificazione

- Primary key
- Non può mai essere riutilizzata
- Progettata per assicurare indipendenza tra versioni
- Progettata per facilitare l'accesso a tabelle collegate
- Progettare per permettere classificazione personalizzate
- Progettata per essere costruita in applicazioni software

3.7.2 ECCMA: organizzazione distributrice di UNSPSC

ECCMA (Electronic Commerce Code Management Association) è un'organizzazione no profit registrata in Delaware (USA) come Code Management Association. E' stata costituita per gestire il mantenimento e la distribuzione di codici usati nel commercio elettronico. Un gruppo di manager (Board of Directors) gestisce l'Associazione. Su richiesta della comunità di utenti il Board of Directors di ECCMA costituisce dei *code group* che comprendono una o più liste di codici che supportano una funzione correlata all'e-commerce. Ogni code group è controllato da un distinto Procedures Review Committee che attribuisce un numero illimitato di membri volontari al Technical Assessment Group (TAG). Il Gruppo di assistenza tecnica di ogni codice è responsabile di valutare le richieste di aggiornamento della lista di codici gestita dal gruppo. L'appartenenza al TAG è elettiva e concessa a tutti i membri di ECCMA. Per ogni code group vi sono due TAG, uno responsabile delle richieste del settore pubblico, l'altro di quelle del settore commerciale.

Vi è inoltre un gruppo dedicato alle traduzioni dei codici, il Language Translation

Group LTG, che per ogni code group è responsabile della traduzione delle modifiche approvate alla lista di codici gestita dal gruppo. L'appartenenza al LTG è elettiva e concessa a tutti i membri di ECCMA.

I membri di ECCMA possono associarsi in uno o più code group.

ECCMA si è dotata di alcuni principi per la gestione degli standard:

1. I codici gestiti da ECCMA devono essere disponibili gratuitamente al pubblico senza restrizioni di copyright.
2. I codici gestiti da ECCMA sono mantenuti e distribuiti esclusivamente su Internet.
3. ECCMA opera sotto un'architettura definita con regole comprensibili e procedure per il mantenimento e lo sviluppo dei codici. Specificatamente, lo sviluppo e la gestione dei codici devono rimanere un processo a consenso aperto, non deve essere dominato da una singola organizzazione o dall'interesse di una categoria. Ogni organizzazione o persona con un interesse materiale o diretto ha il diritto a partecipare nello sviluppo e nel mantenimento dei codici di ECCMA.
4. ECCMA cerca di collaborare in modo chiuso con altri standard correlati per non avere nessuna duplicazione degli sforzi, conflitti, confusione che possano affliggere la comunità di utenti.
5. ECCMA vuole aderire in modo rigoroso ai principi etici.

I codici sono disponibili liberamente al pubblico senza restrizioni anche se l'appartenenza è richiesta per partecipare nella gestione dei codici.

Ogni individuo o organizzazione può essere membro di ECCMA richiedendo di fare parte di uno o più code group e fornendo un indirizzo di posta elettronica accompagnato dal pagamento di una tessera annuale come predisposto dai Direttori di ECCMA.

Ai membri sono assicurati alcuni privilegi:

- Ai membri sono notificate le variazioni al codice;
- Solo ai membri è consentito proporre modifiche o aggiunte;
- Solo i membri possono chiedere di appartenere al TAG;
- Solo i membri possono appartenere al Language Translation Group (LTG);
- Solo i membri sono eleggibili nel Procedures Review Committee.

Per contro i membri del comitato hanno doveri e responsabilità:

- Approvare i membri del TAG;
- Approvare i membri del LTG;
- Assicurare il rispetto delle Code Management Procedures da parte dei membri;
- Raccomandare cambiamenti alle Code Management Procedures.

Il comitato di ogni gruppo di codice deve essere composto da 12 membri inizialmente appartenenti al Board of Directors di ECCMA e poi eletti dai membri alla fine del loro incarico per un determinato periodo nel quale si possono rassegnare le dimissioni. Il comitato deve rappresentare un minimo di quattro organizzazioni pubbliche e un minimo di quattro organizzazioni commerciali. Il comitato deve eleggere un Presidente e un Vice Presidente.

A livello amministrativo ECCMA è gestito da un processo automatico accessibile attraverso Internet. Comprende un accesso pubblico con il quale distribuire liberamente il codice e fornire informazioni su ECCMA, un accesso riservato ai membri di un gruppo di codice che possono accedere al database del codice e proporre modifiche, un accesso ai membri del TAG con il quale discutere le variazioni al codice, un accesso al LTG per concordare le traduzioni, un accesso ai membri del Comitato che gestiscono le richieste di appartenenza al TAG e al LTG, un accesso ai Direttori di ECCMA sul quale definire i doveri dei membri di ECCMA e mantenere una storia completa dei file con lo scopo di consultarli.

ECCMA è una variante del metodo Delphi, nato nel 1944 come metodo di previsione dello sviluppo tecnologico. ECCMA ha adottato il metodo Delphi come strumento per generare un rapido consenso tra i membri del gruppo che in un primo tempo formulano proposte separate, affinché ogni esperto non sia condizionato da altri, e poi lavorano insieme per confrontare le idee e concordare una soluzione. I partecipanti, per essere motivati, devono essere convinti che ogni membro sia in grado di portare un contributo originale sul problema in questione e vogliono ricevere feedback pari alle conoscenze che loro stessi hanno messo in campo.

Tra i membri di ECCMA, molti sono tra loro concorrenti; la competizione è sia orizzontale sia verticale. Questo significa che ogni attività deve essere valutata rispetto alle leggi Antitrust. ECCMA s'impegna a impedire ogni violazione e ogni azione che possa far crescere il sospetto di una violazione. I partecipanti agli incontri di ECCMA

devono sapere che lo scopo del meeting è l'efficienza della competizione dei concorrenti nell'interesse del consumatore e dell'utente. Tuttavia, visto che spesso l'attività di ECCMA accresce la cooperazione tra concorrenti, occorrerà assicurare attenzione alle leggi antitrust, questo significa che la partecipazione deve essere volontaria e una defezione non deve penalizzare alcuna società, non ci devono essere discussioni su prezzi, boicottaggi, allocazione di consumatori, ogni standard sviluppato deve essere uno standard volontario.

3.7.3 Distribuzione di UNSPSC

ECCMA mantiene e distribuisce tre file UNSPSC. Il primo è progettato come file pubblico ed è stato disponibile al sito di accesso pubblico¹⁷ fino all'ottobre 2002 (vedi sezione [3.7]); contiene solamente EGCC, titolo e definizione; viene aggiornato ogni tre mesi. Ha uno scopo educativo e non implementativo, non include EGCI o informazioni per il controllo versione. Gli altri due file sono disponibili ai membri ECCMA.

I member file sono progettati per l'implementazione e vengono aggiornati il 15 di ogni mese. Includono EGCI e le informazioni per il controllo versione; sono riferiti all'Audit file e al Current file. L'Audit file contiene tutta la storia di UNSPSC, il Current file è un subset dell'Audit file e comprende solo quelle classificazioni che non sono state contrassegnate o cancellate nell'Audit file. ECCMA raccomanda che il Current file sia usato per la scrittura di UNSPSC e l'Audit file per la lettura di UNSPSC. Questa raccomandazione assicura compatibilità a ritroso.

UNSPSC è uno standard open source; è di pubblico dominio e non soggetto a copyright. può essere quindi usato e ri-distribuito liberamente senza licenza. ECCMA gestisce il mantenimento e la distribuzione di UNSPSC nell'interesse dei suoi membri. ECCMA mantiene un sito Web per i soli membri che è utilizzato per richiedere cambiamenti al processo e per tenere i membri aggiornati a un numero di content standard, incluso UNSPSC. L'adesione a ECCMA non è richiesta per ottenere o utilizzare UNSPSC; comunque, l'adesione a ECCMA assicura l'aggiornamento e consente l'accesso a consigli per l'implementazione.

¹⁷ Fino all'ottobre 2002 ECCMA ha gestito e distribuito con il sito ufficiale www.eccma.com lo standard UNSPSC. Da quella data UNSPSC è gestito da UNDP e scaricabile dal sito www.un-spssc.net

3.7.4 Rivisitazione e aggiornamento

UNSPSC è soggetto a regolare rivisitazione e modifica. Ogni membro di ECCMA può proporre un cambiamento. La richiesta è approvata con voto a maggioranza del Technical Advisor Group (TAG) composto da membri volontari che hanno conoscenza e esperienza sul campo. I cambiamenti approvati sono aggiunti alle tabelle disponibili sul sito ECCMA il 15 e l'ultimo giorno del mese. ECCMA notifica ai suoi membri i cambiamenti.

Seguono le regole utilizzate dal TAG nel vagliare le richieste di modifica:

Denominazione convenzioni

Regola 1. Tutte le gerarchie sono esplicite, non ce ne sono di implicite.

- I titoli delle commodity non sono usati come implicazione di gerarchia;
- Le virgole non sono usate nei titoli di commodity (Es. “tennis balls” va bene, “balls, tennis” no);
- L'ordine all'interno di un gruppo non è significativo; non ci sono gerarchie implicite.

Regola 2. I titoli delle commodity sono unici.

- Nei titoli delle commodity, la congiunzione “and” può essere usata, “&” no.

Regola 3. Ai livelli Segment, Family, Class “and” può essere usato, “&” no.

- Al livello Segment tutte le parole chiave hanno l'iniziale maiuscola;
- Ai livelli Family, Class, Commodity solo la prima parola è scritta in maiuscolo.

Linee guida per la classificazione:

Regola 1. Una commodity sarà classificata con altre che servono o supportano una funzione comune, un obiettivo, un compito. Questi Segment, Family, Class e Commodity sono normalmente rintracciabili nello stesso settore di distribuzione o approvvigionamento.

Regola 2. Una commodity sarà classificata con altre attraverso un processo simile e spesso prodotta dalla stessa azienda o aziende simili.

Regola 3. Se le prime due regole non sono applicate, allora il materiale con cui sono prodotte le commodity aiuterà a determinare la sua classificazione

ECCMA non fornisce servizi di tagging per UNSPSC; tuttavia mantiene sul sito una lista delle diverse società membri che forniscono codifiche professionali di UNSPSC e servizi di manutenzione. ECCMA non certifica la qualità della codifica, pertanto è consigliabile che l'azienda utilizzatrice confronti prezzi, servizi, qualità di codifica tra i venditori.

3.8 NAICS: dati nuovi per un'economia nuova

Il Nord America ha un sistema tutto nuovo per classificare la attività di business e elaborare statistiche industriali: NAICS. Esso rimpiazza il sistema precedente (SIC) e identifica molte nuove industrie, prevedendo aggiornamenti regolari. E' nato per supportare le statistiche industriali degli Stati Uniti e per gestire il loro utilizzo.

Dopo 60 anni di servizio, il vecchio sistema (Standard Industry Classification) SIC è stato ritirato ed è stato sostituito con il (North American Industry Classification System) NAICS¹⁸. Le prime statistiche industriali basate su NAICS sono i dati schedulati dall'Economic Census¹⁹ nel 1997 per la pubblicazione del 1999.

NAICS è una revisione del sistema usato per classificare le attività di business delle imprese. Diversamente dalle prime revisioni di SIC, i cambiamenti apportati da NAICS sono fondamentali. Vengono riconosciuti centinaia di nuovi business economici, in primo luogo nei settori dei servizi che sono in forte crescita. Le industrie sono classificate secondo un unico consistente principio economico, piuttosto che con svariati. Tali cambiamenti sono stati sviluppati congiuntamente dagli Stati Uniti, dal Canada e dal Messico, estendendo per la prima volta un principio di comparabilità a tutto il Nord America.

NAICS è un prodotto di deliberazione estesa, ufficiale e pubblica. I piani, la concezione, le proposte di revisione e la proposta completa del sistema sono stati dettagliati in numerose notifiche dei registri federali e discussi dal 1994 al 1996. Dopo aver considerato tutti i commenti pubblici, la notifica che ha reso NAICS effettivo negli Stati Uniti è stata pubblicata nel 1997. Il North American Classification System è stato presentato al pubblico a metà del 1998. La classificazione NAICS viene aggiornata

¹⁸ www.ntis.gov/product/naics.htm

¹⁹ www.census.gov

regolarmente per essere al passo con le variabili condizioni di business e con la richiesta di informazioni. Nell'adottare NAICS, tutti e tre gli stati del nord America si sono accordati per revisionare il sistema e il suo utilizzo ogni 5 anni. Questo impegno testimonia che NAICS verrà cambiato nel momento in cui si presenti la necessità di riconoscere nuovi business, aumentare la comparabilità internazionale, e migliorare l'utilità dei dati.

3.8.1 Classificazione delle nuove industrie

NAICS United States fornisce una classificazione dettagliata di 1170 industrie, con un aumento del 15% rispetto a quelle che si potevano trovare in SIC. Allo stesso tempo sono state rimpiazzate o riviste il 60% delle aziende che già erano disponibili in SIC. Sono comprese, inoltre, 358 nuove industrie che SIC non identificava, 390 che sono state revisionate dalle loro controparti su SIC e 422 che sono rimaste sostanzialmente invariate. Il risultato è una classificazione industriale espansa e rivista che rispecchia i business dell'economia moderna. La Tabella 10 fornisce un esempio delle nuove industrie identificate da NAICS.

Semiconductor machinery manufacturing	Temporary help services
Software reproducing	Telemarketing bureaux
Fiber optic cable manufacturing	Hazardous waste collection
Convenience store	HMO medical centers
Gasoline stations with convenience stores	Continuing care retirement communities
	Casinos
Warehouse clubs and superstores	Casino hotels
Food (health) supplement stores	Bread and breakfast inns
Pet and pet supply stores	Limited services restaurant
Pet care services	Automotive oil change and lubrication shops
Cable networks	Diet and weight reducing centers
Satellite telecommunications	
Paging	
Cellular and other wireless telecommunication	
Telecommunications resellers	
Credit card issuing	

Tabella 10 Le nuove aziende di NAICS

Il nuovo catalogo industriale di NAICS mostra i vari modi in cui l'industria è cambiata. Alcune parti riconoscono gli sviluppi *high-tech* come la produzione di cavi per le fibre ottiche, le telecomunicazioni via cellulare, e la riproduzione di software per il computer. Altre riconoscono nuovi business. Inoltre si sono considerati cambiamenti relativi al mondo dei servizi che supportano le attività economiche, come le strutture di Bed&Breakfast, i centri di telemarketing, le società di emissione di carte di credito. Le 422 industrie che non sono state modificate da NAICS hanno la loro diretta controparte in SIC. In più ci sono 38 nuove aziende di NAICS che possono essere identificate in SIC. Queste sono aziende NAICS che riflettono la precedente suddivisione delle aziende di SIC. Per esempio, i sistemi per lo smaltimento dei rifiuti (SIC 4953) sono stati divisi in 5 nuove aziende di NAICS. Tali aziende possono essere combinate per ricreare quella di partenza su SIC. Comunque altre 710 aziende nuove o riviste da NAICS non trovano corrispondenza in SIC e non sono tracciabili con la precedente classificazione di SIC.

3.8.2 Diversi settori economici coinvolti

NAICS raddoppia il numero di raggruppamenti di alto livello della classificazione industriale. Il più alto livello della classificazione è chiamato "settore" ed è simile a quella che in SIC veniva chiamata "divisione". NAICS raggruppa le aziende in 20 settori, paragonabili con le sole 10 divisioni di SIC. La Tabella 11 elenca ogni divisione di SIC e mostra i settori di NAICS cui corrisponde.

Nome della divisione (SIC)	Nome del settore (NAICS)
Agriculture, forestry and fishing	Agriculture, forestry, fishing and hunting
Mining	Mining
Construction	Construction
Manufacturing	Manufacturing
Transportation, communications and public utilities	Transportation and warehousing
Wholesale trade	Wholesale trade
Retail trade	Retail trade
Finance, insurance and real Estate	Finance and insurance Real Estate and rental leasing

Services	Information Professional, scientific and technical services Administrative and support and waste management and remediation services Educational services Health care and social assistance Arts, entertainment and recreation Other services (except public administration)
Public administration	Public administration
None	Management of companies and enterprises

Tabella 11 Settori interessati dalla classificazione NAICS

I cambiamenti introdotti da NAICS sono molto maggiori di quelli che si potrebbero dedurre dalla somiglianza dei nomi di alcuni settori. Nessun settore di NAICS ha la stessa classificazione dei sotto-settori del predecessore in SIC. Per esempio, il settore costruzioni ha tre sottosectori che sono simili a SIC, ma 3 delle sue 28 industrie descritte sono nuove e 14 sono state riviste. Per capire i cambiamenti in NAICS e le loro implicazioni è necessario considerarli settore per settore.

3.8.3 Concetti rivisti della classificazione

NAICS cambia concetti chiave e definizioni della classificazione rispetto a SIC. Questi cambiamenti tecnici riguardano sostanzialmente il modo di classificare le attività di business e il numero e il tipo di business inclusi in particolari raggruppamenti della classificazione. Alcuni di questi cambiamenti possono essere identificati in nuovi titoli e raggruppamenti della classificazione NAICS. Ce ne sono anche altri che non sono evidenti nella nomenclatura. Esempi di cambiamenti nei concetti e nelle definizioni sono:

- *Classificazione della produzione per processo produttivo*: NAICS classifica ogni business in base al processo produttivo industriale usato. Precedentemente alcuni business erano classificati in base ai loro processi produttivi, ma altri venivano classificati usando altri principi, come la classe di consumatori. Per le industrie che

prima non erano classificate in base al loro processo produttivo, la riclassificazione con NAICS può cambiare sostanzialmente quante e quali attività sono incluse.

- *Ridefinizione di commercio al dettaglio e all'ingrosso.* L'uso di un principio di classificazione basato sulla produzione ha cambiato il confine esistente tra i settori di commercio all'ingrosso e al dettaglio. Queste attività nel sistema SIC erano classificate in base al tipo di clienti serviti. NAICS le classifica in base a come opera ciascun business. I commercianti al dettaglio tipicamente vendono merci in piccola quantità usando metodi orientati al pubblico, come la pubblicità mediante mass-media. I commercianti all'ingrosso vendono grandi quantità di beni usando metodi business-oriented come cataloghi specializzati, contatti con i clienti, data warehousing²⁰, office location. Questa ri-definizione fornisce classificazioni e statistiche più ricche di significato per ogni settore, ma ne apporta dei cambiamenti. Ad esempio più della metà delle stazioni per il carico del petrolio che prima erano classificate come vendite all'ingrosso adesso saranno classificate come rivendite al dettaglio.
- *Ri-classificazione delle attività ausiliarie.* Le attività ausiliarie forniscono servizi al personale, processamento dei dati, trasporto ad altre organizzazioni della stessa società. Le attività ausiliarie venivano classificate in relazione con la compagnia a cui prestavano servizio. NAICS classifica le attività ausiliarie in relazione al servizio che forniscono. Questa nuova classificazione rende disponibili dati più dettagliati sulle attività di servizio, ma cambia anche entrambe le attività coinvolte.

3.8.4 I nuovi sistemi di numerazione

I codici numerici di NAICS sono completamente diversi da quelli di SIC. Il nuovo sistema di numerazione sottolinea l'approccio di NAICS e permette ai cambiamenti di venire pienamente sviluppati. Il nuovo sistema si compone di cinque livelli di classificazione (invece dei 4 di SIC), dettaglia il codice fino ad arrivare ad un massimo

²⁰ Data Warehouse in ambito IT è inteso come sede centrale per le parti significative di dati che i vari sistemi di un'impresa vogliono immagazzinare.

di sei cifre, la classificazione dettagliata viene chiamata “US Industry” (invece di Industry, come veniva chiamata in SIC). Questi cambiamenti richiedono nuovi numeri anche per le aziende che procedono senza cambiare. Ad esempio NAICS e SIC includono entrambi “passenger car rental”. In NAICS il codice è 532111 e in SIC è 7514.

Benché tutti i codici numerici di NAICS siano nuovi, come in SIC, sono organizzati in gruppi gerarchici multi livello di classificazione costruiti a partire dal livello industriale dettagliato. Il livello di classificazione dettagliato e l’aumentato numero di cifre permettono a NAICS di comprendere un maggior numero di classificazioni, garantire flessibilità nello strutturare i gruppi di classificazione industriale e permettono variazioni specifiche di uno stato nella classificazione dei dettagli.

La Tabella 12 mostra la struttura gerarchica di NAICS e i codici numerici per un settore mantenuto e uno nuovo. Il sesto carattere della classificazione NAICS può essere usato in modo diverso negli Stati Uniti, in Canada e in Messico; è per uso opzionale da parte di ogni Paese e riflette differenze economiche e informative. Gli Stati Uniti e il Canada hanno deciso di usare lo stesso sesto carattere per aziende che sono comparabili, in modo da evitare di ottenere codici identici per aziende non comparabili.

	Settore mantenuto		Settore nuovo	
Settore	31-33	Manufacturing	51	Information
Sotto-settore	334	Computer and electronic product manufacturing	513	Broadcasting and telecommunications
Gruppo d’industrie	3346	Manufacturing and reproducing of magnetic and optical media	5133	Telecommunications
Industria	33461	Manufacturing and reproducing of magnetic and optical media	51332	Wireless telecommunication carriers (except Satellite)
Industria US	334611	Reproducing of software	513321	Paging

Tabella 12 Struttura gerarchica di NAICS e codici numerici per un settore

3.8.5 Le prospettive di NAICS

Le agenzie di statistica del Canada, del Messico e degli Stati Uniti si occuperanno dello sviluppo, dell'implementazione e del mantenimento di NAICS. Quando sarà pienamente implementato, le statistiche industriali di tutti e tre gli stati saranno pienamente comparabili ai livelli di dettaglio concordati, sarà possibile un'analisi migliore della performance economica nazionale e collettiva, saranno disponibili nuove informazioni riguardo i flussi commerciali internazionali e il mercato dei business.

Per ottenere i benefici di NAICS, gli utilizzatori delle statistiche industriali dovranno rivedere ed aggiustare le loro sorgenti di dati, i metodi e le analisi. NAICS United States mantiene 422 industrie di SIC senza sostanziali cambiamenti e crea 38 industrie che sono tracciabili alla loro controparte SIC. I dati per queste industrie devono rimanere comparabili e derivabili in entrambi i sistemi di classificazione. Comunque, anche per queste industrie, le abitudini legate ai numeri dei codici di classificazione e alla nomenclatura ufficiale devono essere rivisti.

Dagli utilizzatori dei dati saranno richiesti aggiustamenti addizionali per le 390 industrie che NAICS United States ha ridefinito e per le 320 che sono nuove e non tracciabili. Per queste industrie non ci sono dati direttamente comparabili basati su SIC e NAICS. L'uso di dati per queste industrie richiederà nuovi metodi per stimare dati comparabili e in alcuni casi non sono possibili stime adeguate.

Il manuale di NAICS United States è il più utile per chi vuole capire e adattarsi al nuovo sistema. Questo manuale rende possibile un'analisi dettagliata di NAICS per tutte le nuove classificazioni.

Le informazioni più importanti del manuale sono le tabelle di corrispondenza che dettagliano le relazioni tra NAICS Stati Uniti e SIC, facendo la traduzione tra statistiche basate su NAICS e SIC.

La Tabella 13 mostra le classificazioni di NAICS United States e quelle di SIC o una loro parte che comprende la classificazione. Le classificazioni NAICS United States che non sono comparabili per Stati Uniti, Canada e Messico sono identificate con superscript e quelle SIC che sono solo parzialmente incluse in NAICS United States sono identificate con parentesi e asterischi.

Codice NAICS	Descrizione NAICS	Status code	Codice SIC	Descrizione SIC
5111	Newspaper, periodical, book and database publisher			
51111	Newspaper publisher	E	2711	Newspaper publishing, or publishing and printing
51112	Periodical publisher	R	2721 *2741	Periodicals: publishing or publishing and printing miscellaneous publishing (shopping news)
51113	Book publisher	R	2731 *2741	Books publishing or publishing and printing miscellaneous publishing (Technical manuals and book)
51114	Database and directory publisher	N	*2741 *7331	Miscellaneous publishing (database publishers) direct mail advertising services (mailing list compilers)
51119	Other publisher			
511191	Greeting card publisher-	R	*2771	Greeting cards (publishing greeting cards)
5111919	All other publisher-	R	*2741	Miscellaneous publishing

Tabella 13 NAICS (1997) VS SIC (1987)

Capitolo 4

Integrazione delle Informazioni

Il World Wide Web ha drasticamente cambiato la disponibilità on-line delle informazioni e la quantità delle stesse scambiate elettronicamente. Il Web ha rivoluzionato l'accesso personale alle informazioni e la gestione della conoscenza nelle organizzazioni più ampie. Inoltre ha permesso l'inizio dello scambio di relazioni commerciali on-line nei marketplace Business to Business, come luogo di incontro tra venditori e compratori, ciascuno dei quali può utilizzare il proprio formato per rappresentare i prodotti all'interno del catalogo prodotti.

Un'importante frazione delle transazioni B2B è ancora realizzata sulle reti tradizionali non Internet, come i sistemi EDI. Questi modelli non sfruttano appieno le possibilità del commercio elettronico e saranno presto rimpiazzati da tipi di transazione basati sul Web, che assicurano un alto livello di flessibilità e chiarezza, necessarie per ottimizzare le relazioni di business.

Gli eMarketplace sono un middleware, che aiuta i clienti a contattare un numero potenzialmente alto di fornitori senza incorrere nel problema di implementare svariati canali di comunicazione. Un marketplace di successo deve integrare varie piattaforme software e hardware e fornire un protocollo comune per lo scambio di informazioni.

Tuttavia il problema dell'eterogeneità degli standard di descrizione per prodotti, cataloghi e documenti si pone come ostacolo al loro sviluppo e può essere risolto studiando opportuni sistemi per l'integrazione delle sorgenti di dati [Omelayenko, 01b].

I capitoli successivi proporranno una soluzione relativa all'eterogeneità degli standard di descrizione dei prodotti, studiando un progetto di integrazione tra content standard nell'ambito del progetto MOMIS, un sistema per l'integrazione intelligente delle informazioni, sviluppato con la collaborazione tra l'Università di Modena e Reggio Emilia e l'Università di Milano.

4.1 Tre livelli di eterogeneità dei dati

La gestione efficiente dei diversi stili di descrizione degli elementi che partecipano a una transazione B2B è un'attività chiave dei Marketplace. Tra le descrizioni eterogenee vi sono seri problemi di integrazione che si sviluppano in tre direzioni [Ding, 02]:

1. Diverse ontologie di prodotto (**content standard**), definite nella sezione [2.3];
2. Diversi standard per descrivere la struttura dei **cataloghi di prodotto**, collegati ai content standard;
3. Diversi standard per descrivere i **documenti di business** (document business), quali ordini e fatture.

Il contenuto delle informazioni scambiate deve essere modellizzato. Storicamente sono stati sviluppati diversi modi di descrivere e categorizzare i prodotti; spesso ogni venditore ha un proprio modo per indicare un prodotto. La strutturazione e la standardizzazione della descrizione dei prodotti è un compito significativo nel commercio elettronico B2B, che assicura ai differenti attori di comunicare con gli altri in modo che il loro cliente possa individuare il prodotto che lo interessa. In questo senso i provider di soluzioni per il Content Management (definito nella sezione [2.3.1]) possono offrire valore aggiunto nell'aiutare i venditori a costruire un'ontologia per un determinato dominio di prodotto.

L'e-Commerce si occupa dello scambio elettronico di informazioni B2B di cui le descrizioni dei prodotti sono solo un elemento. Le descrizioni di prodotto sono le unità elementari di un catalogo elettronico, insieme alle informazioni su venditore, produttore, lead time e numerose altre considerazioni (controllo della qualità, versione del catalogo, data e codice identificativo). Il complesso di queste informazioni elementari è chiamato Struttura del Catalogo.

Vediamo l'uso attuale del catalogo: il cliente dopo aver visionato il catalogo manda un ordine di acquisto, il venditore deve dare una conferma che dà inizio al processo di

vendita; per entrambi processare i documenti relativi alla transazione richiede l'uso di un linguaggio condiviso, che sarà oggetto di studio delle prossime pagine.

4.2 Cataloghi di prodotti: descrizione della struttura

I venditori hanno cataloghi che descrivono i loro prodotti ai potenziali clienti e attraverso un B2B marketplace li mettono a disposizione on-line; questo passaggio non è facile come sembra, poiché esistono già molti cataloghi elettronici e vi è una situazione di “confusione”.

Inoltre i cataloghi sono progettati per il lettore umano; ne segue che estrarre le informazioni, classificare i prodotti, normalizzare e archiviare un formato strutturato è la prima attività da svolgere manualmente.

L'aspetto più critico, nel convertire le descrizioni, specialmente con l'idea di automatizzare alcuni passi, è trasformare un testo non strutturato in un documento formale e strutturato che sia machine-readable (Figura 12).

The 2001 CLK320 Cabriolet

Base Price \$48,900*

Vehicle Specifications

Engine and Drivetrain

Engine 3,199-cc SOHC 18-valve V-6, High-pressure die-cast alloy cylinder block, alloy heads.

Net Power 215 hp @ 5,700 rpm

Net torque 229 lb-ft @ 3,000-4,600 rpm

Compression ratio 10.0:1

Fuel Requirement Premium unleaded, 91 pump octane.

Fuel and ignition systems ME 2.8 engine management. Integrated sequential multipoint fuel injection and phased twin-spark ignition includes individual cylinder control of fuel spray, spark timing and phase, and antiknock control. Two high-energy ignition coils and two spark plugs per cylinder and 100,000-mile spark plug intervals. Electronically controlled throttle.

Intake system Magnesium 2-stage resonance intake manifold increases the intake runner length at lower rpm for improved response.

Extract schema and data

Type	Name	Color	Manufacturer	Engine	Net power
Car	CLK320	Gray	DaimlerChrysler	3,299 cc	215 hp

Net	Compression	Unleaded fuel
229 lb-ft	10.0	yes

Figura 12 Estrazione di informazioni da un catalogo prodotti

Questo processo prevede generalmente due attività:

- Definire le categorie di prodotti e i loro attributi (definizione dello schema);
- Estrarre i dati, ovvero gli attuali valori degli attributi precedentemente definiti.

Un terzo passo potrebbe essere meccanizzare l'estrazione dei dati, vale a dire derivare in modo manuale o automatico (con tecniche di machine-learning) delle regole o delle procedure di estrazione. Si può considerare l'estrazione dei dati come un'attività che genera una sorgente di informazioni strutturata (come database o fogli di calcolo) a partire da una sorgente di informazioni non strutturata. Ma da questa meccanizzazione nascono significativi problemi, quali il formato di rappresentazione delle informazioni inadeguato e le irregolarità delle informazioni di rappresentazione. Ogni catalogo differisce sostanzialmente nello stile di presentazione, quando proprio le informazioni multimediali sono difficili da estrarre in modo automatico (per esempio, l'estrazione del colore di un prodotto dall'immagine presente sul catalogo).

La combinazione di tecniche multiple dell'Intelligenza Artificiale può consentire un buon risultato nell'estrazione. Tuttavia persistono due problemi principali: l'incompletezza - solo le informazioni presenti nelle descrizioni di prodotto non strutturate possono essere raccolte - e i valori errati, che potrebbero venire estratti per piccoli, ma inevitabili gradi di inaccuratezza.

Le ontologie di prodotto (content standard) possono giocare un ruolo importante nel ridurre questi problemi, in quanto aiutano nell'individuare le cause e a scegliere tra diverse operazioni, supportano gli utenti a recepire conoscenze aggiuntive.

Un esempio di integrazione di cataloghi prodotto può interessare Ariba²¹ VS CommerceOne²², due standard per la fornitura di applicazioni per attività di business basate sull'XML.

4.3 Descrizione del contenuto dei prodotti

A questo punto del processo di Content Management si può assumere che l'informazione di prodotto sia strutturata in modo tabulare. Ogni venditore usa strutture e vocabolari differenti nella descrizione; ciò non creerebbe problemi se le relazioni

²¹ www.ariba.com

²² www.commerceone.com

fossero di tipo 1 a 1, ma il commercio elettronico attiva la possibilità di relazioni n a m e pertanto non può contare su tale assunto. Le informazioni devono essere classificate secondo uno standard che aiuti venditori e acquirenti a comunicare tra loro.

Classificare i prodotti con uno schema come UNSPSC è difficile, richiede conoscenza del dominio e spesso avviene manualmente; ciò rende l'operazione costosa. Si deve ricercare la manutenibilità e la visibilità delle informazioni sul prodotto. Ci sono problemi a raggiungere un elevato livello di accuratezza a causa dell'eterogeneità dei vocaboli usati nei cataloghi e del limitato numero di termini usati per descrivere le categorie negli standard di classificazione dei prodotti.

Alcuni enti di ricerca (privati o pubblici) hanno provato a meccanizzare la classificazione. Fra questi ProCat²³ offre il supporto di un tool che cataloga automaticamente le descrizioni dei prodotti, fornendo e ottimizzando vari metodi di catalogazione a partire dal reperimento delle informazioni attraverso tecniche di machine learning. La versione corrente utilizza una metafora per svolgere le attività che visualizza la descrizione prodotto come un'interrogazione e lo schema di classificazione come una collezione di documenti.

In generale gli attuali case study mostrano seri problemi nel raggiungere un alto grado di accuratezza. Tipicamente le difficoltà nascono dall'eterogeneità dei vocabolari utilizzati e dal numero limitato di parole usato per descrivere le categorie negli standard di classificazione.

4.4 Descrizione di prodotto a seguito di riclassificazione

Ci sono troppi standard nessuno dei quali è attuale. Alla maggior parte degli standard mancano importanti caratteristiche.

Si consideri ad esempio UNSPSC:

- non descrittivo: non definisce nessun attributo per descrivere i prodotti;
- non intuitivo: venditori e compratori non trovano facilmente i prodotti;
- superficiale: non fornisce abbastanza distinzioni per il mercato verticale che gestisce molti prodotti appartenenti a un certo dominio.

²³ ProCat, ambiente software per automatizzare la catalogazione prodotti, sviluppato presso la Vrije Univeriteit di Amsterdam.

RosettaNet è un esempio di standard che funziona all'opposto, ovvero descrive nel dettaglio prodotti hardware e software relativi a un particolare dominio, ma ne traslascia molti.

Ecl@ss potrebbe rappresentare un'alternativa alla non descrittività di UNSPSC, è un codice gerarchico ed è intuitivo dal punto di vista dell'utilizzatore.

Dato che i diversi venditori e acquirenti usano differenti schemi di classificazione:

- Si devono stabilire dei **mapping** tra i diversi schemi di classificazione, vale a dire dei collegamenti che legano i diversi concetti e attributi;
- Si devono **riclassificare** i prodotti, non c'è bisogno di una corrispondenza 1 a 1 tra i concetti in diversi schemi di classificazione.
- Le vecchie descrizioni devono essere trasformate in un **nuovo stile** di descrizione.

Dal punto di vista dell'implementazione gli standard si muovono attraverso una rappresentazione basata su XML e richiedono la connessione con architetture di basso livello per l'integrazione, fornite dal consorzio W3C con XSLT [Clark, 99] e Xpath [Clark&DeRose, 99].

4.5 Allineamento dei content standard

Gli standard di contenuto danno una gerarchia di descrizioni di prodotto e forniscono una sottoclasse di relazioni tra categorie di prodotti. Si possono dividere in:

- **Standard orizzontali:** forniscono una descrizione ad alto livello di *tutti* i possibili prodotti e domini;
- **Standard verticali:** descrivono più approfonditamente una gamma di prodotti limitata. Imprese di grandi dimensioni tendono a sviluppare speciali schemi di classificazione dei prodotti e ciò si riflette sui sistemi di codifica usati in azienda.

Come mostrato in Figura 13 uno standard verticale espande diverse categorie di uno standard orizzontale di livello gerarchico basso (bottom).

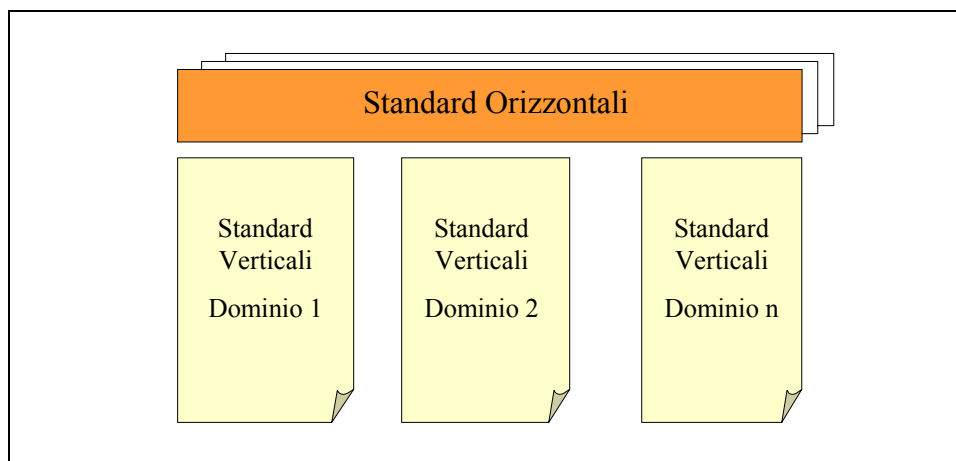


Figura 13 Relazione tra standard verticali e orizzontali

Per riclassificare i prodotti un B2B marketplace deve essere in grado di agire per migliorare tre diversi tipi di mapping:

1. Allineamento di standard orizzontali;
2. Collegamento tra standard orizzontali e verticali;
3. Allineamento di standard verticali.

4.5.1 Allineamento di standard orizzontali

Ci si aspettano molte equivalenze nel mappare questi standard, spesso si può dire che il mapping viene fornito da chi genera lo standard. I principali problemi incontrati sono:

- Solo pochi standard orizzontali hanno ufficialmente pubblicato il mapping;
- Gli standard differiscono nella loro classificazione a causa dell'assenza di uno schema condiviso per la classificazione di tutti i prodotti;
- Gli standard differiscono per il livello di granularità nella classificazione di un particolare gruppo di prodotti, pertanto i mapping pubblicati presentano concetti descritti con vari livelli di generalità;
- L'equivalenza delle categorie non è evidente dalla loro descrizione.

4.5.2 Allineamento tra standard orizzontali e verticali

Gli standard verticali forniscono una classificazione approfondita e ravvicinata, al contrario di quelli orizzontali che ne danno una estesa ma superficiale. Normalmente uno standard verticale espande uno o più nodi di uno standard orizzontale. Questo tipo di allineamento (tra standard orizzontali e standard verticali) richiede:

1. La mappatura di un ristretto numero di concetti *top level* dello standard verticale con concetti più generali dello standard orizzontale;
2. La mappatura di concetti che sono fuori dal focus dello standard verticale con i corrispondenti concetti dello standard orizzontale.

Il collegamento tra standard è più facile dell'allineamento orizzontale e ha un evidente struttura top down.

4.5.3 Allineamento di standard verticali

Richiede il collegamento delle categorie come si è fatto per gli standard orizzontali; in più gli standard hanno un set esteso di attributi che può essere maggiore del set di classi, come nello standard RosettaNet.

L'allineamento di standard verticali richiede:

- Mapping del nome degli attributi e delle abbreviazioni
- Trasformazione dei tipi di attributi
- Mapping della possibile lista dei valori
- Definizione del dominio degli attributi, o il set di categorie alle quali applicare gli attributi.

4.6 Integrazione dei documenti

Gli standard per lo scambio di documenti del B2B contengono un grande numero di documenti che possono essere scambiati tra i partecipanti dell'eMarketplace.

Nella sezione [4.1] si distinguevano tre problemi di integrazione tra dati eterogenei: il primo, relativo ai content standard, è già stato introdotto e sarà oggetto di discussione nei prossimi capitoli; di seguito si analizzeranno le eterogeneità tra cataloghi prodotto e

business document, parlando più in generale di integrazione dei documenti. Gli standard di documenti forniscono solo la struttura dei documenti e non si occupano della semantica.

Si studiano in particolare i frammenti di due cataloghi. I cataloghi contengono una parte statica che comprende descrizioni non aggiornate di frequente e una parte dinamica con informazioni aggiornate frequentemente e mandate su richiesta all'utente.

4.6.1 Esempio di integrazione di documenti

Per chiarire cosa si intende con il concetto di integrazione dei documenti, proponiamo come esempio il problema dell'integrazione degli indirizzi (*address*) descritto in [Omelayenko, 01c]. Un indirizzo è un semplice concetto di business che ricorre frequentemente nell'e-commerce ed è parte rilevante di ogni sistema di mediazione per il B2B. A differenza della maggior parte dei prodotti, la struttura di un indirizzo e il significato dei suoi componenti sono noti a chiunque e rendono l'esempio comprensibile. L'integrazione degli indirizzi coinvolge varie problematiche che ricorrono nell'integrazione dei prodotti. Si prendono in considerazione quattro standard per la descrizione degli indirizzi.

Il primo standard analizzato è xCBL 3.0 sviluppato da CommerceOne. Esso fornisce un set completo di formati di documenti XML standardizzati, che permettono ai compratori, ai fornitori e a chi fornisce i servizi di integrare, nel mercato elettronico, i loro sistemi esistenti in modo veloce ed efficiente [CommerceOne, 00]. La Document Type Definition (DTD) per un indirizzo nello standard xCBL è presentata di seguito:

```
<!ELEMENT OrganizationAddress ( (AddressType)?, (ExternalAddressID),  
(POBox)?, (Street)?, (HouseNumber)?, (StreetSupplement1)?,  
(StreetSupplement2)?, (PostalCode)?, (City), (Country), (Region)?,  
(District)?, (County)?, (TradingPartnerTimezone)? )>  
<!ELEMENT AddressType ( (AddressTypeCoded), (AddressTypeCodedOther)? )>
```

Il secondo standard è Internet Open Trading Protocol (IOTP) [IOTP, 99], sviluppato in collaborazione con il consorzio Internet Engineering Task Force (IETF²⁴), fornisce un formato standard per le operazioni di pagamento nelle transazioni svolte via Internet. È indipendente da qualsiasi specifico sistema di pagamento. IOTP fornisce le strutture ed i

²⁴ www.ietf.org

protocolli di comunicazione per le transazioni di pagamento: acquisto, autenticazione, re-found, deposito e altri protocolli che ricorrono nel commercio elettronico. La sicurezza, l'autenticazione e la firma elettronica sono i principali obiettivi, ne consegue una descrizione dell'indirizzo non elaborata come nel caso precedente. La DTD per un indirizzo nello standard IOTP è:

```
<!ELEMENT PostalAddress EMPTY>
<!ATTLIST PostalAddress
xml:lang NMTOKEN #IMPLIED
AddressLine1 CDATA #IMPLIED
AddressLine2 CDATA #IMPLIED
CityOrTown CDATA #IMPLIED
StateOrRegion CDATA #IMPLIED
PostalCode CDATA #IMPLIED
Country CDATA #IMPLIED
LegalLocation (True | False) "False">
```

Open Applications Group è un consorzio industriale che pubblica specifiche per i contenuti di business nel settore delle applicazioni su Open Applications Group Integration Specification (OAGIS)²⁵, specifiche per messaggi di business e scenari di integrazione. Fornisce strutture di dati, formati e protocolli per le integrazioni di business [OAGIS, 00]. OAGIS definisce un vocabolario per i termini di business e permette di scambiare più di 90 differenti tipi di documenti di business.

Lo standard OAGIS definisce il seguente indirizzo:

```
<!ELEMENT ADDRESS (ADDRLINE*, ADDRTYPE?, CITY?, COUNTRY?,
COUNTY?, DESCRIPTN?, FAX*, POSTALCODE?, REGION?, STATEPROVN?,
TAXJRS DCTN?, TELEPHONE*, URL?, USERAREA?)>
```

La Versione 1.0 del Real Estate Data Interchange Standard (RETS, 2000) definisce le informazioni per lo scambio di beni. Mostra un'interfaccia standard attraverso la quale un programma client può comunicare con un server di dati proprietario o no. Le specifiche definiscono un protocollo per implementare le transazioni, incorporano una specifica XML per transazioni generiche; inoltre, forniscono un formato per lo scambio di dati compressi e specifiche per permettere lo scambio di informazioni proprietarie interpretabili dalla macchina. Le strutture dati per lo scambio sono definite nel Real Estate Transaction Markup Language (RETML)²⁶, gli indirizzi sono così rappresentati:

```
<!ELEMENT MailingAddress (StreetAddress)>
```

²⁵ www.openapplications.org

²⁶ www.rets-wg.org

```
<!ELEMENT StreetAddress ((StreetNumber?, BoxNumber?, StreetDirPrefix?,
StreetName,
StreetAdditionalInfo?, StreetDirSuffix?, StreetSuffix?, UnitNumber?,
City?,
StateOrProvince?, Country?, PostalCode?, CarrierRoute?) |
Unstructured)>
```

La rappresentazione dello stesso concetto, “address”, è differente nei diversi cataloghi. La descrizione dei prodotti può essere espressa in XML utilizzando i tag in modi diversi. Concettualmente uguali, le proprietà dei prodotti possono essere descritte tramite elementi XML con nomi diversi. Gli elementi contrassegnati con gli stessi tag possono avere semantiche differenti. Anche l’ordine dei tag è importante nell’XML. Alla fine, alcune proprietà dei prodotti possono essere descritte con differenti livelli di granularità a seconda delle richieste dell’applicazione. Per esempio, un sistema di spedizione richiede la divisione dell’indirizzo di una società in nome della via, numero civico, distretto e così via, mentre per un sistema di pagamento elettronico questi dettagli non sono così importanti, come si può vedere nei sistemi generici xCBL, OAGIS e nel sistema di pagamento IOPT. Allo stesso tempo alcune agenzie richiedono maggiori dettagli nella descrizione degli indirizzi, come si vede dallo standard RETML.

Esempio dal catalogo IOPT:

```
<PostalAddress AddressLine1="Division of Mathematics and Computer
Science"
AddressLine2="De Boelelaan 1081a" CityOrTown="Amsterdam"
Country="Netherlands"
LegalLocation="True" PostalCode="1081 hv" StateOrRegion="Horth
Holland"
xml:lang="en"/>
```

Esempio dal catalogo OAGIS:

```
<ADDRESS>
<ADDRLINE>De Boelelaan, 1081a</ADDRLINE>
<ADDRTYPE>office</ADDRTYPE>
<CITY>Amsterdam</CITY>
<COUNTRY>Netherlands</COUNTRY>
<COUNTY>North Holland</COUNTY>
<DESCRIPTN/>
<FAX>3120444765</FAX>
<POSTALCODE>1081 hv</POSTALCODE>
<REGION/>
<STATEPROVN/>
<TAXJRS DCTN/>
<TELEPHONE>3120444765</TELEPHONE>
<URL>www.cs.vu.nl</URL>
<USERAREA/>
</ADDRESS>
```

Se il marketplace media tra n venditori e m compratori, deve essere in grado di mappare ciascuno degli n cataloghi dei fornitori negli m formati dei compratori, svolgendo $n*m$ mapping (Figura 14).

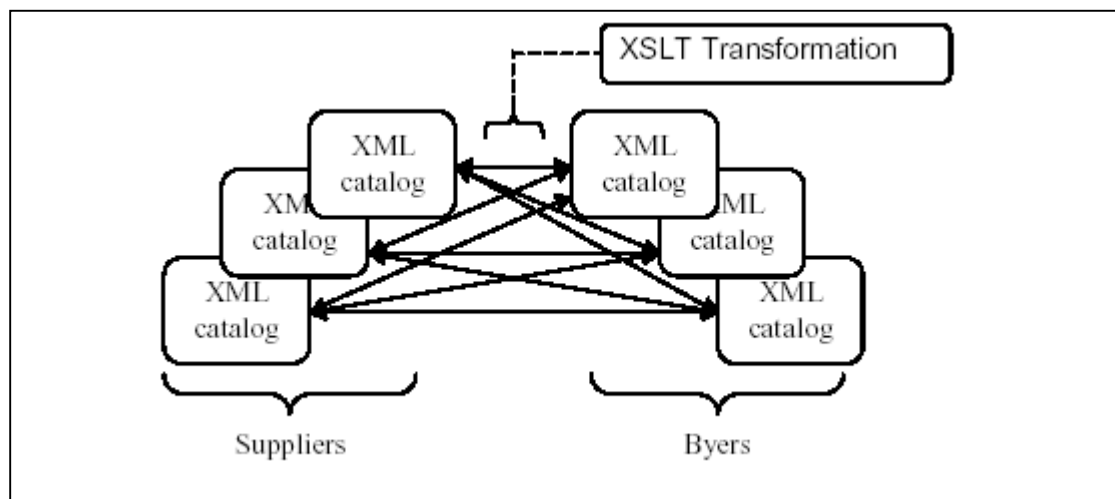


Figura 14 Integrazione con $n*m$ mapping

L'introduzione di un catalogo intermedio, chiamato Unified Catalog (UC), richiede solo al marketplace di svolgere il mapping tra ogni catalogo dei venditori o dei compratori e l'UC, quindi, richiede solo $n+m$ mapping (Figura 15).

Ci sono due opposte strategie per selezionare gli elementi da includere nell'UC:

1. Il catalogo unificato contiene il numero minimo di attributi per ogni prodotto;
2. Il catalogo unificato contiene il numero massimo di prodotti.

In entrambe le strategie l'UC può cambiare se viene introdotto un nuovo catalogo. Seguendo la prima strategia, l'aggiunta di un catalogo più dettagliato non cambierà l'UC, ma l'introduzione di uno meno dettagliato ne ridurrà il livello di granularità. Il risultato di questa strategia è limitare il livello di granularità dell'UC al catalogo meno dettagliato, cosa che è inaccettabile per la maggior parte dei sistemi di B2B. Nella seconda strategia, l'aggiunta di un catalogo meno dettagliato non influenza l'UC, mentre quella di uno più dettagliato ne richiede l'aggiornamento.

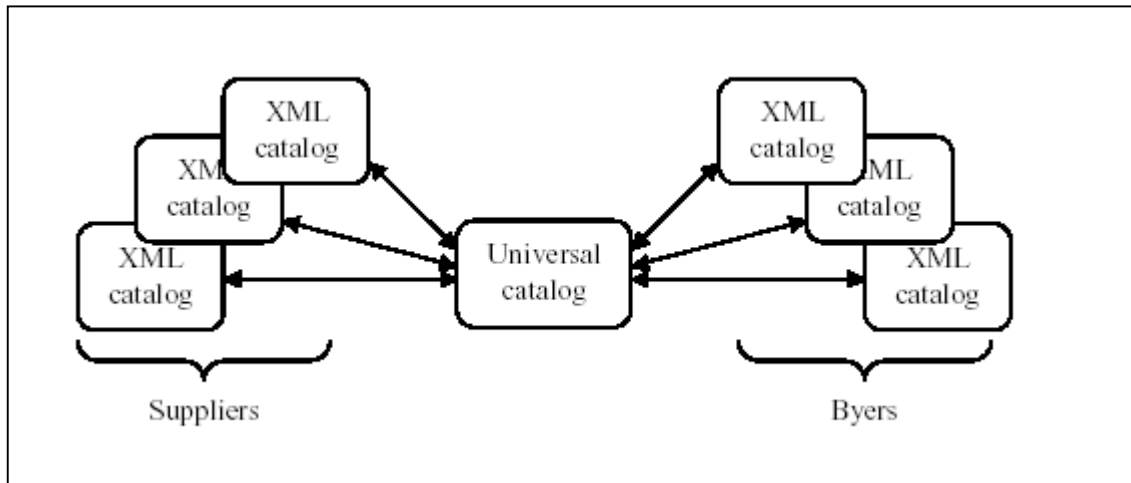


Figura 15 Integrazione con n+m mapping

CBL	IOTP	OAGIS	RETML	UC
OrganizationAddress	PostalAddress	ADDRESS	StreetAddress	address
Street?			StreetName	Street
HouseNumber?			StreetNumber?	House
	AddressLine1	ADDRLINE		
	AddressLine2			
POBox?	-	-	BoxNumber?	PObox
PostalCode?	PostalCode	POSTALCODE ?	PostalCode?	Postcode
City	CityOrTown	CITY?	City?	City
Region?	StateOrRegion	STATEPROVN?	StateOrProvince?	Province
District?	-	REGION?	-	Region
County?	-	COUNTY?	-	District
Country	Country	COUNTRY?	Country?	Country
StreetSupplement1?	LegalLocation	DESCRIPTN?	StreetDirPrefix?	Description
StreetSupplement2?		ADDRTYPE?	StreetAdditionalInfo?	
			StreetDirSuffix?	
			StreetSuffix?	
			UnitNumber?	
-	-	FAX?	-	Fax
-	-	TELEPHONE?	-	Phone
-	-	URL ?	-	URL
-	-	-	CarrierRoute ?	Route
-	xml:lang	-	-	xml:lang
-	-	TAXJRSCTN?	-	TaxDistrict
-	-	USEAREA?	-	Area
TradingPartner	-	-	-	TimeZone
Timezone?				

Figura 16 Mapping relativo all'esempio Address

4.6.2 Integrazione a un solo livello

Ci sono molte differenze nei formati di documenti che riportano le stesse informazioni:

- Diverse terminologie;
- Diversa scelta di usare attributi o elementi;
- Diversi formati dei valori;
- Diverse scale di valori.

Gli standard sono spesso rappresentati in XML. L'integrazione di documenti in XML si potrebbe effettuare con XSL-T²⁷ attraverso tre passaggi:

1. Allineando la terminologia;
2. Allineando la granularità;
3. Trasformando i valori degli attributi.

Tuttavia sorgono problemi nell'integrazione ad un solo livello, poiché avvengono contemporaneamente due operazioni:

- Traduzione sintattica;
- Mapping semantico.

Questi due tipi di trasformazione appartengono a diversi livelli di rappresentazione. Il livello sintattico individua un modo per serializzare il contenuto delle informazioni in una sequenza di caratteri; il livello semantico fornisce un modello concettuale delle informazioni.

Per evitare confusione tra le due trasformazioni si è arrivati alla teorizzazione di un processo di integrazione multi-livello, presentato nel prossimo paragrafo

4.6.3 Integrazione multilivello (layered approach)

Il concetto di integrazione con approccio multilivello per la rappresentazione delle informazioni sul Web è stato proposto inizialmente in [Melnik&Decker, 00] con la costruzione di un modello su tre livelli (Figura 17):

²⁷ XSL Transformations (XSL-T) è un modo standard di descrivere come trasformare la struttura di un documento XML in un documento XML con struttura diversa. XSL-T è una raccomandazione del W3C.

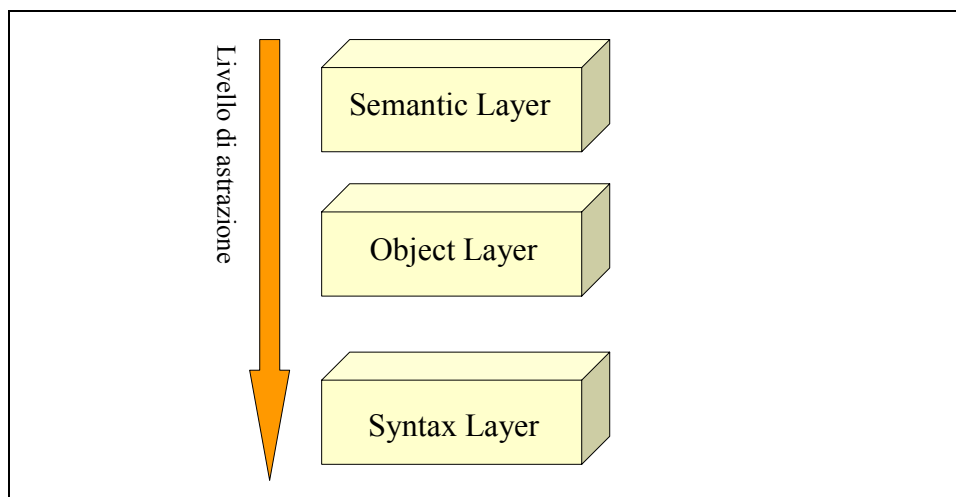


Figura 17 Approccio di integrazione multi-livello

Il livello sintattico (Syntax layer) fornisce un modo di serializzare il contenuto delle informazioni in una sequenza di caratteri utilizzando ad esempio XML.

Il livello ad oggetti (Object layer) ha lo scopo di offrire una vista object-oriented delle informazioni con un modello dei dati normalizzato.

Il livello semantico (Semantic layer) fornisce un modello concettuale delle informazioni.

Sulla base di questo modello è nato un altro approccio multilivello descritto in [Omelayenko, 01a]:

- **Livello sintattico:** corrisponde ai documenti rappresentati con una serializzazione XML che specificano elementi e attributi usati nel loro ordine;
- **Livello data-model:** è un ponte tra il livello ontologico e il livello sintattico. Le rappresentazioni sono astratte a partire dalla differenze imposte dal livello sintattico, gli oggetti sono rappresentati come triple oggetto-proprietà-oggetto.
- **Livello ontologico:** corrisponde all'ontologia del documento usato per rappresentare i prodotti, questo livello specifica il documento con il maggiore grado di dettaglio, sufficiente a specificare le trasformazioni tra cataloghi.

4.7 Integrazione dei modelli di dati

Le difficoltà di un approccio ad un solo livello derivano dall'integrazione contemporanea di vari task che operano insieme.

Per risolvere questo problema, in base ai modelli descritti nella sezione [4.6] si propone un approccio multilivello con il quale l'integrazione è resa performante passando almeno attraverso due livelli: il livello sintattico dei documenti XML e il livello dei modelli di dati normalizzati (codificati) per cataloghi.

Nel realizzare l'integrazione si effettuano tre step (Figura 18):

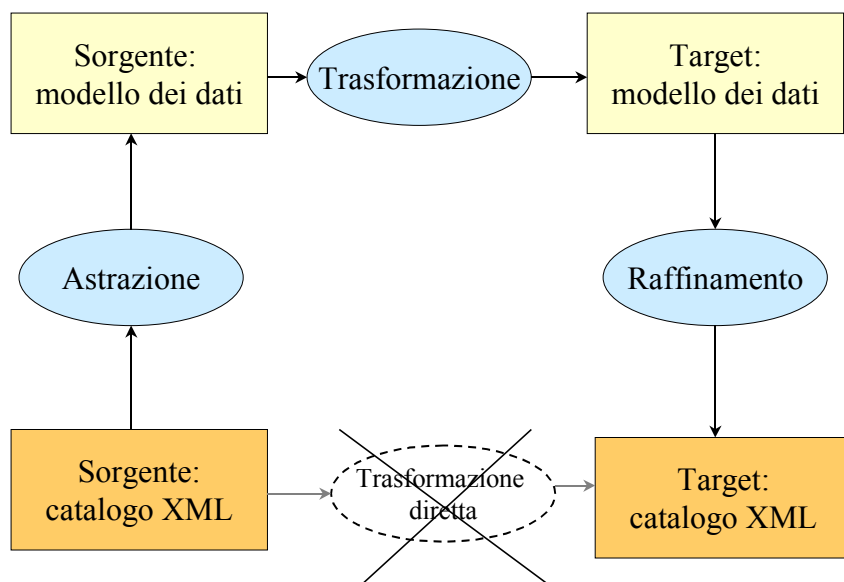


Figura 18 Il modello per la trasformazione dei dati

1. **Step di astrazione** (*Abstraction Step*): traduzione di un catalogo sorgente XML in un modello di dati codificato con triple RDF attraverso i seguenti passaggi:
 - traduzione di ogni elemento o attributo XML in tripla;
 - split di un elemento XML in due o più triple;
 - concatenazioni di descrizioni multifide in un solo file;
 - inclusione di elementi opzionali XML nel modello RDF.
2. **Step di trasformazione** (*Translation Step*): traduzione da un modello di dati normalizzato al modello di dati obiettivo (target).

3. **Step di raffinamento** (*Refinement Step*): traduzione dal modello di dati obiettivo al catalogo XML obiettivo attraverso successivi passaggi:
- ogni tripla RDF è tradotta nella corrispondente XML;
 - gli elementi target di XML sono creati nell'ordine appropriato;
 - una o più triple di RDF sono riunite in un singolo elemento XML.

Questa formalizzazione conclude la presentazione dell'integrazione di documenti. Nei prossimi capitoli mi occuperò di integrazione di content standard nell'ambito del progetto MOMIS, un sistema per l'Integrazione Intelligente delle Informazioni.

Capitolo 5

MOMIS: un sistema intelligente di integrazione

Le problematiche connesse all'integrazione di informazioni sono legate alla grande eterogeneità dei dati disponibili, sia per quanto riguarda la natura (testi, immagini, ecc.), sia il modo in cui vengono descritti nelle diverse sorgenti. Gli standard esistenti (TCP/IP, ODBC, OLE, CORBA, SQL, ecc.) risolvono parzialmente i problemi relativi alle diversità hardware e software, ai protocolli di rete e di comunicazione tra i moduli; rimangono però irrisolti quelli relativi alla modellazione delle informazioni. Difatti i modelli e gli schemi dei dati sono differenti e questo crea un'eterogeneità semantica (o logica) non risolvibile da questi standard.

Problematiche aggiuntive sono dovute al cosiddetto sovraccarico di informazioni, o *information overload*, infatti l'enorme mole di dati disponibili impedisce all'utente di discernere e isolare le informazioni significative. Altri problemi sono dovuti ai tempi di accesso, alla salvaguardia della sicurezza e agli elevati costi per il mantenimento e la consistenza delle informazioni.

Per far fronte alla molteplicità e complessità degli aspetti appena descritti, le architetture dedicate all'integrazione di sorgenti eterogenee devono essere necessariamente flessibili e modulari.

Gli approcci all'integrazione presentano diverse metodologie: la reingegnerizzazione delle sorgenti mediante standardizzazione degli schemi e la creazione di un database distribuito; i data warehouse che realizzano delle viste presso l'utente finale, replicando fisicamente i dati e utilizzando algoritmi di allineamento per assicurarne la consistenza a fronte di modifiche nelle sorgenti.

Nel seguito verrà descritto un tipo di approccio differente, che non ricorre alla duplicazione fisica dei dati, indicato in letteratura come Integrazione di Informazioni (I^2) [Wiederhold, 96].

5.1 Integrazione Intelligente di Informazioni

L'Integrazione delle Informazioni va dunque distinta da quella dei dati (e dei database), per ottenere risultati essa richiede conoscenza ed intelligenza volte all'individuazione delle sorgenti e dei dati, nonché alla loro fusione e sintesi.

Quando l'Integrazione di Informazioni fa uso di tecniche di Intelligenza Artificiale (IA) si parla allora di Integrazione Intelligente di Informazioni (*Intelligent Integration of Information*, I^3).

5.1.1 Il programma I^3

Dal 1992 è operativo il Programma I^3 [Hull, 95] un progetto fondato e sponsorizzato dall'ARPA (Advanced Research Projects Agency), che si prefigge di individuare un'architettura di riferimento che realizzi in maniera automatica l'integrazione di sorgenti di dati eterogenee. I^3 propone l'introduzione di architetture modulari sviluppabili secondo i principi proposti da uno standard che ponga le basi dei servizi da soddisfare dall'integrazione ed abbassi i costi di sviluppo e mantenimento. Questo renderebbe possibile ovviare ai problemi di realizzazione, manutenzione, adattabilità, inoltre la riutilizzazione della tecnologia già sviluppata, rende la costruzione di nuovi sistemi più veloce e meno difficoltosa, con conseguente abbassamento dei costi. Per poter sfruttare un'elevata riusabilità bisogna disporre di interfacce ed architetture standard. Il paradigma suggerito per la suddivisione dei servizi e delle risorse nei diversi moduli si articola su due dimensioni:

- orizzontale, divisa in tre livelli: livello utente, moduli intermedi che fanno uso di tecniche di IA, risorse di dati;
- verticale, molti domini, con un numero limitato (e minore di 10) di sorgenti.

I domini nei vari livelli non sono strettamente connessi, ma si scambiano dati ed informazioni la cui combinazione avviene a livello dell'utente, riducendo la complessità totale del sistema e permettendo lo sviluppo di applicazioni con finalità diverse. I^3 si concentra sul livello intermedio della partizione, quello che media tra gli utenti e le sorgenti. Questo livello deve offrire servizi dinamici quali la selezione delle sorgenti, la gestione degli accessi e delle interrogazioni, l'analisi e sintesi dei dati.

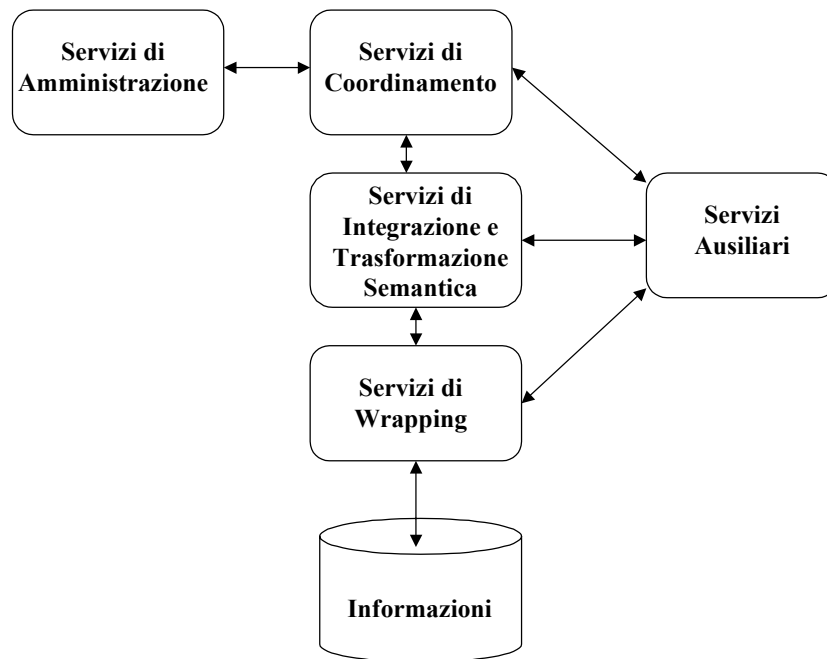


Figura 19 Diagramma dei servizi I^3

5.1.2 Architettura di riferimento per sistemi I^3

L'obiettivo del programma I^3 è di ridurre il tempo necessario per la realizzazione di un integratore di informazioni, fornendo una raccolta e una formalizzazione delle soluzioni prevalenti finora nel campo della ricerca. Come abbiamo visto, la complessità del processo di integrazione è tale da rendere estremamente utile la proposta di un'architettura di riferimento standard, che rappresenti alcuni dei servizi che un

integratore di informazioni deve contenere e le possibili interconnessioni fra di loro. Il programma I^3 individua cinque famiglie di attività omogenee, illustrate in Figura 19 unitamente ai loro legami. La reciproca interazione tra queste attività consente di eseguire le operazioni di comunicazione, traduzione ed integrazione dei dati nelle sorgenti. Analizziamo nel dettaglio questi servizi:

I **Servizi di Coordinamento** sono servizi ad alto livello che costituiscono l'interfaccia con l'utente, dandogli l'impressione di trattare con un sistema omogeneo.

Grazie alle funzionalità messe a disposizione dalle altre famiglie, essi permettono l'individuazione delle sorgenti di dati interessanti, ovvero delle sorgenti che possono fornire una risposta ad una determinata richiesta dell'utente.

I moduli di coordinamento possono essere più o meno complessi in funzione delle modalità di selezione delle sorgenti:

- **Facilitator e Brokering Services:** forniscono una selezione dinamica delle sorgenti in grado di soddisfare la richiesta dell'utente. Il sistema usa un deposito di metadati per individuare il modulo che può trattare direttamente questa richiesta, in particolare si parla di Brokering quando è coinvolto un modulo alla volta, oppure di Facilitatori o Mediatori se vi sono più moduli interessati. In quest'ultimo caso la query iniziale viene decomposta in un insieme di sottoquery da inviare a differenti moduli che gestiscono sorgenti distinte, successivamente vengono integrate le risposte per fornirne una presentazione globale all'utente.
- **Matchmaking:** il mapping fra informazioni integrate e locali è effettuato manualmente da un operatore in fase di inizializzazione. In questo caso tutte le richieste vengono trattate allo stesso modo.

I **Servizi di Amministrazione** sono utilizzati da quelli di Coordinamento per individuare le sorgenti, determinarne le capacità, creare ed interpretare *template*.

I *template* sono strutture dati che descrivono i servizi, le sorgenti ed i moduli da utilizzare per portare a termine un determinato task; essi servono per ridurre al minimo le possibilità di decisione del sistema, consentendo di definire a priori le azioni da eseguire a fronte di una determinata richiesta. In alternativa ad essi si possono utilizzare le *yellow pages*: servizi di directory che mantengono le informazioni sul contenuto delle sorgenti e sul loro stato (attiva, inattiva, occupata), consentendo al Mediatore di inviare la richiesta di informazioni alla sorgente giusta o, se non fosse disponibile, ad una

equivalente. Fa parte di questa famiglia di servizi il modulo denominato *Browser* che permette appunto di “navigare” tra le descrizioni degli schemi delle sorgenti, recuperando informazioni.

Altri moduli interessanti sono gli *Iterative Query Formulation*, che aiutano l’utente a rilassare o specificare meglio alcuni vincoli dell’interrogazione al fine di ottenere risposte più precise.

I **Servizi di Integrazione e Trasformazione Semantica** hanno come input una o più sorgenti dati, tradotte dai servizi di Wrapping, e generano una vista integrata o trasformata di queste informazioni. Essendo tipici dei moduli mediatori vengono indicati spesso come servizi di mediazione. I principali sono:

- Servizi di integrazione di schemi: creano il vocabolario e le ontologie condivise dalle sorgenti, integrano gli schemi in una vista globale, mantengono il mapping tra schemi globali e sorgenti;
- Servizi di integrazione di informazioni: aggregano, riassumono ed astraggono i dati per fornire presentazioni analitiche significative;
- Servizi di supporto al processo di integrazione: sono utilizzati quando una query deve essere scomposta in più sottoquery da inviare a fonti differenti, con la necessità di integrare poi i loro risultati.

I **Servizi di Wrapping** si comportano come traduttori dai sistemi locali ai servizi di alto livello dell’integratore e viceversa quando si interroga la sorgente dati. Essi realizzano il primo passo verso l’integrazione, rendendo le informazioni provenienti dalle sorgenti omogenee grazie alla loro traduzione in un linguaggio standard. Inoltre, fornendo interfacce che seguono gli standard più diffusi (ad esempio il linguaggio SQL come linguaggio di interrogazione di basi di dati e CORBA come protocollo di scambio di oggetti), consentono alle sorgenti estratte da questi wrapper “universali” di essere accedute dal maggior numero possibile di moduli mediatori.

I **Servizi Ausiliari** aumentano le funzionalità degli altri servizi; vanno dai semplici servizi di monitoraggio del sistema ai servizi di propagazione degli aggiornamenti (mantenimento della consistenza dei dati), dai servizi di arricchimento semantico (già accennati in precedenza) a quelli di ottimizzazione.

In Figura 19 risultano evidenti i due assi orizzontale e verticale, che rappresentano le

diverse interazioni fra i servizi appena descritti. Percorrendo l'asse orizzontale, si nota il rapporto fra servizi di Coordinamento ed Amministrazione, mentre sull'asse verticale viene messo in evidenza lo scambio di informazioni all'interno del sistema: i dati estratti dalle sorgenti per mezzo dei servizi di Wrapping sono integrati dai servizi di Integrazione e Trasformazione semantica per poi essere passate ai servizi di Coordinamento che ne avevano fatto richiesta.

Concludendo questa rapida descrizione dell'architettura di riferimento per sistemi I^3 è opportuno sottolineare che, essendo la casistica dei campi applicativi interessata molto vasta, le funzionalità di base descritte non sono esaustive, ogni ambiente di sviluppo individuerà obiettivi e problematiche differenti e di conseguenza anche funzionalità specifiche.

5.1.3 Il mediatore

Secondo la definizione proposta da Wiederhold [Duschka, 96]:

“un mediatore è un modulo software che sfrutta la conoscenza su un certo insieme di dati per creare informazioni per un'applicazione di livello superiore...Dovrebbe essere piccolo e semplice, così da poter essere amministrato da uno, o al più pochi, esperti.”

Un mediatore presenta allora i seguenti compiti:

- assicurare un servizio stabile, anche nel caso di cambiamento delle risorse;
- amministrare e risolvere le eterogeneità delle diverse fonti;
- integrare le informazioni ricavate da più risorse;
- presentare all'utente le informazioni attraverso un modello scelto dall'utente stesso.

Il progetto MOMIS, di cui si parlerà nella seconda parte del capitolo ha come obiettivo la progettazione e realizzazione di un Mediatore.

L'ipotesi di avere a che fare esclusivamente con sorgenti di dati strutturati e semistrutturati, ha consentito di restringere il campo applicativo del sistema con una conseguente diminuzione delle problematiche riscontrate in fase di progettazione e realizzazione. L'approccio architetturale scelto quello classico, che si sviluppa su tre livelli principali:

1. utente: attraverso un'interfaccia grafica l'utente pone delle query su uno schema globale e riceve un'unica risposta, come se stesse interrogando un'unica sorgente di informazioni;
2. mediatore: il mediatore gestisce l'interrogazione dell'utente, combinando, integrando ed eventualmente arricchendo i dati ricevuti dai wrapper, ma usando un modello (e quindi un linguaggio di interrogazione) comune a tutte le fonti;
3. wrapper: ogni wrapper gestisce una singola sorgente, convertendo le richieste del mediatore in una forma comprensibile dalla sorgente, e le informazioni da essa estratte nel modello usato dal mediatore.

Rispetto al progetto I^3 l'architettura del mediatore MOMIS esamina i servizi di Integrazione e Trasformazione Semantica; sono forniti dal mediatore servizi che facilitino l'integrazione sia degli schemi che delle informazioni.

Parallelamente a questa impostazione architetturale il progetto si vuole distaccare dall'approccio strutturale, cioè sintattico, tuttora dominante tra i sistemi presenti sul mercato.

5.2 Il sistema MOMIS

MOMIS (Mediator EnvirOnment for Multiple Information Sources) [Bergamaschi, 99] è un sistema per l'integrazione intelligente delle informazioni, che nasce dalla collaborazione tra l'Università di Modena e Reggio Emilia e l'Università di Milano e Brescia.

Come si può osservare in Figura 20 la struttura di MOMIS ricalca le specifiche I^3 ; lo schema evidenzia l'organizzazione a tre livelli utilizzata.

Al livello dei dati (il più basso) abbiamo i wrapper, con la funzione di interfaccia tra il mediatore e le sorgenti, che per il sistema MOMIS possono essere sia strutturate (database), sia semistrutturate, come i documenti XML.

I compiti del wrapper sono essenzialmente due:

1. nella fase di integrazione: raccolgono le informazioni sulle strutture dei dati delle sorgenti e ne forniscono una descrizione nel linguaggio ODL₁₃ [Bergamaschi, 98] utilizzato dal sistema;

2. nella fase di query processing: traducono la query, inviata dal mediatore nel linguaggio OQL₁₃, in un linguaggio interpretabile dalla singola sorgente.

Quindi operano il processo inverso per restituire al mediatore i risultati della query. Al livello centrale, troviamo il Mediatore, il cuore dell'intero sistema.

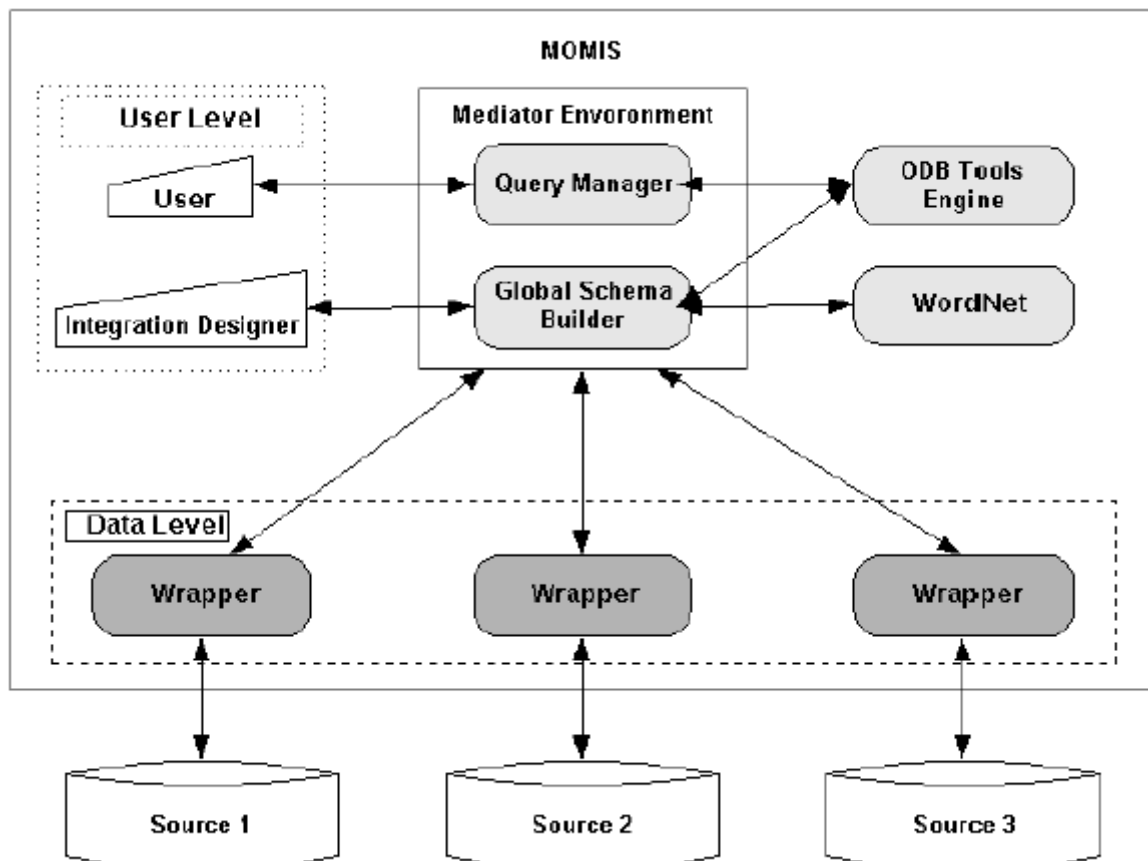


Figura 20 Architettura del sistema MOMIS

Il mediatore di MOMIS è costituito da due moduli:

- il Global Schema Builder (GSB) che integra gli schemi delle diverse sorgenti in base alle descrizioni in ODL₁₃ fornite dai wrapper e crea, con l'intervento del progettista, uno schema globale al quale farà riferimento nell'interazione con l'utente;
- il Query Manager [Zaccaria, 98] che, con l'uso di tecniche di logica descrittiva, scompone e ottimizza la query che fornirà in linguaggio OQL₁₃ ai wrapper che le applicheranno alle singole sorgenti.

L'ultimo livello che troviamo è quello in cui interviene l'utente, il quale opera

interrogazioni sullo schema globale, query che passando poi per il query manager e i wrapper saranno sottoposte direttamente alle sorgenti originali.

5.2.1 Il linguaggio ODL_{I3}

Il linguaggio ODL_{I3}, realizzato in accordo con le raccomandazioni ODMG [Cattel, 94] [Cattel, 97], estende le capacità del sistema ODL che già permetteva:

- definizione di tipi-classe e tipi-valore;
- distinzione fra intensione ed estensione di una classe di oggetti;
- definizione di attributi semplici e complessi;
- definizione di attributi atomici e collezioni (set, list, bag);
- definizione di relazioni binarie con relazioni inverse;
- dichiarazione della signature dei metodi.

A queste sono state aggiunte le seguenti estensioni:

- il costruttore **union**: per esprimere le strutture dati alternative nelle definizioni delle classi ODL_{I3};
- il costruttore **optional(*)**: per specificare se l'attributo è opzionale per l'istanza;
- le regole per l'integrità: al fine di esprimere le regole di integrità *if-then* nei livelli intra e inter-schema;
- le relazioni intensionali: che sono relazioni terminologiche di sinonimia **SYN**, ipernimia **BT**, iponimia **NT** e omonimia **RT**, che esprimono la conoscenza interschema;
- le relazioni estensionali: le relazioni SYN, BT, NT, RT estese alle classi;
- le regole di mapping: per esprimere le relazioni tra le grandezze globali e quelle locali.

5.2.2 Gli strumenti di MOMIS

Per il suo funzionamento il sistema MOMIS fa uso di due tool esterni:

- ODB tools: sia in fase di integrazione che in fase di query processing;

- WordNet: durante lo sviluppo del Thesaurus comune durante il processo di integrazione.

ODB tools [Beneventano, 97a] è un software che verifica la validità degli schemi dei database e che ottimizza semanticamente le query su basi di dati ad oggetti e si basa su due elementi:

1. l'Object Language with Complements allowing Description cycles (OLCD): linguaggio usato per esprimere gli schemi, le query e i vincoli di integrità; dotato di tecniche di inferenza basate sulla sussunzione, tecniche tratte dalle Logiche Descrittive per l'Intelligenza Artificiale;
2. l'espansione semantica dei tipi, sempre attraverso l'algoritmo di sussunzione.

ODL Interface	OQL Interface
Schema Validator	Query Optimizer
Graphic Interface	
ODB-Designer	ODB-QOptimizer

Figura 21 Struttura di ODB tools

Come si osserva in Figura 21 ODB tools è costituito da due parti:

- ODB Designer,
- ODB Qoptimizer [Beneventano, 97b].

ODB Designer acquisisce e valida schemi OODB e dopo averli trasformati in schemi OLCD, con l'algoritmo di sussunzione, determina le relazioni di specializzazione tra i tipi.

ODB Optimizer serve per ottimizzare semanticamente le query, ossia in virtù delle ipotesi fatte, sui vincoli di integrità che possono essere utilizzati per ottimizzare le interrogazioni, si trasforma una query in una equivalente, al limite più complessa, ma maggiormente efficiente.

Sia il processo di validazione che quello di ottimizzazione semantica sono basati sul concetto di espansione semantica di un tipo che permette di incorporare ogni possibile

restrizione che non è presente nel tipo originale, ma che è logicamente implicata dallo schema. Questo processo determina l'inserimento nello schema di nuove relazioni *isa* e quindi nuove gerarchie di ereditarietà.

Il secondo tool usato da MOMIS è WordNet [Miller, 95], un database lessicale realizzato dal Cognitive Science Laboratory di Princeton e già descritto in [2.6].

5.3 Wrapping degli schemi sorgente

Per gestire l'eterogeneità delle informazioni, il mediatore raccoglie ogni sorgente in un wrapper, che converte logicamente (mediante algoritmi) la struttura dei dati sottostanti in un modello di dati comuni, il linguaggio ODL_{B} con cui operano i tool di MOMIS. L'architettura del wrapper e le interfacce di MOMIS sono cruciali per la gestione delle diverse sorgenti, poiché, in parte agendo in maniera automatica, in parte attraverso l'interazione con il progettista, riescono a costruire la vista integrata degli schemi delle sorgenti connesse [Roth, 97].

Nelle prove fatte su MOMIS per il progetto studiato in tesi, si sono utilizzate sorgenti di dati semistrutturati, in particolare file XML. Di seguito, quindi, mi soffermerò nella descrizione di questo tipo di sorgenti.

5.3.1 I dati semistrutturati

I dati semistrutturati sono “self-describing”, pertanto l'informazione associata allo schema è specificata all'interno dei dati [Buneman' 97]. Esempi possono essere l'XML (eXtensible Markup Language) e l'HTML (Hypertext Markup Language).

MOMIS presenta le sorgenti di dati semistrutturati come un grafo con rami e etichette, nel quale i dati sono i nodi e le etichette sono nei vertici.

Un oggetto semistrutturato può essere visto come una tripla di forma:

$\langle id, label, value \rangle$

dove *id* è l'identificatore oggetto, *label* è la stringa che descrive cosa rappresenta l'oggetto, *value* è il valore, che può essere atomico (es. integer, real, ...) o complesso (set di dati semistrutturati, con oggetto “padre” e oggetti “figli”).

Per rappresentare lo schema di una sorgente di dati semistruzzurati si introduce la nozione di *object pattern*: tutti gli oggetti di una sorgente sono partizionati in set disgiunti, tali che tutti gli oggetti appartenenti allo stesso set hanno la stessa etichetta. Quindi l'*object pattern* è estratto da ogni set per rappresentare tutti gli oggetti nel set; in sostanza è rappresentativo di tutti gli oggetti che all'interno di una sorgente descrivono lo stesso concetto.

5.3.2 Sorgenti XML

XML è nato come linguaggio per descrivere le sorgenti di informazioni usando un formato universale. Uno dei principali obiettivi di questo standard è di scambiare file attraverso Internet.

Come sorgente semistruzzurata XML presenta alcune caratteristiche salienti:

- Attributo di *object pattern* \Rightarrow XML tag
- *object pattern* \Rightarrow DTD element
- Valore atomico di un attributo di *object pattern* \Rightarrow PCDATA value

Dalla schematizzazione si deduce che il file DTD (Document Type Definition) rappresenta il modello strutturale dei dati ed è necessario per la costruzione del file XML vero e proprio contenente i dati.

Ciascun documento XML ha una struttura logica e fisica. Fisicamente, il documento è composto di unità dette entità. Un'entità può indirizzare altre entità per includerle nel documento. Un documento inizia con una "radice" o entità documento, di cui nessuna parte occorre nel contenuto di qualche altro elemento. Logicamente, il documento è composto di dichiarazioni, elementi, commenti, riferimenti a caratteri, e istruzioni di elaborazione, ciascuno dei quali è indicato nel documento da espliciti markup.

Un documento XML è ben formato se:

1. Il documento contiene uno o più elementi.
2. C'è un solo elemento, detto radice, o "document element", di cui nessuna parte occorre nel contenuto di qualche altro elemento. Per tutti gli altri elementi, se il tag di partenza si trova nel contenuto di un altro elemento allora anche il tag di fine si

- trova nel contenuto dello stesso elemento. Più semplicemente si può dire che, gli elementi, delimitati dal tag-di-inizio e dal tag-di-fine si annidano uno dentro l'altro²⁸.
3. Ciascuna entità analizzata che viene riferita direttamente o indirettamente all'interno del documento risulta ben-formata.

5.3.3 Wrapping di sorgenti XML

XML permette di descrivere dati semistrutturati in accordo con il modello di dati da rappresentare.

Il Wrapper XML analizza la DTD associata ad ogni file XML ben formato e genera la traduzione degli statement XML in statement ODL_{IS}; questo passaggio implica alcuni aspetti critici dovuti alla mancanza di semantica di XML rispetto a ODL_{IS}, tra cui:

- l'ordine con il quale gli attributi sono descritti nella DTD;
- la traduzione del concetto di attributo da XML a ODL_{IS};
- la debole semantica tra relazioni intra-schema (in questo caso è necessario il supporto del progettista).

Occorre tenere presente XML Schema, uno standard W3C recente, che consente di esprimere meglio la semantica e la struttura con il linguaggio XML.

5.4 Il processo di integrazione

La prima parte del processo di integrazione del sistema MOMIS, come si vede in Figura 22, è quella che porta alla generazione del Thesaurus Comune in cui viene raccolta la conoscenza delle informazioni semantiche relative al contesto e alla struttura dei vari schemi sorgente.

²⁸ Tag: parola compresa tra due parentesi angolari, "<" e ">". Letteralmente "etichetta", nel contesto si fa riferimento a tag utilizzati per evidenziare, in un format gerarchico, i differenti componenti del documento.

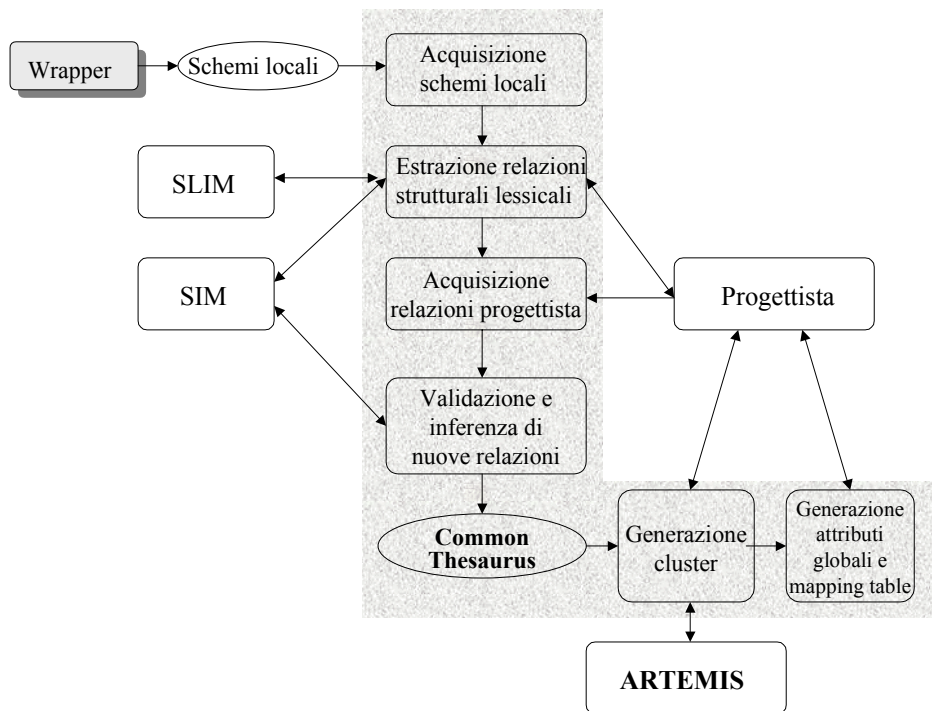


Figura 22 Prima fase del processo di integrazione

Terminata la costruzione del Thesaurus Comune si prosegue con il calcolo delle affinità tra classi e termini, il cui risultato viene sfruttato dal modulo ARTEMIS per la formazione dei cluster in cui vengono raggruppate le classi; cluster per ognuno dei quali nella fase conclusiva dell'integrazione viene generata una classe globale e una *mapping table*.

5.5 Generazione del Thesaurus comune

Questa parte dell'integrazione viene operata, come mostra Figura 22, per mezzo dei moduli SIM, SLIM, e ARTEMIS con l'aiuto dei due tool esterni già citati. La determinazione del thesaurus si ottiene attraverso cinque passi:

1. estrazione delle relazioni intra-schema;
2. estrazione delle relazioni inter-schema;
3. arricchimento dell'insieme delle relazioni;
4. validazione delle relazioni;
5. inferenza di nuove relazioni.

5.5.1 Estrazione delle relazioni intra-schema

Una volta che il sistema ha acquisito gli schemi delle diverse sorgenti, può estrarne le gerarchie di aggregazione ed ereditarietà, in particolare quelle definite tramite foreign key. Per una sorgente relazionale, se la foreign key è chiave primaria sia per la classe di partenza che per quella referenziata, il modulo SIM trasforma la relazione tra le due classi in una relazione terminologica di tipo BT/NT, in caso contrario la caratterizza come una generica RT. Per una sorgente a oggetti viene invece generata una relazione BT/NT per le eredità tra classi, una relazione RT per le gerarchie di aggregazione.

5.5.2 Estrazione delle relazioni inter-schema

Il secondo processo di estrazione di relazioni è quello realizzato dall'analisi e dal confronto di tutti gli schemi ODL_{I3} che portano all'identificazione di relazioni lessicali tra i nomi delle classi e degli attributi usati nei diversi schemi.

In questa fase interviene il già citato database lessicale che permette di assegnare significati appropriati alle forme base. Il compito più delicato ricade comunque sul progettista che deve cercare di specificare il maggior numero di significati, selezionandoli, però, in modo non equivoco, per evitare la generazione di relazioni errate.

Le relazioni estratte da WordNet vengono poi convertite in relazioni tipiche del Thesaurus Comune:

- la sinonimia diventa una SYN;
- l'iperonimia una BT;
- l'iponimia una NT;
- l'olonimia e la correlazione vengono invece tradotte in relazioni RT.

5.5.3 Arricchimento dell'insieme delle relazioni

In questa fase è il progettista stesso che inserisce nuove relazioni, non ricavate precedentemente e arricchisce il Thesaurus sulla base della sua personale conoscenza. Il

passaggio è importante perché le relazioni che vengono aggiunte devono essere corrette per evitare di generare uno schema errato.

5.5.4 Validazione delle relazioni

Una volta estratte le relazioni intra e inter-schema, queste devono essere validate per evitare, in fase di integrazione, l'uso di relazioni errate.

Questo compito come già accennato viene affidato al sistema ODB tools che analizza la compatibilità tra i domini degli attributi che partecipano a una relazione, e nel caso questi verificano i criteri di validità, le valida e li inserisce a tutti gli effetti nel Thesaurus Comune.

I criteri di validità sono fondamentalmente tre:

1. una relazione SYN è valida se i domini dei due attributi che vi prendono parte sono equivalenti, oppure se sono l'uno la specializzazione dell'altro;
2. una relazione BT è corretta se il dominio del primo attributo contiene o è equivalente a quello del secondo;
3. quando il dominio di un attributo è definito utilizzando il costruttore *union* la relazione che coinvolge quell'attributo è corretta se i criteri già citati sono rispettati da almeno uno dei suoi domini.

5.5.5 Inferenza di nuove relazioni

Anche questa parte del processo di generazione del Thesaurus Comune, come la precedente, ricade su ODB tools che con tecniche di sussunzione derivate dall'intelligenza artificiale inferisce nuove relazioni.

Il sistema compie un'operazione di scomposizione e riorganizzazione degli schemi locali per individuare nuove relazioni tra le varie classi che vengono collegate in questa struttura provvisoria per mezzo delle relazioni già esistenti:

- ogni BT/NT dà luogo a una gerarchia di ereditarietà;
- una SYN genera una doppia gerarchia di ereditarietà;
- una RT produce un'aggregazione.

Sulla base di queste informazioni viene creata una struttura di collegamenti tra le classi

dalla quale ODB tools desume nuove relazioni di aggregazione ed ereditarietà, concludendo il processo di arricchimento del Thesaurus Comune.

5.6 Il calcolo delle affinità

Una volta conclusa la formazione del Thesaurus Comune, per proseguire nell'integrazione è necessario calcolare l'affinità tra le classi locali per poterle raggruppare nei cluster che porteranno alla formazione degli schemi globali.

Questo calcolo viene realizzato, dal modulo ARTEMIS, con l'uso di due parametri:

1. Structural Affinity Coefficient che viene determinato tra due classi, in base alle relazione tra i loro attributi;
2. Name Affinity Coefficient che si basa sulle relazione che legano le coppie di classi.

La combinazione di questi due coefficienti produce il Global Affinity Coefficient, il vero elemento di confronto per stabilire la similarità tra due classi. Per un raffronto numerico delle affinità tra i termini viene assegnato un peso σ a ogni relazione, peso che sarà tanto maggiore quanto più stringente sarà il legame imposto dalla relazione ai due termini considerati:

$$\sigma_{syn} \geq \sigma_{nt/bt} \geq \sigma_{rt}$$

Tipicamente vengono considerati i seguenti valori:

$$\sigma_{syn} = 1$$

$$\sigma_{nt/bt} = 0.8$$

$$\sigma_{rt} = 0.5$$

5.7 La generazione dei cluster

Calcolato il grado di affinità tra le classi, tutte quelle che hanno un'affinità superiore a una soglia prestabilita vengono riunite in gruppo dal mediatore, secondo tecniche di clustering. La procedura di clustering è iterativa e inizia allocando un cluster per ogni classe, quindi ad ogni successiva iterazione fonde i cluster delle due classi che hanno il

Global Affinity Coefficient massimo. La procedura ha termine quando tutte le classi appartengono a un unico cluster.

Il risultato più interessante di questo processo non è però il cluster finale, ma l'albero che viene generato, in cui le foglie sono le classi da raggruppare.

In quest'albero ogni nodo rappresenta un livello di clusterizzazione ed è caratterizzato da un coefficiente di affinità tra i due sottoalberi che collega. Come detto, la formazione dei cluster avviene confrontando i coefficienti di affinità dei nodi con un valore soglia, tipicamente viene usato 0,5; tutte le classi dei sottoalberi che hanno un grado di affinità superiore alla soglia vengono fuse in un unico cluster, ossia in un'unica classe globale.

5.8 La costruzione delle classi globali

Una volta che sono stati realizzati i cluster in cui sono state raccolte le classi locali, si passa alla fase di integrazione vera e propria degli schemi. Per ogni cluster realizzato si crea una classe globale caratterizzata da:

- un nome che fa da identificatore;
- un insieme di attributi;
- una *mapping table* che gestisce la corrispondenza tra gli attributi globali e gli attributi degli schemi locali.

Dopo l'implementazione delle classi globali, il processo di integrazione prosegue con altre due fasi: la fusione degli attributi e la creazione della *mapping table*. Poiché le regole di mapping definite nella seconda fase dipendono dalle fusioni operate nella prima, il sistema MOMIS ne tiene traccia operando i due passi finali contemporaneamente.

Per comprenderne meglio il funzionamento verranno, però, analizzati separatamente.

5.9 Fusione degli attributi

La fusione degli attributi nasce dall'esigenza di eliminare le ridondanze e di integrare completamente gli schemi. La modalità di fusione degli attributi all'interno di una

classe globale è dipendente dal tipo di relazioni che li legano e dal fatto che queste siano o meno validate.

5.9.1 Fusione degli attributi di relazioni validate

Come già detto, i modi di fondere gli attributi sono funzione dei tipi di relazione che li legano, vediamoli quindi caso per caso.

Consideriamo due attributi di classi locali differenti, legati da una relazione di sinonimia (SYN), questo significa che esprimono lo stesso significato e quindi possono essere fusi. A questo punto del processo il mediatore analizza i domini dei diversi attributi: se coincidono li fonde nell'attributo globale il cui dominio contiene quello di tutti gli altri; se sono diversi, propone un elenco di domini, che generalizzano quelli degli attributi locali, dal quale il progettista dovrà scegliere quello da utilizzare.

Per quanto riguarda gli attributi legati da relazioni di specializzazione BT/NT, il sistema procede fondendo l'attributo che sta più in basso, nella gerarchia di generalizzazione, in quello che sta più in alto. Il comportamento riguardo ai domini è il medesimo descritto per le relazioni di sinonimia. Le relazioni di associazione RT non vengono considerate in questa fase, poiché non esprimono legami forti gestibili dal sistema in modo automatico, il loro utilizzo, per la fusione degli attributi, viene lasciato al progettista.

Un caso particolare di fusione si verifica quando attributi legati da relazioni di sinonimia, compaiono anche in relazioni di specializzazione.

Il sistema è realizzato in modo tale da risolvere comunque questo problema in automatico, dando priorità alle relazioni più stringenti, quindi fondendo prima gli attributi legati da SYN. Ovviamente al termine del processo automatico il progettista può intervenire e modificare lo stato generato dal sistema.

5.9.2 Fusione di attributi di relazioni non validate

Per gli attributi legati da relazioni non validate il sistema non riesce a identificarne la compatibilità e quindi non può operarne la fusione in modo automatico. Questo caso ricade quindi sulla conoscenza del progettista che si farà carico delle fusioni e delle scelte dei domini degli attributi globali creati.

5.10 La generazione delle mapping table

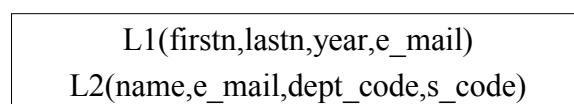
Le *mapping table* sono le tabelle degli schemi globali che conservano le informazioni necessarie per passare dagli attributi globali a quelli locali, cioè ai dati veri e propri. Sono evidentemente uno strumento fondamentale soprattutto in fase di query processing.

Sono ammesse le seguenti tipologie di corrispondenze tra attributi:

1. **corrispondenza semplice**: l'attributo globale assume banalmente il valore dell'attributo locale;
2. **corrispondenza in AND**: questo caso si può avere quando due attributi di una stessa classe locale sono stati fusi insieme, quindi si avrà che l'attributo globale assumerà i valori concatenati degli attributi locali considerati;
3. **corrispondenza in UNION**: è analoga alla precedente con la differenza che i valori degli attributi locali sono in alternativa, cioè l'attributo globale assumerà solo il valore di uno o dell'altro attributo locale;
4. **valore di default**: è una costante assegnata dal progettista per valorizzare informazioni presenti negli schemi sottoforma di metadato, ma che non sono state prese in considerazione dal sistema;
5. **valore NULL**: serve chiaramente per identificare la mancanza di corrispondenza tra un attributo globale e gli attributi di un determinato schema locale.

Terminiamo questo paragrafo con un semplice esempio di integrazione per chiarire il concetto di *mapping table* su MOMIS, diverso da quello di mapping tra schemi che sarà presentato nel capitolo [6]; vari esempi di integrazione, relativi al problema della classificazione dei prodotti, saranno illustrati nei capitoli seguenti.

Classi Locali (relazionali)



INTEGRAZIONE

Unica Classe Globale con Attributi :

Name,E_mail,Year,Dept,Section

Sia data una classe globale G con m attributi globali (A1, A2, ...Am) e rappresentativa di n classi locali (L1, L2, ... Ln).

La *mapping table* associata a G (Figura 23) è una tabella con n righe e m colonne, nella quale l'elemento (i,j) rappresenta in che modo l'attributo globale Ai è mappato nella classe locale Lj.

Mapping Table di G:

	Name	E_mail	Section	Year	Dept
L1	firstn and lastn	e_mail	null	year	null
L2	name	e_mail	s_code	null	dept_code

Figura 23 Esempio di Mapping Table

5.11 SI-Designer: il tool per la creazione dello schema globale

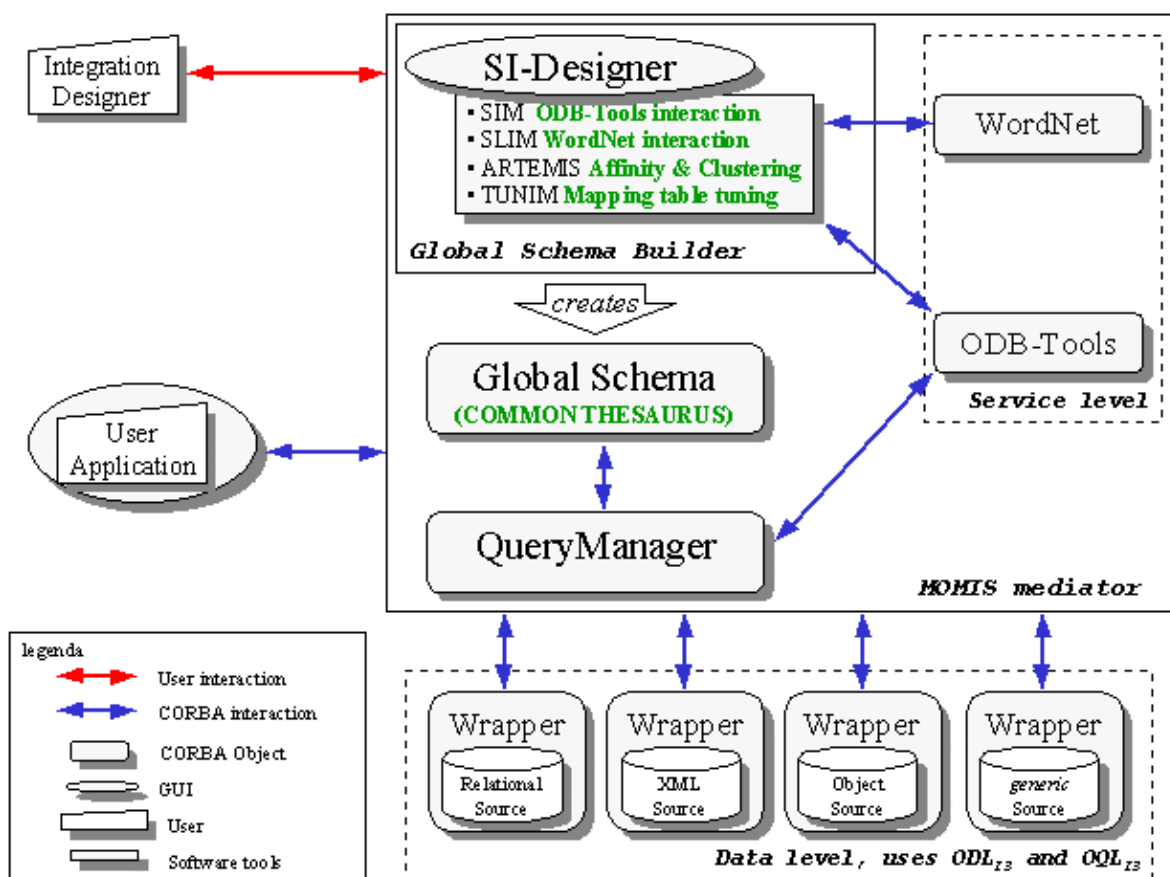


Figura 24 Architettura di SI-Designer

SI-Designer [Beneventano, 00] è l'Interfaccia Utente Grafica (GUI) del sistema MOMIS; è a questa che accede il progettista per l'integrazione intelligente delle sorgenti e la costruzione delle classi globali.

Come si vede in Figura 24 SI-Designer è un contenitore per gli altri moduli, necessari alla costruzione dello schema globale:

- SIM (Source Integrator Module)
- SLIM (Sources Lexical Integrator Module)
- ARTEMIS (Analysis and Reconciliation Tool Environment for Multiple Information Sources)
- TUNIM [Guidetti, 00] (TUNIng of the Mapping table)

La raffinazione delle informazioni, fatta nel passaggio attraverso questi moduli porta all'integrazione finale delle sorgenti locali. Vediamo quindi come si usa SI-Designer.

5.11.1 L'automa di utilizzo di SI-Designer

Prima di entrare nel dettaglio dei vari elementi presenti nella GUI di MOMIS vediamo come si possono susseguire le fasi di utilizzo dei vari moduli. Si inizia col pannello *Source* nel quale si caricano i wrapper per acquisire le informazioni sulle sorgenti, quindi si passa al pannello *Thes.Rel* in cui il modulo SIMA estrae le relazioni intra-schema. Fissate queste prime relazioni si passa al pannello di SLIM che per mezzo del modulo WordNet consente la costruzione di nuove relazioni lessicali tra i termini che caratterizzano classi e attributi.

Si torna quindi al pannello *Thes.rel* dove il progettista può intervenire aggiungendo o rimuovendo delle relazioni. Al termine di queste operazioni il progettista richiama SIMB che per mezzo di ODB tools inferisce nuove relazioni e controlla e valida le relazioni di tutto il Thesaurus. Validate le operazioni si passa al pannello Cluster dove per mezzo del modulo ARTEMIS le classi locali vengono raggruppate in cluster; cluster che determineranno le *mapping table*. Nel pannello TUNIM si modificano le *mapping table* fondendo gli attributi per rimuovere le ridondanze. Completata questa operazione si giunge alla formazione dello schema globale. Come si può vedere dalle linee

tratteggiate dell'automa è permesso al progettista di tornare indietro, nel processo di integrazione, per operare delle revisioni e delle correzioni:

- può ritornare a SLIM per modificare il significato o la forma base di un termine per il quale erano state generate relazioni non corrette;
- può tornare a *Thes.Rel* e inserire relazioni che si accorge, a posteriori, essere state trascurate;
- è possibile fare un passo indietro nel pannello cluster per modificare i parametri di ARTEMIS poiché erano stati generati cluster non corretti o poco soddisfacenti.

5.11.2 Source

Il primo passo da compiere nell'integrazione è, come abbiamo visto nell'automa, il caricamento dei wrapper e quindi delle informazioni sulle sorgenti.

La prima finestra che ci presenta SI-Designer è quindi il pannello *Source*, che possiamo vedere in Figura 25.

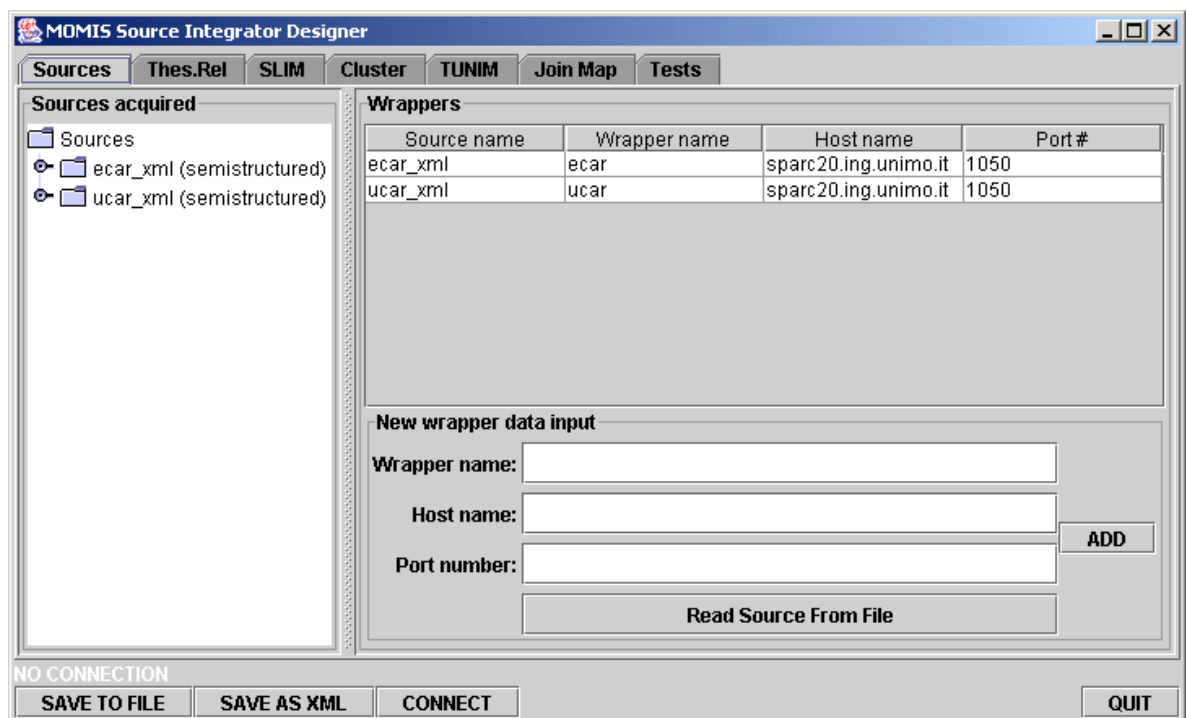


Figura 25 Pannello Source

La parte in basso di questo pannello, denominata *New wrapper data input*, è quella necessaria al caricamento dei wrapper, dove bisogna inserire:

- il nome del wrapper;
- la macchina su cui il wrapper è installato;
- il numero della porta attraverso cui vi si può accedere.

Inseriti i dati e cliccando sul tasto ADD si carica il wrapper.

Questa operazione produce la comparsa dei dati relativi al wrapper e alla sorgente a cui si interfaccia nella finestra di *label Wrappers* e la visualizzazione della struttura della sorgente, con la specificazione del tipo (relazionale, a oggetti, semistrutturata), nella finestra di destra denominata *Source acquired*.

5.11.3 SIM A

Una volta caricate le sorgenti si prosegue nel processo di integrazione con la formazione del Thesaurus Comune il cui primo passo consiste nell'estrazione delle relazioni intra-schema.

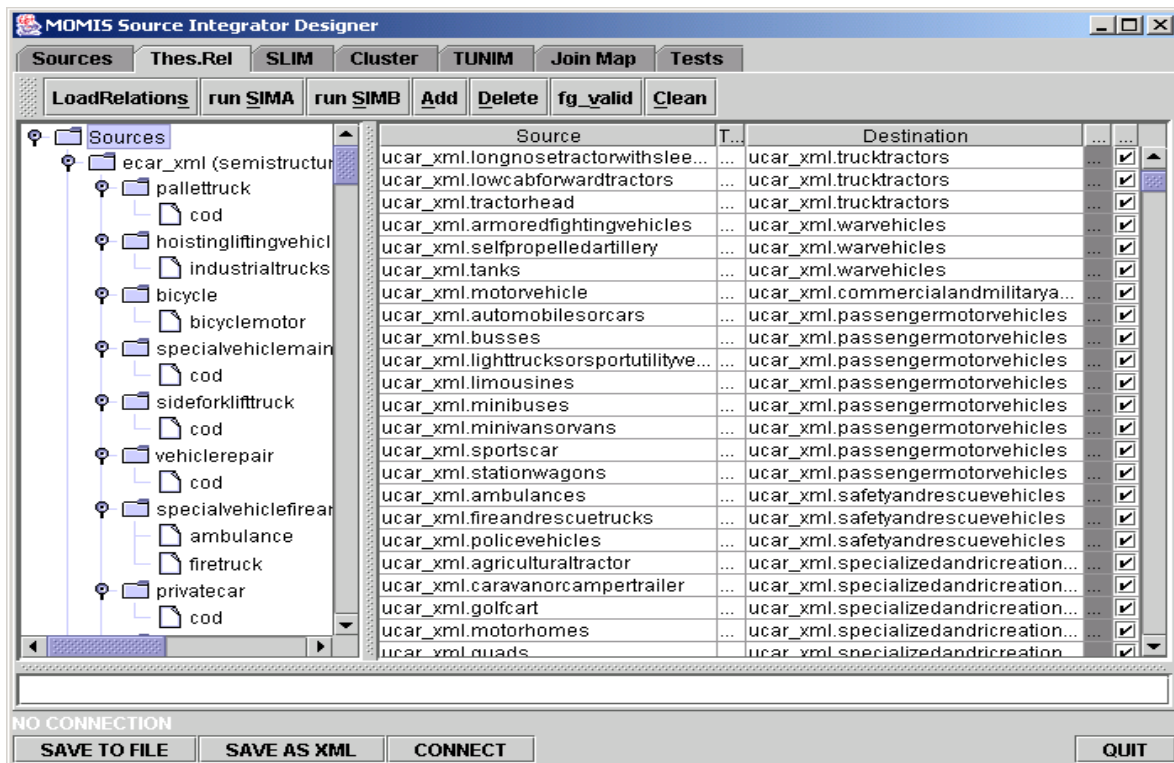


Figura 26 Pannello Thes.Rel: run SIM A

Per fare questo si accede al pannello *Thes.Rel*, Figura 26, e si seleziona il pulsante run SIM A. Il risultato di questa operazione sarà la comparsa nella finestra di destra delle relazioni tra le classi e gli attributi, di una stessa sorgente, dovute alla presenza delle

foreign key negli schemi.

Nella finestra di destra si possono osservare i due termini che partecipano alla relazione: *Source*, *Destination* separati dal tipo di relazione che li lega: SYN, BT,NT RT. Questi elementi sono seguiti da una colonna in cui viene identificato, per mezzo di un colore e di un codice, chi ha costruito la relazione. Per terminare sulla destra si può osservare una colonna che segnala se la relazione è stata o meno validata.

5.11.4 SLIM

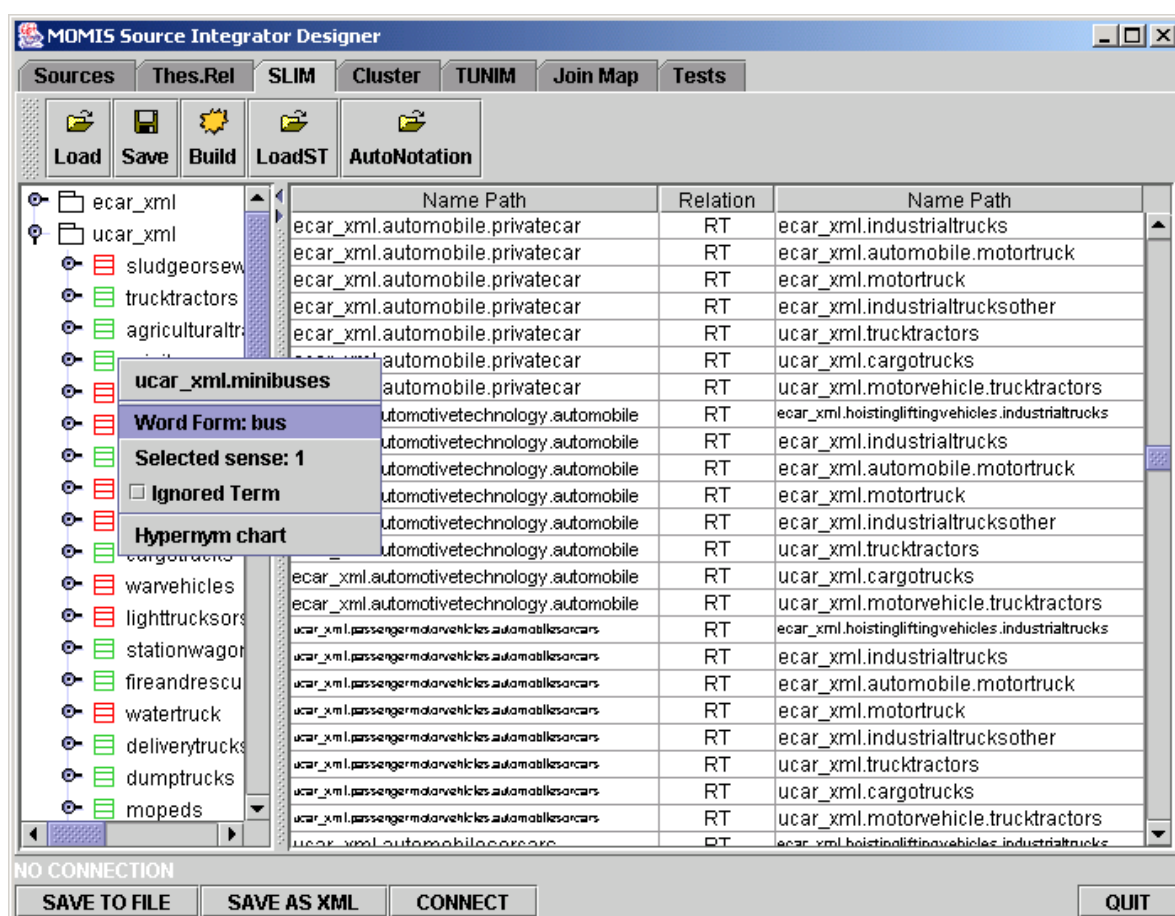


Figura 27 Modulo SLIM

L'operazione successiva all'estrazione delle relazioni intra-schema consiste nella costruzione delle relazioni lessicali tra i termini degli attributi e delle classi dei diversi schemi, questo si compie nel pannello *SLIM* (Figura 27).

Vediamo quindi come si fa ad assegnare un significato a un termine:

- per prima cosa si seleziona col tasto sinistro del mouse il termine che

interessa; quindi vi si clicca sopra col tasto destro, questo provoca la comparsa di una tendina, Figura 27, con i seguenti elementi che conducono alla selezione del significato del termine:

- word form: selezionando questa voce si apre una finestra in cui si può inserire la forma base per il termine considerato;
 - select sense: questo elemento causa l'apertura di una finestra contenente i significati di WordNet relativi alla forma base scelta per quel termine. In questa finestra è possibile selezionare il significato o i significati più adeguati;
 - ignored term: questa voce permette di obbligare SLIM a non considerare il termine. Questa operazione può essere necessaria, per esempio, quando la forma base che si assegna al termine avrebbe un significato eccessivamente generale che produrrebbe la formazione di molte relazioni lessicali, la maggior parte delle quali sarebbe sbagliata;
 - Hypernym: selezionando questo elemento il progettista può scegliere il significato del termine selezionando l'ambito a cui il significato deve appartenere e quindi la specificazione più corretta in questo ambito;
- BUILD: quando per tutti i termini è stata compiuta questa operazione si clicca sul tasto BUILD che spinge SLIM alla costruzione di tutte le relazioni lessicali che riesce a desumere in base ai significati scelti per i vari termini.

Poiché le sorgenti hanno spesso un lungo elenco di termini e quindi occorre molto tempo per lo svolgimento di questa operazione sono stati rivisti due tasti, LOAD e SAVE, per caricare e salvare i significati già assegnati ai termini. In alternativa al pannello SLIM si può utilizzare il pannello ARM, che svolge le medesime funzioni, solamente con un'interfaccia a pannelli invece che a menù.

5.11.5 SIM B

Costruite le relazioni dal modulo lessicale si ritorna al pannello *Thes.Rel*, qui il progettista può inserire nuove relazioni frutto della sua personale conoscenza, per mezzo del tasto ADD che fa comparire una finestra in cui inserire il termine *source*, quello *destination* e la relazione che li lega.

Inoltre per mezzo del tasto DELETE il progettista può rimuovere delle relazioni, che ritiene errate, dal Thesaurus. Una volta concluse queste operazioni di arricchimento, modifica e correzione delle relazioni, il progettista preme il tasto *run SIMB* che fa compiere al sistema il processo di inferenza di nuove relazioni e il controllo e la validazione di tutte le relazioni contenute nel Thesaurus (Figura 28).

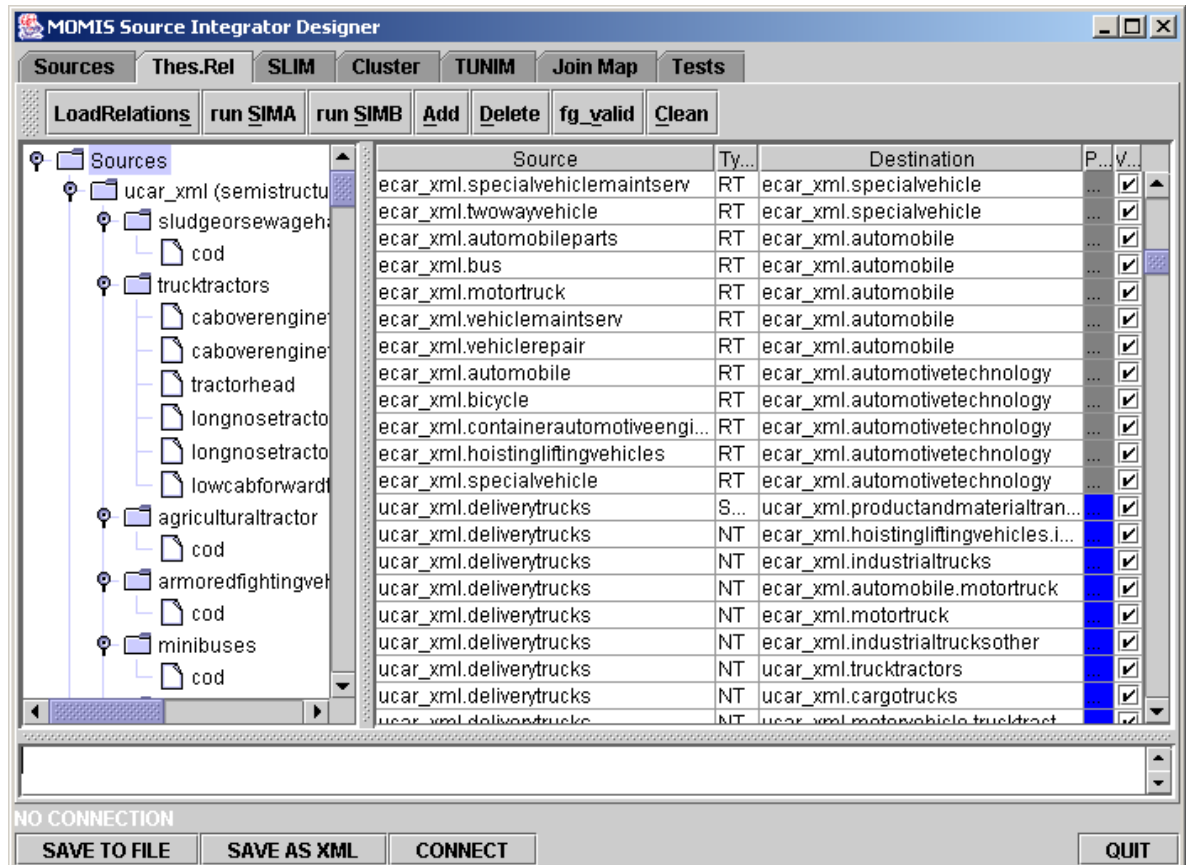


Figura 28 Modulo SIM B

5.11.6 Cluster

Il pannello Cluster è l'interfaccia grafica del modulo ARTEMIS e serve, come già detto, alla costruzione dei cluster, nei quali vengono raggruppate le classi locali dei vari schemi che devono essere integrati. Questo pannello presenta due sheets:

- *Cluster Tuning*;
- *ARTEMIS Configuration*.

Nel pannello *ARTEMIS Configuration*, si configura il modulo ARTEMIS settando:

- i pesi delle relazioni del Thesaurus;

- i valori di soglia per le affinità e il clustering;
- i coefficienti per l'affinità globale.

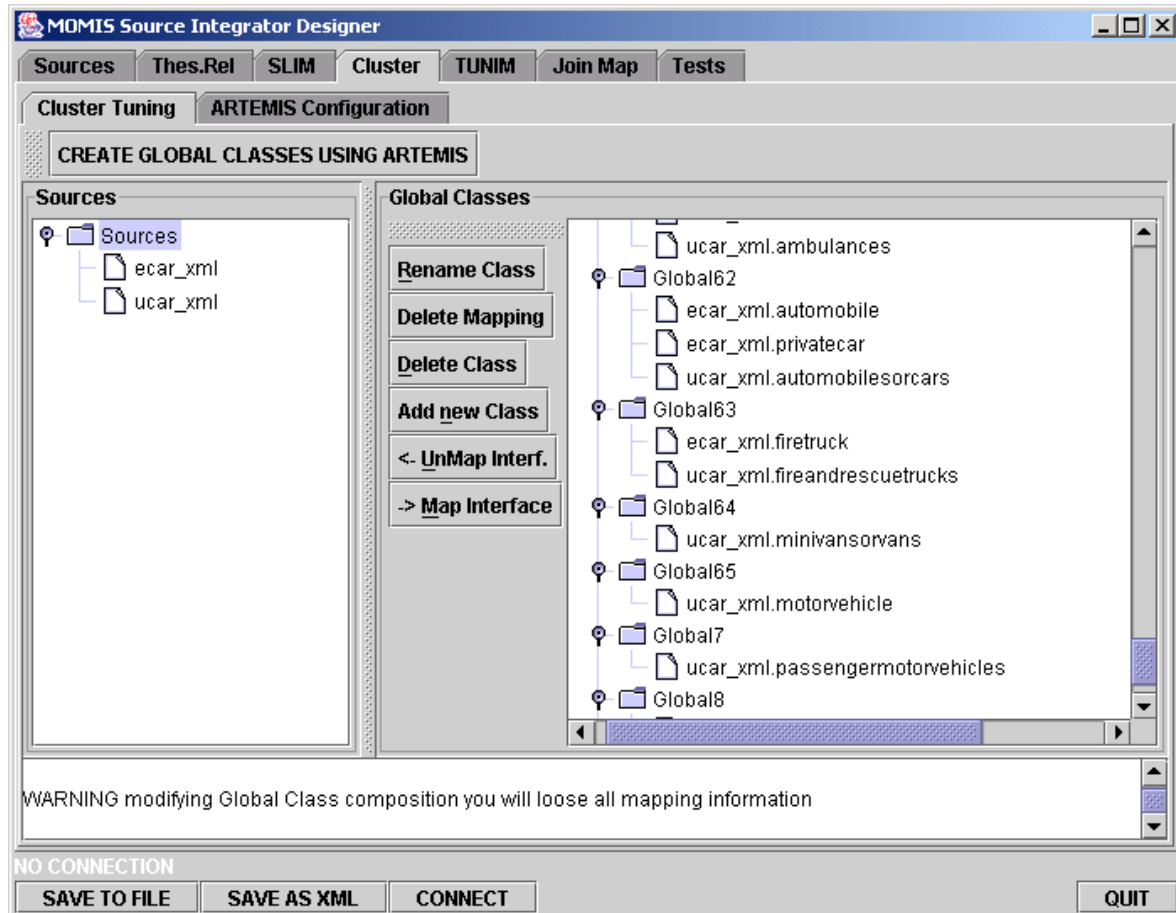


Figura 29 Pannello Cluster

Una volta settate le variabili di ARTEMIS si procede premendo il pulsante CREATE GLOBAL CLASSES che fa eseguire dal sistema il modulo ARTEMIS e fornisce nella finestra in basso i cluster costruiti secondo le specifiche. Se non si vogliono modificare i valori standard dei parametri di ARTEMIS si può eseguire la costruzione dei cluster anche dal pannello Cluster Tuning (Figura 29) invocando il modulo per mezzo del pulsante CREATE GLOBAL CLASSES USING ARTEMIS.

Il risultato viene presentato nella finestra di sinistra del pannello di Figura 29. Su questo risultato il progettista può intervenire con i pulsanti che si trovano al centro, per una modifica manuale.

Selezionando una classe e premendo il pulsante *Rename Class* è possibile cambiare il nome a una classe globale; questa è un'operazione consueta poiché i cluster creati non

hanno nomi significativi. Con il pulsante *Delete Mapping* è possibile cancellare la mappa costruita.

I pulsanti *Delete Class* e *Add new Class* servono per rimuovere o aggiungere una classe globale. Con i pulsanti *UnMap Interf.* e *Map Interface* il progettista può rimuovere o mappare un attributo in una classe globale selezionata.

5.11.7 TUNIM

Il pannello TUNIM, che possiamo osservare in Figura 30, è quello che permette la costruzione degli schemi globali dopo che le classi locali sono state raggruppate in cluster dal modulo ARTEMIS.

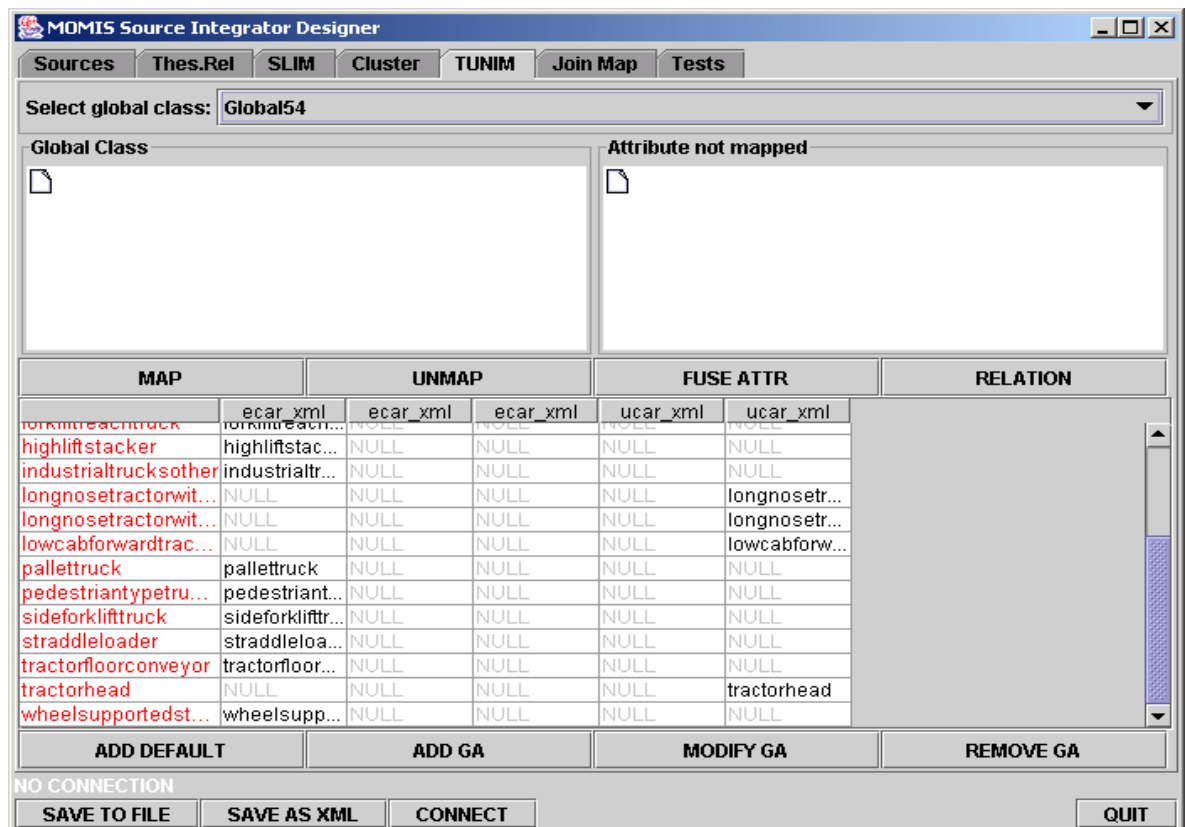


Figura 30 Modulo TUNIM

La prima operazione da compiere è premere il tasto MAP ALL che provoca la creazione di un attributo globale per ogni attributo locale, nelle rispettive classi globali determinate dai cluster. Successivamente per mezzo del tasto FUSE ATTR si compie la

fusione degli attributi locali uguali in un unico attributo globale rimuovendo così le ridondanze.

Terminate queste operazioni automatiche il progettista può intervenire manualmente per modificare lo schema, da qui la presenza di altri pulsanti nel pannello:

- *map*: produce la mappatura dell'elemento scelto nella finestra *Attribute not mapped*, nell'attributo globale selezionato;
- *unmap*: rimuove un attributo locale dalla mappatura e lo pone nella finestra *Attribute not mapped*;
- *relation*: crea una finestra in cui visualizza le relazioni a cui partecipa l'attributo locale selezionato tra quelli non mappati;
- *add ga*: permette l'inserimento di un nuovo attributo globale, per il quale il progettista dovrà scegliere nome e dominio;
- *add default*: consente di assegnare un valore di default a un determinato attributo locale;
- *remove ga*: rimuove l'attributo globale selezionato dalla mapping table. Se l'attributo non è vuoto, ma in esso vi sono mappati degli attributi locali, questi vengono spostati nella finestra *Attribute not mapped*;
- *modify ga*: consente di modificare nome e domini di un attributo globale.

Capitolo 6

Integrazione di standard di classificazione con il sistema MOMIS

Nella prima parte di questa trattazione si è contestualizzato l'ambito del progetto, definendo da un lato le principali caratteristiche dell'e-commerce con le problematiche correlate, dall'altro illustrando lo stato dell'arte della ricerca e le possibili risoluzioni teoriche necessarie all'esplosione definitiva del B2B.

Il commercio elettronico soffre la mancanza (o la sovrabbondanza) di standard per la catalogazione dei prodotti scambiati sui marketplace elettronici; gli studi relativi a Ontologia e Web Semantico propongono modelli e schematizzazioni della realtà utili a formalizzare il problema, in particolare indicano l'integrazione tra gli standard di classificazione dei prodotti esistenti come risposta da preferire rispetto alla creazione di un nuovo standard unico e universalmente riconosciuto.

Questa tesi si colloca all'interno di un progetto che vuole implementare l'integrazione tra i codici di prodotto con il sistema MOMIS, per realizzare un mapping tra due standard di classificazione, come proposto dal Challenge lanciato dall'équipe di OntoWeb²⁹ che descriverò in seguito.

²⁹ OntoWeb: progetto UE di Network tematico per lo scambio di informazioni basate sull'ontologia in ambito E-Commerce e Knowledge Management
www.ontoweb.org

6.1 Il Challenge di OntoWeb

Nel numero speciale di IEEE Intelligent Systems Magazine, dedicato all'Intelligent E-Business (07-08/2001) il gruppo di ricerca di OntoWeb (in particolare Heloise Ontology Associates) ha lanciato un challenge internazionale nel campo dell'intelligent e-business con l'obiettivo di trovare un modello generale e delle soluzioni operative capaci di mappare in modo (semi-)automatico una descrizione di prodotto data da due diversi standard di classificazione prodotto [Schulten, 01].

6.1.1 Ipotesi metodologica

Alla ricerca di una Lingua Franca o della Stele di Rosetta

Per iniziare un progetto di ricerca nell'ambito dell'ITC occorre fare una premessa fondamentale:

le interazioni umane che comportano consumo di tempo e fatica possono essere rimpiazzate da processi (semi-)automatici basati su Internet;

E' evidente che quest'affermazione diventa rilevante perché individua nella ricerca non solo un'appassionante sfida dell'intelligenza umana, ma comporta importanti implicazioni economiche che possono muoverla e incentivarla.

Nel guardare le attuali implementazioni vediamo sia applicazioni semplici del tipo ricerca prodotti, sia soluzioni più sofisticate quali comunicazione tra server per la ricostituzione delle scorte.

Tuttavia l'e-business si scontra con la lentezza che varie imprese mostrano nell'adattarsi alle transazioni economiche e guarda alla complessità di sostituzione delle interazioni umane con i computer. Nel linguaggio umano i dialoghi sono strutturati dalle regole della grammatica, della semantica e della sintassi, inserite in un contesto di convenzioni sociali e culturali.

Il nuovo mondo dell'e-commerce manca ancora di un background di comunicazione ricco ed è lontano dal conseguimento della visione del Web come definito dal W3C, ("Universe of Network Accessibile Information"). Il protocollo esistente e gli standard di ipertesto appaiono insufficienti per lo scambio di informazioni non ambigue tra reti,

ma questa, come ribadito più volte, sarebbe la condizione fondamentale per il commercio elettronico. In particolare lo è nel business tra imprese, dove vi sono problemi di interoperabilità tra differenti sistemi di integrazione di sistemi back-end, di processi di business complessi con regole stringenti.

Per la mancanza di standard i soggetti economici popolano il Web in numerose comunità, intranet, extranet, marketplace; di conseguenza sono nati vari linguaggi.

Il multi-linguismo non è un problema in sé, al contrario, spesso dà luogo a soluzioni creative; ma nel caso del B2B e-commerce diventa un ostacolo alla comunicazione tra le imprese.

Nel contesto B2B ritorna spesso una metafora: industria e comunità scientifica devono unire i loro sforzi in iniziative di standardizzazione, come nell'episodio biblico della Torre di Babele ogni individuo portava il proprio blocco.

Varie iniziative quali nuovi modelli di Lingua Franca (spesso usata per riferirsi all'XML) o di Stele di Rosetta, da cui deriva lo standard RosettaNet, costellano la letteratura.

Lingua Franca era il linguaggio commerciale usato nel 18^{esimo} secolo nell'area del Mediterraneo per rendere possibile il commercio internazionale. Stele di Rosetta è la pietra basaltica che portava incisi alcuni messaggi in tre lingue, rivelatesi codici utili per la traduzione dei geroglifici.

Ho ripreso questi due esempi non per interesse storico, ma perché indicano due diverse strade percorribili per raggiungere l'interoperabilità semantica: Lingua Franca rappresenta la ricerca di un nuovo standard universale, Stele di Rosetta l'obiettivo di integrare standard esistenti.

L'esigenza di un modello condiviso nella comunità economica cresce a vari livelli e si riflette su differenti progetti di sviluppo di queste iniziative di armonizzazione su cataloghi di prodotto o business document. Ma ci sono poche iniziative che mirano all'armonizzazione, apparentemente più semplice, delle unità elementari di una transazione commerciale: le descrizioni dei prodotti.

A livello base vi sono due funzioni associabili con lo standard di descrizione dei prodotti (content standard):

1. A partire da un set di richieste, il cliente ha bisogno di visualizzare una completa gamma di un prodotto;
2. Il cliente ha bisogno della descrizione di un prodotto precisa per un'applicazione specifica.

Per rispondere a queste esigenze fondamentali OntoWeb suggerisce di concentrare lo studio dei sistemi di classificazione sulla standardizzazione delle descrizioni di prodotto.

Idealmente si dovrebbe avere un sistema di classificazione dei prodotti unico e assoluto, ma questo non è possibile per almeno due ragioni. Prima di tutto perché il ciclo di vita dei prodotti è breve; secondariamente perché le imprese non possono smettere di crescere e innovare la loro gamma di prodotti a causa di uno standard globale.

Ma le imprese hanno già popolato il Web di troppe soluzioni per sistemi di classificazione (sovrrabbondanti), a volte ispirate da principi di progettazione, altre volte nate per rispondere a bisogni industriali, altre ancora basate sulle convenzioni della “old economy”, ...; in generale si sono rivelate inopportune e poco *user-friendly*.

Pertanto è giunto il momento, al fine di trovare risposte alla domanda di sviluppo e adozione di soluzioni e-commerce, di prendere in considerazione gli standard di prodotto emergenti e armonizzarli; da questa esigenza ha preso vita l'idea del Challenge per l'integrazione dei content standard.

6.1.2 Mapping tra schemi

In [Bernstein, 01] l'operazione di manipolazione delle informazioni contenute in schemi viene detta “match” e consiste nel prendere due schemi come input per produrre in output un mapping tra elementi che corrispondono semanticamente l'uno all'altro (nel seguito della trattazione utilizzerò il lessico “mapping” o “matching” in modo indifferenziato).

Il matching tra schemi gioca un ruolo centrale in numerose applicazioni, quali integrazione di dati *Web-oriented*, e-commerce, integrazione di schemi, sviluppo di schemi e applicazioni, data warehousing, progetto di database, creazione e gestione di siti Web.

La realizzazione manuale del mapping tra schemi presenta significative limitazioni:

- Dispendio di tempo;
- Facilità di errore;
- Processi costosi.

Questi fattori sono un problema crescente, visto il rapido incremento di sorgenti Web e E-business da integrare; inoltre più aumenta l'abilità dei sistemi a gestire database e

applicazioni complessi, più gli schemi diventano ampi e cresce il numero di match da realizzare.

Pertanto diventa necessario un approccio all'integrazione più veloce e meno labor-intensive che richiede un supporto automatico al mapping tra schemi.

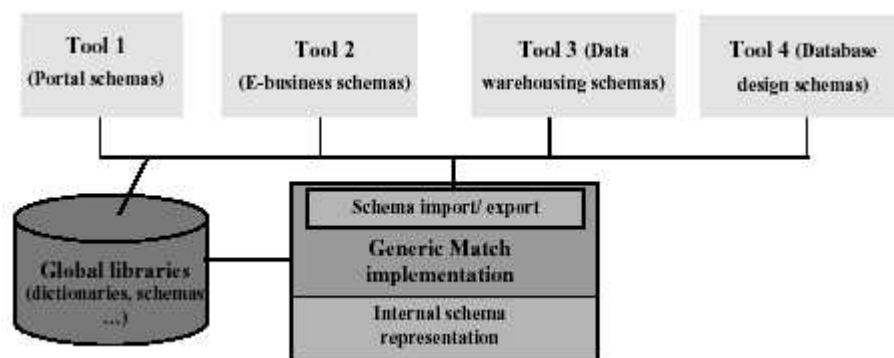


Figura 31 Architettura di un generico mapping

6.1.2.1 Un dominio di applicazione del mapping: l'e-commerce

Nel decennio scorso l'e-commerce ha lasciato una precisa motivazione a sostegno del matching tra schemi: la transazione di documenti. I partner commerciali frequentemente scambiano dati che contengono transazioni economiche e, come detto ripetutamente, ogni attore utilizza il proprio formato. Il formato può variare per sintassi, come EDI e XML, o per struttura dati dei clienti. Quindi per permettere ai sistemi di scambiare informazioni, occorre sviluppare un processo che converta i documenti in un formato condiviso e di conseguenza bisogna prima di tutto fare un mapping tra gli schemi che possono presentare nomi e tipi di dato diversi, valori ammessi e range incongruenti.

6.1.2.2 Schematizzazione del mapping

Come sostegno al lavoro del progettista che manualmente o in modo (semi-)automatico deve realizzare un mapping tra sorgenti di dati eterogenee è utile formalizzare una rappresentazione grafica degli schemi.

Comunemente si utilizza la rappresentazione ad albero perché visualizza la gerarchia tra le informazioni contenute.

Nel caso particolare dei content standard (oggetto della mia tesi) si indica con la radice (root) il primo livello del codice e si scende attraverso i rami fino alle foglie che individuano i beni indifferenziati (commodity).

6.1.3 Come arrivare a un'ontologia comune: *case study* sul dominio “Writing Paper”

I gruppi di ricerca che partecipano al Challenge do OntoWeb hanno il compito di progettare un modello (ontologia, metodo problem-solving, soluzione implementata) per arrivare a un modo di descrivere la conoscenza del prodotto e di stabilire un mapping semi-automatico tra i diversi sistemi di classificazione prodotto.

Il risultato deve rispettare alcune specifiche, deve essere:

- Efficace in modo computazionale;
- Utilizzabile nella pratica;
- Fondato dal punto di vista teorico.

Questo lavoro può essere interessante sia dal punto di vista dei ricercatori, sia da quello delle industrie.

In [Schulten, 01] OntoWeb spiega i termini del Challenge attraverso un case study sul dominio di prodotto “writing paper” classificato da due diversi sistemi di classificazione prodotti per l'e-commerce. L'utilità dell'esempio è quella di evidenziare gli aspetti strategici che sono emersi da lavori precedenti come input per la modellazione futura e di puntualizzare lo stato dell'arte su sistemi intelligenti, ingegneria della conoscenza, ontologia e Web semantico.

Inoltre circoli culturali, librerie di settore e professionisti di musei e esposizioni hanno già effettuato una lunga pratica nel campo della catalogazione digitale; risulta interessante vedere come il loro contributo possa essere recepito nell'area del commercio elettronico.

Un ultimo compito, forse il più rilevante, che il challenge vuole portare avanti è di contribuire a superare il gap che divide mondo scientifico e industriale.

6.1.3.1 Descrizione dell'esempio di OntoWeb

Immaginiamo che un venditore o acquirente che ha optato per uno schema di classificazione UNSPSC non riesca a comunicare con un venditore o acquirente che ha scelto una codifica concorrente a UNSPSC, ad esempio ecl@ss e che il dominio di prodotto in questione sia “writing paper”.

Si riportano di seguito i rispettivi frammenti di codice relativi a “writing paper” così come illustrati sull'articolo di OntoWeb.

UNSPSC

Dominio: Paper
Find: 14-11-15-11
14 Paper Materials and Products
11 Paper products
15 Printing and writing paper
11 Writing paper
000000303 Type
000000304 Length
000000305 Width
000000306 Weight
000000307 Colour
000000308 Composition

ecl@ss

Dominio: Papier
Find: 24-11-05-34 und Standardmerkmalleisten
24 Kommunikationstechnik, Bürotechnik
11 Büromaterial
05 Büromaterial (sonstiges)
34 Kanzleipapier, Schreibpapier
AAA474001 Alterungsbeständigkeit
AAA889001 EAN Code
DDA081001 Farbe
AAA001001 Hersteller
AAA252001 Hersteller-Artikelnummer
AAA457001 Papierformate
AAA458001 Papiergewicht
AAA003001 Produkt Name
AAA002001 Produkt Typ
AAA475001 Recyclinganteil
AAA215001 Zertifikate
AAA216001 Zulassung

Nel mapping tra i codici si è utilizzata la versione in inglese di ecl@ss e si sono trascurati gli attributi, non presenti in UNSPSC, come rappresentato nello schema di Figura 32. Nello schema compaiono diverse linee che rappresentano per ogni colore un mapping tra i due standard. Ad esempio i connettori verdi costruiscono un mapping tra

il codice prodotto “Printing&Writing Paper” di UNSPSC e i codici “Writing Paper” e “Printer Paper” di ecl@ss.

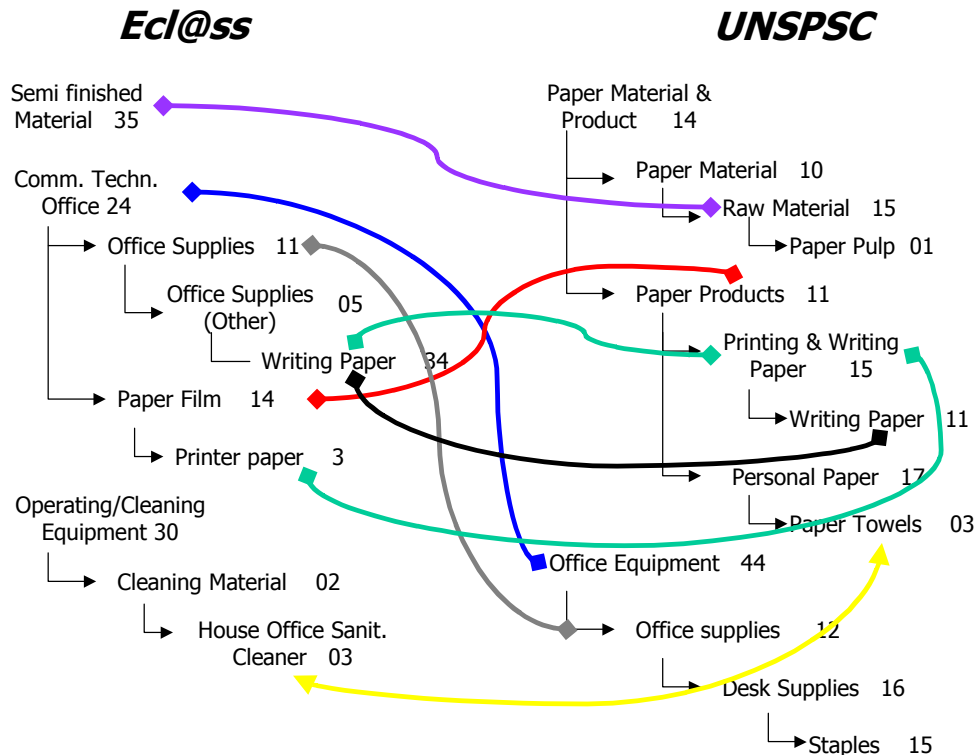


Figura 32 Esempio di mapping di OntoWeb (il caso “writing paper”)

L’articolo di OntoWeb in [Schulten, 01] non specifica il sistema con cui è stato ottenuto lo schema di mapping, ma riporta solo il risultato finale di Figura 32 dal quale comunque si possono trarre alcune indicazioni pratiche importanti di rapporto tra i due codici:

- Vi sono divergenze costruttive tra le gerarchie: ad esempio in ecl@ss “Office supplies (other)” è una sottoclasse di “Office supplies”, in UNSPSC “Paper pulp” (cellulosa) è inteso quale commodity, ma in ecl@ss si parla genericamente di semilavorati;
- Anche in esempi piccoli si può richiedere il mapping tra diversi rami della classificazione. In UNSPSC tutti i materiali e prodotti di carta sono organizzati in un singolo ramo, mentre in ecl@ss sono organizzati per funzioni. Ad esempio “paper pulp” in UNSPSC necessita di una connessione con un “Semi finished material” in

ecl@ss, mentre “paper towel” trova corrispondenza in un altro ramo di ecl@ss “Operation/cleaning equipment”;

- Alcune categorie non hanno o hanno più di una classe equivalente in differenti schemi di classificazione. Ad esempio “printing & writing paper” costituisce un’unica categoria in UNSPSC, mentre in ecl@ss vi sono due categorie.

Da un esempio corpuscolare per dimensioni (rispetto al numero di prodotti coinvolti nelle iniziative di e-commerce) emergono già importanti sottolineature sulle divergenze delle codifiche e sulla difficoltà della procedura di mapping.

Ai progettisti che parteciperanno al challenge aspetta pertanto un compito piuttosto arduo!

In conclusione è possibile formalizzare il tema di ricerca proposto. Il challenge richiede:

di progettare un modello generale (ontologia, metodo, implementazione di un campione, ...) per giungere a un mapping (automatico o semi-automatico) tra ecl@ss e UNSPSC, quali standard di riferimento per la classificazione delle descrizioni di prodotto.

6.2 Un framework di integrazione dati per la classificazione prodotti

Nel contesto di questo Challenge si colloca l’articolo [Bergamaschi, 02] nel quale vengono delineate alcune idee preliminari di una metodologia semiautomatica per l’integrazione degli standard di classificazione prodotto.

Tale metodologia è basata sul sistema MOMIS, illustrato nel capitolo precedente.

L’articolo [Bergamaschi, 02] riprende i frammenti di codice di UNSPSC e ecl@ss relativi a “Writing Paper” utilizzati nell’esempio di OntoWeb e vi applica una metodologia di integrazione preliminare basata su MOMIS.

Di seguito descriverò questo approccio metodologico, perché fornisce un riferimento di procedura di integrazione applicata nella fase sperimentale della tesi per ampliare lo studio ai due standard nel loro complesso.

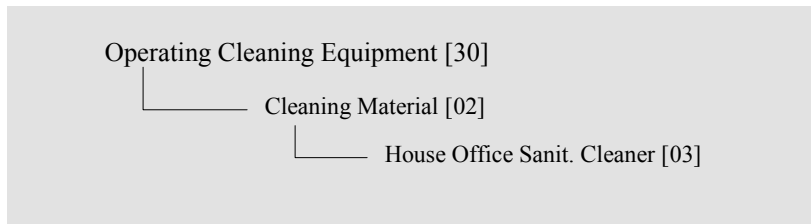
6.2.1 L'esempio "Writing Paper" con MOMIS

Nell'integrare i codici prodotto relativi a "writing paper" in UNSPSC e ecl@ss occorre tenere presente che, all'ultimo livello, ecl@ss contiene un set di attributi, mentre UNSPSC non è descrittivo al livello attributi e pertanto il mapping avviene tra le commodity.

Il progettista trasforma i frammenti di codice selezionati in file di tipo XML, che si è deciso di utilizzare perché è riconosciuto come formato standard dei dati; quindi, con la procedura di wrapping descritta nella sezione [5.3], fornisce a MOMIS le sorgenti di dati da integrare.

E' interessante soffermarsi su questa fase per capire come tradurre i livelli gerarchici delle codifiche in XML. A tale proposito si propone un piccolo esempio.

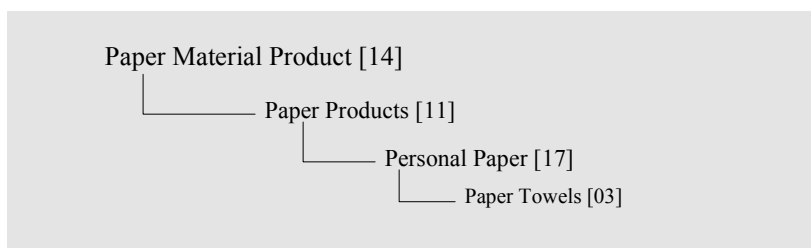
Si consideri in ecl@ss la famiglia di prodotti "Operating Cleaning Equipment" e si scenda fino al codice prodotto "House Office Saint. Cleaner" come visualizzato nello schema:



Per tradurlo in una sorgente di dati semistrutturati di tipo XML si scrive il file DTD corrispondente. La struttura della DTD utilizzata nella prova preliminare sul dominio di prodotto "Writing Paper" è la seguente:

```
<!ELEMENT PAPER (semifinishedmaterials | communicationTechnologyOffice | operatingCleaningEquipment)>
  <!ELEMENT operatingCleaningEquipment (cleansingMaterialCleaningMaterial)>
    <!ELEMENT cleansingMaterialCleaningMaterial (houseOfficeSanitCleaner)>
      <!ELEMENT houseOfficeSanitCleaner (FhouseOfficeSanitCleaner)>
    <!ELEMENT FhouseOfficeSanitCleaner (#PCDATA)>
```

In UNSPSC si individua una possibile corrispondenza nella classe di prodotto "Personal Paper" all'interno della famiglia "Paper Material Product":



In UNSPSC per trovare una corrispondenza a “House Office Saint. Cleaner” occorre scendere nella gerarchia di un ulteriore livello in cui si incontra “Paper Towel”; quindi la DTD dovrà adattarsi alla nuova struttura:

```
<!ELEMENT PAPER (paperMaterialsProducts |
officeEquipmentAndAccessoriesAndSupplies)*>
  <!ELEMENT paperMaterialsProducts (paperMaterials | paperProducts |
industrialUsePaper)>
    <!ELEMENT paperProducts (printingAndWritingPaper | noveltyPaper |
personalPaperProduct | businessPaper)>
      <!ELEMENT personalPaperProduct (paperTowels | toiletTissue)>
        <!ELEMENT paperTowels (FpaperTowels)>
          <!ELEMENT FpaperTowels (#PCDATA)>
```

Si noti che, affinché MOMIS interpreti come entità anche l’ultimo livello gerarchico, si aggiunge nella DTD un livello fittizio dichiarato (#PCDATA) che viene interpretato da MOMIS come attributo dell’entità; nell’esempio si ripete il nome espresso dall’ultimo livello e se ne modifica la forma con l’aggiunta in testa del carattere “F”.

Nell’ambito di un approccio semantico all’integrazione di dati (qual è quello adottato nel progetto MOMIS) è di fondamentale importanza la conoscenza delle informazioni semantiche relative al contesto e alla struttura dei vari schemi sorgenti. Tale conoscenza è contenuta nel cosiddetto Common Thesaurus un dizionario all’interno del quale sono presenti un insieme di relazioni terminologiche che legano tra loro classi di attributi (vedi sezione [5.5]).

Come primo passo verso la costruzione del Thesaurus c’è l’acquisizione degli schemi locali, dai quali successivamente il sistema è in grado di estrarre in modo automatico alcune relazioni. Il primo modulo che si preoccupa di estrarre le relazioni è SIM. Le relazioni estratte in questa fase sono tutte intra-schema e non esistono relazioni di sinonimia; infatti, all’interno di uno stesso schema, due classi strutturalmente identiche sono necessariamente la stessa classe. Nell’esempio di integrazione sul dominio “Writing Paper” suggerito in [Bergamaschi, 02] le sorgenti di dati inserite in input al sistema sono in formato semistutturato (XML) e pertanto le relazioni intra-schema estraibili sono di tipo:

- BT/NT ogni volta che c’è ereditarietà tra classi;
- RT quando si è in presenza di gerarchie di aggregazione.

Quindi si passa alla generazione delle relazioni inter-schema, vale a dire, relazioni che legano i termini (nomi di classi o attributi) di una sorgente con quelli di un’altra. Il modulo preposto a questa estrazione è SLIM, dove l’ontologia lessicale WordNet permette di annotare ogni nome incluso nelle sorgenti a partire dal significato del

termine e su questi genera delle relazioni. I tipi di relazioni lessicali estratte con l'utilizzo di WordNet, prima di essere inserite nel Thesaurus, vengono tradotte in relazioni semantiche tenendo presente la seguente corrispondenza:

- Sinonimia: corrisponde ad una relazione SYN;
- Iperonimia: corrisponde ad una relazione BT;
- Omonimia: corrisponde ad una relazione RT;
- Correlazione: corrisponde ad una relazione RT.

Oltre alle relazioni (intra-schema, inter-schema) estratte automaticamente, il progettista può arricchire ulteriormente il Thesaurus aggiungendo manualmente relazioni.

Sulla base di tali considerazioni e applicandole alle sorgenti di "Writing Paper", descritte con la struttura delle DTD presentata sopra, MOMIS costruisce una rappresentazione dei dati in questa forma:

```
OperatingCleaningEquipment
├── CleaningMaterial
CleaningMaterial
├── HouseOfficeSanitCleaner
HouseOfficeSanitCleaner
├── FHouseOfficeSanitCleaner
```

Quindi, in definitiva, i livelli gerarchici delle codifiche vengono rappresentati in MOMIS con una relazione Classe-Attributo e non con una relazione ISA del tipo:

```
class OperatingCleaningEquipment
class CleaningMaterial ISA OperatingCleaningEquipmentclass
HouseOfficeSanitCleaner ISA CleaningMaterial
```

in quanto in questo modo MOMIS raggrupperebbe le voci di codice in un'unica classe e si perderebbe di vista l'obiettivo di costruire mapping puntuali tra prodotti corrispondenti, riscontrabili nelle classi globali del GlobalSchema. Va precisato che le relazioni ISA permetterebbero di risparmiare tempo in fase di annotazione dei significati con WordNet perché, senza queste relazioni, il sistema interpreta lo stesso nome sia come entità, che come attributo e richiede la doppia annotazione.

In sintesi, con l'ausilio dei moduli SIM e SLIM e grazie al supporto del progettista, SI-Designer costruisce le relazioni inter-schema e intra-schema. Per ultimo, nel modulo di clusterizzazione, viene generato lo schema globale che fornisce come risultato le classi globali contenenti i mapping tra i codici prodotto.

In Figura 33 si visualizzano i frammenti degli standard utilizzati nella prova preliminare in [Bergamaschi, 02].

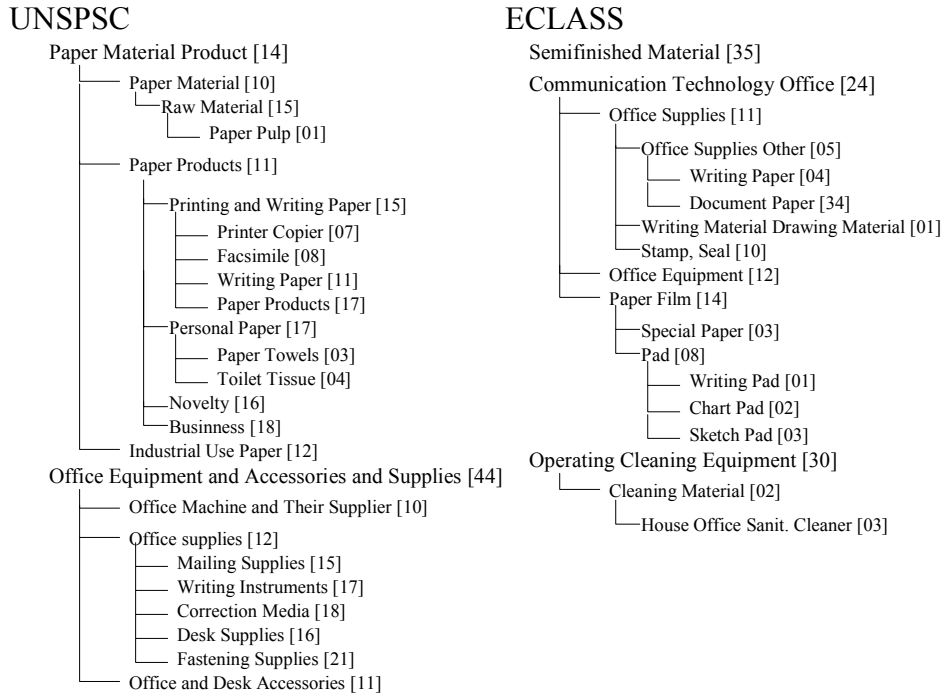


Figura 33 Frammenti di codice UNSPSC e ecl@ss integrati con MOMIS

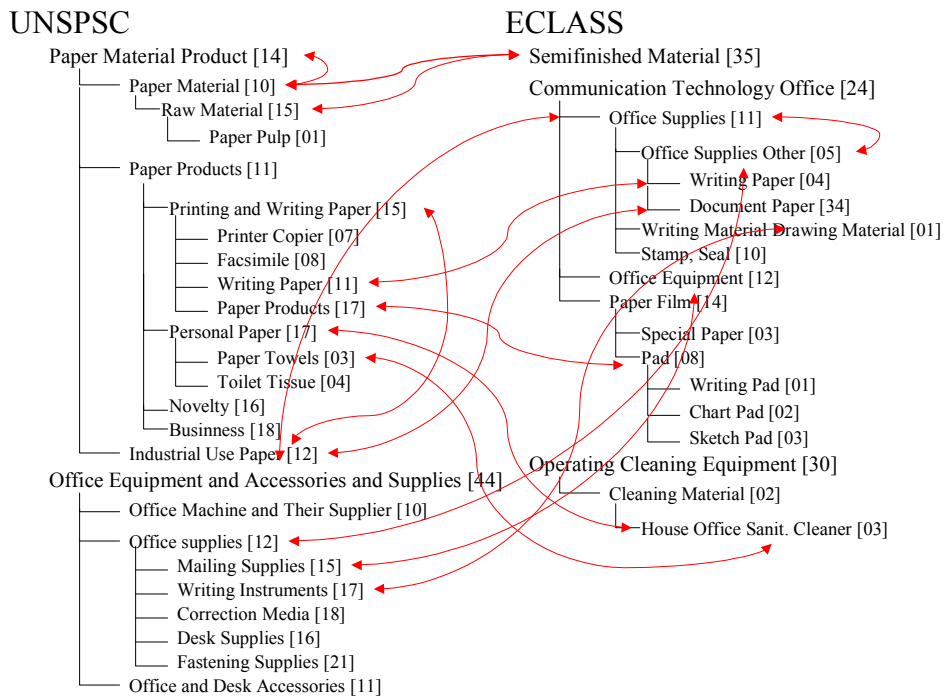


Figura 34 Risultato del mapping sul dominio "Writing Paper"

In Figura 34 si riporta la rappresentazione del mapping ottenuto sull'esempio "Writing Paper". I mapping possono essere confrontati con quelli di Figura 32 solo per la parte di dominio in comune, visto che in [Schulten, 01] si sono considerati un numero minore di codici prodotto. In ogni caso i mapping rappresentati in Figura 32 sono stati ottenuti anche con la metodologia di integrazione proposta in [Bergamaschi, 02], che inoltre ha individuato molti altri mapping, perché estesa a un dominio più ampio.

Da queste prove preliminari si intravede il sistema MOMIS come strumento interessante per l'integrazione semi-automatica di standard di classificazione prodotti nel commercio elettronico; nel seguito della trattazione si descriveranno i test effettuati su MOMIS per ampliare lo studio di integrazione sulle classificazione a tutte le voci di codice.

6.3 Estensione del processo di integrazione

dal dominio “Writing paper” all’analisi completa degli standard di classificazione scelti

All’inizio del progetto di integrazione, oltre ai due standard suggeriti dal Challenge (UNSPSC e ecl@ss), si è preso in considerazione l’altro codice presentato in [3.8]: NAICS. Tuttavia NAICS classifica le attività produttive o di vendita e non il prodotto, ciò lo rende molto diverso da UNSPSC e da ecl@ss e pertanto difficile da comparare agli altri.

L’obiettivo era ampliare l’analisi per testare se il sistema MOMIS, utilizzato nell’integrazione di una piccola porzione di codice, si confermasse strumento valido e adeguato per trattare in modo completo il matching tra gli standard.

Pertanto sono state svolte una serie di prove, prendendo in considerazione di volta in volta varie parti di codice che presentassero tra loro qualche affinità semantica. La selezione dei frammenti è stata molto difficoltosa in quanto è complicato individuare frammenti appartenenti alle diverse codifiche che siano fra loro mappabili.

Il progetto è stato affrontato con due punti di vista: da un lato si sono approfondite le problematiche emerse durante la fase sperimentale, ricercandone le origini, dall’altro, una volta evidenziate le criticità, si sono individuati possibili ambiti risolutivi per definire una nuova metodologia sulla base dei test e delle caratteristiche di MOMIS. Il secondo punto di vista è l’oggetto di studio di questa tesi e di conseguenza scopo del mio lavoro.

6.3.1 Studio delle codifiche

Si è detto che le prove di integrazione fatte su MOMIS devono riguardare l’intero dominio classificato dagli standard; quindi è stato necessario studiare i tre standard in esame, a livello di segmento (il primo della gerarchia), per cercare delle affinità e in base a queste definire un metodo con cui trattarle per testare le corrispondenze.

Nelle tabelle seguenti si riportano i primi livelli, rispettivamente delle tre codifiche, per aiutare la comprensione degli approcci scelti nell’integrazione.

Ecl@ss


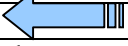


class	preferred name
20000000	Packing material
21000000	Tools, machine tools
22000000	Construction technology
23000000	Machine element, fixing
24000000	Communication technology, office -
25000000	Service
26000000	Energy, basic chemicals, aux. agents 
27000000	Automation, electrical-engineering, PLT
28000000	Automotive technology
29000000	Home economics, Home technology
30000000	Operating, cleaning material, equipment
31000000	Polymers 
32000000	Laboratory material, Laboratory technology
33000000	Installation (complete, turnkey)
34000000	Medicine, medical technology
35000000	Semi finished products, materials
36000000	Machine, apparatus
37000000	Pipeline technology
38000000	Chemical product (inorganic) 
39000000	Chemical product (organic) 
40000000	Occupational safety, accident prevention
41000000	Advertising

Tabella 14 Segmenti di ecl@ss

NAICS

class	preferred name
11	Agriculture, Forestry, Fishing and Hunting
21	Mining
22	Utilities
23	Construction
42	Wholesale Trade
51	Information
52	Finance and Insurance
53	Real Estate and Rental and Leasing
54	Professional, Scientific, and Technical Services
55	Management of Companies and Enterprises
56	Administrative and Support and Waste Management and Remediation Services
61	Educational Services
62	Health Care and Social Assistance
71	Arts, Entertainment, and Recreation
72	Accommodation and Food Services
81	Other Services (except Public Administration)
92	Public Administration

Tabella 15 Segmenti di NAICS

UNSPSC

class	preferred name
10000000	Live Plant and Animal Material and Accessories and Supplies
11000000	Mineral and Textile and Inedible Plant and Animal Materials
12000000	Chemicals including Bio Chemicals and Gas Materials
13000000	Resin and Rosin and Rubber and Foam and Film and Elastomeric Materials
14000000	Paper Materials and Products
15000000	Fuels and Fuel Additives and Lubricants and Anti corrosive Materials
20000000	Mining Machinery and Accessories
21000000	Farming and Fishing and Forestry and Wildlife Machinery and Accessories
22000000	Building and Construction Machinery and Accessories
23000000	Industrial Manufacturing and Processing Machinery and Accessories
24000000	Material Handling and Conditioning and Storage Machinery and their Accessories and Supplies
25000000	Commercial and Military and Private Vehicles and their Accessories and Components
26000000	Power Generation and Distribution Machinery and Accessories
27000000	Tools and General Machinery
30000000	Structures and Building and Construction and Manufacturing Components and Supplies
31000000	Manufacturing Components and Supplies
32000000	Electronic Components and Supplies
39000000	Lighting and Electrical Accessories and Supplies
40000000	Distribution and Conditioning Systems and Equipment and Components
41000000	Laboratory and Measuring and Observing and Testing Equipment
42000000	Medical Equipment and Accessories and Supplies
43000000	Information Technology
44000000	Office Equipment and Accessories and Supplies
45000000	Printing and Photographic and Audio and Visual Equipment and Supplies
46000000	Defence and Law Enforcement and Security and Safety Equipment and Supplies
47000000	Cleaning Equipment and Supplies
48000000	Service Industry Machinery and Equipment and Supplies
49000000	Musical Instruments and Recreational Equipment and Supplies and Accessories
50000000	Food Beverage and Tobacco Products
51000000	Drugs and Pharmaceutical Products
52000000	Domestic Appliances and Supplies and Consumer Electronic Products
53000000	Apparel and Luggage and Personal Care Products
54000000	Timepieces and Jewellery and Gemstone Products
55000000	Published Products
56000000	Furniture and Furnishings
70000000	Farming and Fishing and Forestry and Wildlife Contracting Services
71000000	Mining and Oil and Gas Services
72000000	Building and Construction and Maintenance Services
73000000	Industrial Production and Manufacturing Services
76000000	Industrial Cleaning Services
77000000	Environmental Services
78000000	Transportation and Storage and Mail Services
80000000	Management and Business Professionals and Administrative Services
81000000	Research and Science Based Services
82000000	Editorial and Design and Graphic and Fine Art Services
83000000	Public Utilities and Public Sector Related Services
84000000	Financial and Insurance Services
85000000	Healthcare Services

86000000	Education and Training Services
90000000	Travel and Food and Lodging and Entertainment Services
91000000	Personal and Domestic Services
92000000	National Defence and Public Order and Security and Safety Services
93000000	Politics and Civic Affairs Services
94000000	Organizations and Clubs

Tabella 16 Segmenti di UNSPSC

A colpo d’occhio emerge subito un dato importante: il dominio coperto da UNSPSC è vastissimo, gli altri sono limitati, ecl@ss alle categorie di interesse dei partner (in particolare “Chemistry”), NAICS ai settori importanti ai fini statistici (si ricorda che NAICS è un’iniziativa promossa da U.S. Census Bureau per fornire dati comparabili in attività statistiche e di business).

Per “navigare” tra i codici occorre dotarsi di una strategia che permetta di confrontare i segmenti che presentano analogia semantica e allo stesso tempo non precluda il mapping tra categorie apparentemente non individuate in una codifica, ma poi comprese nei livelli inferiori.

6.3.2 Due approcci per la navigazione tra le codifiche

Il primo passo verso l’integrazione è l’individuazione di frammenti di codice confrontabili, adatti all’integrazione. Lo studio degli standard evidenzia alcune divergenze, da tenere presenti nella scelta della strategia:

- Diverso numero di segmenti
 - $N^{\circ}_{\text{segmenti}} \text{ ecl@ss (22)} \ll N^{\circ}_{\text{segmenti}} \text{ UNSPSC (54)}$
 - $N^{\circ}_{\text{segmenti}} \text{ NAICS (20)} \ll N^{\circ}_{\text{segmenti}} \text{ UNSPSC (54)}$
- Diversi punti di vista
 - UNSPSC, ecl@ss classificano i prodotti
 - NAICS classifica le attività produttive e di vendita
- Diverse scelte di approfondimento dei domini
 - UNSPSC, fra l’altro, approfondisce i servizi (18 segmenti), ecl@ss dedica molta attenzione al settore chimico
- Diverso grado di accuratezza nell’approfondimento dei domini
 - UNSPSC è molto dettagliato sui prodotti
 - Ecl@ss è poco dettagliato sui prodotti

- NAICS è dettagliato sui fattori produttivi

In base a queste considerazioni emerge che vi sono segmenti che non trovano corrispondenze in tutti gli standard. Conviene distinguere tra due punti di partenza: si possono scorrere le codifiche cercando analogie tra i segmenti (ad esempio le categorie evidenziate con lo stesso colore in Tabella 15 e Tabella 16), per poi procedere all'integrazione dei frammenti corrispondenti, oppure si può scegliere un tipo di prodotto e cercarlo in entrambi i sistemi di classificazione.

Nel corso delle prove si è scelto di denominare il primo approccio di tipo Top-down, il secondo di tipo Bottom-up, come schematizzati in Figura 35 e Figura 36.

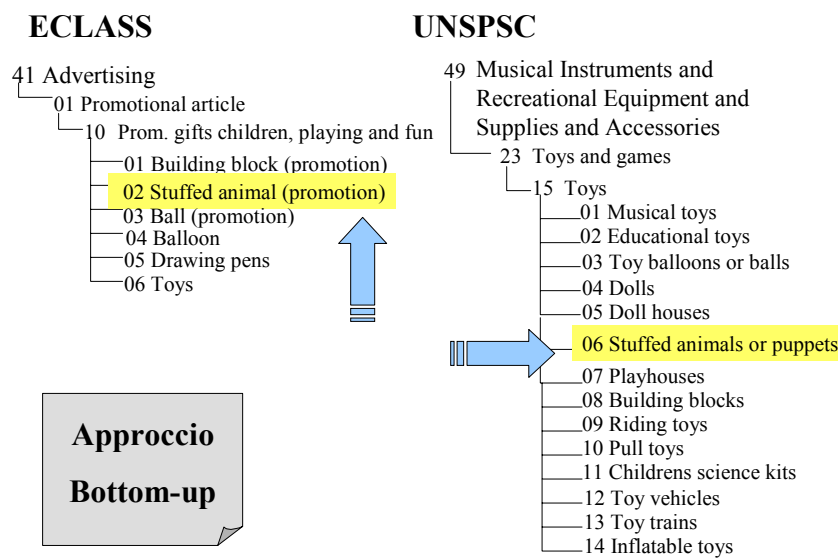


Figura 35 Approccio Bottom-Up

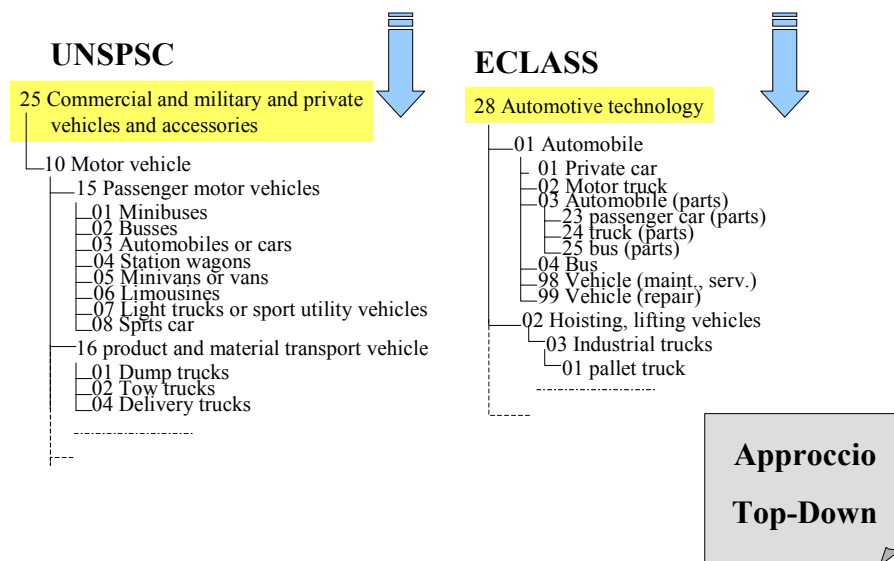


Figura 36 Approccio Top-Down

6.3.2.1 L'approccio Bottom-Up

Il punto di partenza è la scelta di un prodotto da cercare all'interno delle codifiche, ad esempio "bottle" (bottiglia); tra i possibili frammenti individuati con la ricerca si sceglie il più pertinente o quello che, essendo ad un livello superiore comprende anche gli altri. Una volta individuato il codice prodotto cercato, ci si muove contemporaneamente in due distinte direzioni, si sale lungo la gerarchia fino ad arrivare al segmento cui il codice appartiene e nello stesso tempo si scende attraverso le diramazioni che derivano dal codice, comprendendo tutti i prodotti che vengono classificati nei livelli inferiori. Nel caso il prodotto si trovasse già nel livello più basso della gerarchia si considerano tutti i "fratelli" del prodotto in questione.

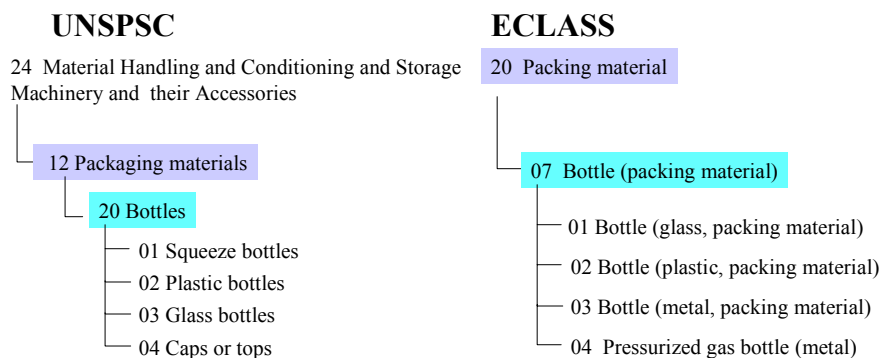


Figura 37 Frammenti di codice su cui applicare l'approccio Bottom-Up

6.3.2.2 L'approccio Top-Down

Si parte esaminando il primo livello della gerarchia dei sistemi di classificazione, che nel caso di UNSPSC ed ecl@ss è costituito dai segmenti, mentre nel caso di NAICS è costituito dai settori. Si devono cercare similitudini tra i diversi standard; se vengono individuate, i frammenti si costruiscono scendendo nella gerarchia di uno o due livelli a partire dal segmento/settore scelto.

Per rispondere alle differenze di numerosità dei segmenti contenuti negli standard la soluzione a questo problema sarebbe considerare anche delle unioni, per esempio si potrebbe individuare un segmento di ecl@ss che corrisponde a 2 o 3 segmenti di UNSPSC. Questo tentativo è stato da principio poco utilizzato poiché complica notevolmente la gestione dell'integrazione.

S'intuisce che l'approccio top-down deve essere usato solo quando c'è una corrispondenza evidente tra due segmenti, quindi non è uno strumento per ottenere il mapping completo degli standard, ma solo per svolgere prove di integrazione.

Tra le prove effettuate con questo approccio si può citare l'esempio che mappa i frammenti di codice relativi a "Farming" negli standard UNSPSC e NAICS, in particolare i segmenti:

- "Live Plant and Animal Material and Accessories and Supplies" in UNSPSC
- "Agriculture, Forestry, Fishing and Hunting" in NAICS

Si sono scesi due livelli nella gerarchia, considerando un ampio spettro di prodotti. Nell'esempio ci si ferma al terzo livello per evitare di gestire una grande quantità di dati eterogenei con termini troppo specifici, che sarebbero di difficile comprensione per un dizionario generico come WordNet.

Il principale inconveniente di questo metodo è che non può essere applicato su tutta la classificazione, ma solo su quelle parti tra le quali si riesce ad individuare una corrispondenza sin dal livello dei segmenti/settori. Un altro ostacolo è che ci si affida alle capacità del progettista nell'individuare corrispondenze a prima vista. Un metodo per rendere questa tecnica più affidabile è verificare, una volta individuate le corrispondenze tra i segmenti, che in entrambi i rami sottostanti delle gerarchie siano presenti una serie di parole chiave specifiche del dominio selezionato; la presenza di questi termini comuni dà la conferma della corrispondenza tra segmenti/settori.

Nell'esempio si può verificare che in entrambi i frammenti si trovano parole chiave come: "crop", "seed", vedi Figura 38 e Figura 39.

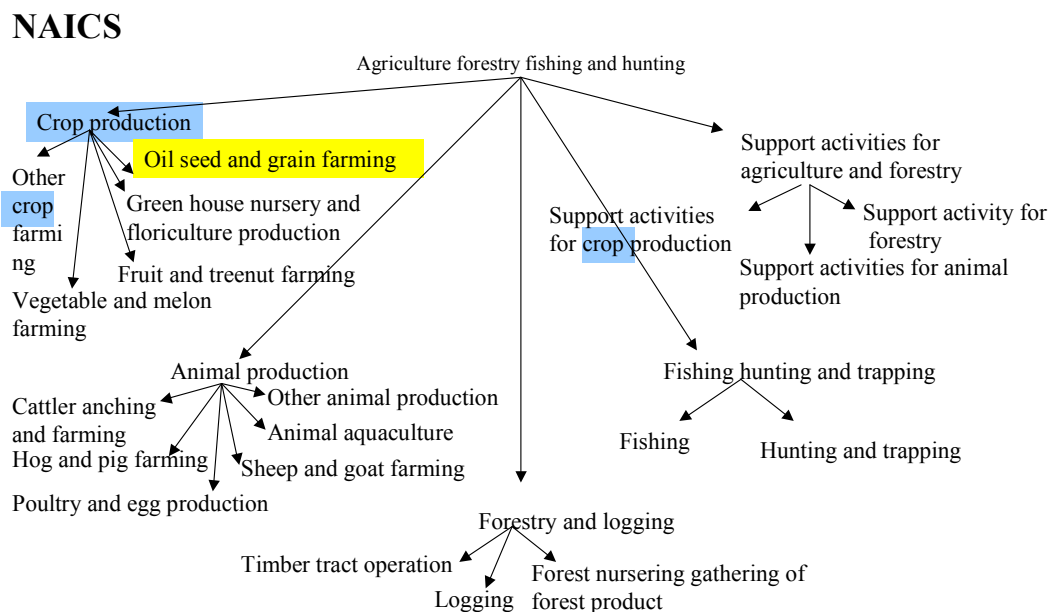
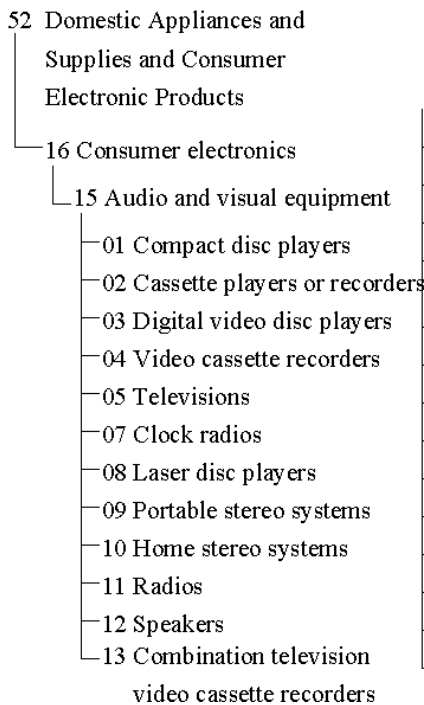


Figura 38 Segmento di NAICS sul dominio Farming

UNSPSC



ECLASS

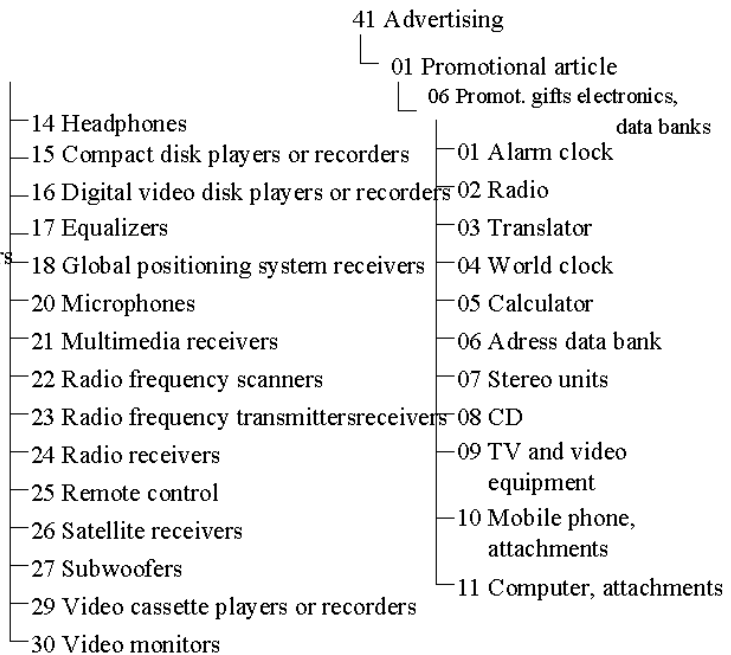


Figura 40 Approccio Bottom-Up sul dominio RadioTV

Dopo la selezione dei frammenti adatti all'integrazione, si passa alla loro formalizzazione mediante la costruzione del file XML.

A partire dagli schemi in Figura 40, come primo passo, si preparano i file DTD, uno per ogni standard, in modo da mantenere la gerarchia indicata.

Riporto il listato dei file uradiotv.dtd relativo a UNSPSC e il file eradiotv.dtd relativo a ecl@ss.

File DTD del domino "RadioTV" in UNSPSC:

```
<!ELEMENT DomesticAppliancesandSuppliesandConsumerElectronicProducts
(ConsumerElectronics)>
  <!ELEMENT ConsumerElectronics (AudioandVisualEquipment)>
    <!ELEMENT AudioandVisualEquipment (CompactDiscPlayers |
      CassettePlayersorRecorders | DigitalVideoDiscPlayers |
      VideoCassetteRecorders | Televisions | ClockRadios |
      LaserDiscPlayers | PortableStereoSystems |
      HomeStereoSystems | Radios | Speakers |
      CombinationTelevisionVideoCassetteRecorders | Headphones |
      CompactDiskPlayersorRecorders |
      DigitalVideoDiskPlayersorRecorders | Equalizers |
      GlobalPositioningSystemReceivers | Microphones |
      MultimediaReceivers | RadioFrequencyScanners |
      RadioFrequencyTransmittersreceivers | RadioReceivers |
      RemoteControl | SatelliteReceivers | Subwoofers |
      VideoCassettePlayersorRecorders | VideoMonitors)>
    <!ELEMENT CompactDiscPlayers (cod)>
```

```

<!ELEMENT CassettePlayersorRecorders (cod)>
<!ELEMENT DigitalVideoDiscPlayers (cod)>
<!ELEMENT VideoCassetteRecorders (cod)>
<!ELEMENT Televisions (cod)>
<!ELEMENT ClockRadios (cod)>
<!ELEMENT LaserDiscPlayers (cod)>
<!ELEMENT PortableStereoSystems (cod)>
<!ELEMENT HomeStereoSystems (cod)>
<!ELEMENT Radios (cod)>
<!ELEMENT Speakers (cod)>
<!ELEMENT CombinationTelevisionVideoCassetteRecorders (cod)>
<!ELEMENT Headphones (cod)>
<!ELEMENT CompactDiskPlayersorRecorders (cod)>
<!ELEMENT DigitalVideoDiskPlayersorRecorders (cod)>
<!ELEMENT Equalizers (cod)>
<!ELEMENT GlobalPositioningSystemReceivers (cod)>
<!ELEMENT Microphones (cod)>
<!ELEMENT MultimediaReceivers (cod)>
<!ELEMENT RadioFrequencyScanners (cod)>
<!ELEMENT RadioFrequencyTransmittersreceivers (cod)>
<!ELEMENT RadioReceivers (cod)>
<!ELEMENT RemoteControl (cod)>
<!ELEMENT SatelliteReceivers (cod)>
<!ELEMENT Subwoofers (cod)>
<!ELEMENT VideoCassettePlayersorRecorders (cod)>
<!ELEMENT VideoMonitors (cod)>
<!ELEMENT cod (#PCDATA)>

```

File DTD del domino “RadioTV” in ecl@ss:

```

<!ELEMENT Advertising (PromotionalArticle)>
  <!ELEMENT PromotionalArticle (PromotGiftsElectronicsDataBanks)>
    <!ELEMENT PromotGiftsElectronicsDataBanks (AlarmClock | Radio |
      Translator | WorldClock | Calculator | AdressDataBank |
      StereoUnits | CD | TVandVideoEquipment |
      MobilePhoneAttachments | ComputerAttachments)>
  <!ELEMENT AlarmClock (cod)>
  <!ELEMENT Radio (cod)>
  <!ELEMENT Translator (cod)>
  <!ELEMENT WorldClock (cod)>
  <!ELEMENT Calculator (cod)>
  <!ELEMENT AdressDataBank (cod)>
  <!ELEMENT StereoUnits (cod)>
  <!ELEMENT CD (cod)>
  <!ELEMENT TVandVideoEquipment (cod)>
  <!ELEMENT MobilePhoneAttachments (cod)>
  <!ELEMENT ComputerAttachments (cod)>
<!ELEMENT cod (#PCDATA)>

```

Dai file DTD è possibile costruire i file XML, come visualizzati di seguito:

File XML del domino “RadioTV” in UNSPSC:

```

<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE DomesticAppliancesandSuppliesandConsumerElectronicProducts SYSTEM
"uradiotv.dtd">
<DomesticAppliancesandSuppliesandConsumerElectronicProducts>
  <ConsumerElectronics>
    <AudioandVisualEquipment>
      <CompactDiscPlayers>
        <cod/>
      </CompactDiscPlayers>
    </AudioandVisualEquipment>
  </ConsumerElectronics>
</DomesticAppliancesandSuppliesandConsumerElectronicProducts>

```

File XML del domino “RadioTV” in ecl@ss:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Advertising SYSTEM "eradiotv.dtd">
<Advertising>
  <PromotionalArticle>
    <PromotGiftsElectronicsDataBanks>
      <AlarmClock>
        <cod/>
      </AlarmClock>
    </PromotGiftsElectronicsDataBanks>
  </PromotionalArticle>
</Advertising>
```

6.3.4 Inizializzazione dei Wrapper XML

Per caricare le sorgenti XML in MOMIS bisogna ricorrere all’ausilio dei wrapper che devono essere inizializzati a partire dai file XML creati.

Il wrapper XML presiede sorgenti di dati di tipo XML valid. In particolare viene effettuata la traduzione della struttura del documento XML così come espressa nella DTD in una struttura avente un’analogia semantica nel linguaggio ODL₁₃. Una particolare attenzione durante la progettazione di questa parte del sistema MOMIS è stata posta nella realizzazione di un’interfaccia grafica che permetta al progettista la scelta di chiavi primarie, candidate e esterne, non contemplate in XML, ma fortemente utilizzate in MOMIS per la realizzazione del Global Schema.

La fase di attivazione del wrapper avviene in connessione con *sparc20.ing.unimo.it*, il server di MOMIS su cui è installato il sistema operativo Solaris 2.7; per farlo in remoto si devono rispettare alcuni requisiti:

- Si effettua la connessione a *sparc20.ing.unimo.it* da un PC utilizzando emulatori di terminale che supportano SSH³⁰ tramite l’account messo a disposizione agli utenti MOMIS (ad es. bonini@sparc20.ing.unimo.it);
- Si attiva l’X server³¹;
- Si settano le variabili ambientali per interagire in locale (*source 0setVars*) e si ridirige l’output grafico sulla macchina corrente per effettuare le operazioni successive (ad es. *setenv DISPLAY terminale.ing.unimo.it:0.0*)

³⁰ Secure Shell (SSH) è un programma per connettersi in rete a un altro computer, per eseguire comandi su una macchina remota e per spostare file da un computer all’altro. Fornisce comunicazioni sicure grazie all’autenticazione utente

³¹ Xserver è un server di connessione a un Terminale X, una rete distribuita che usa X Windows System (sistema aperto client/server per la gestione delle interfacce grafiche a finestra in una rete).

Quindi, tramite collegamento FTP, si spostano i file XML e DTD, precedentemente copiati nella propria directory su *sparc20*, nella cartella *wrapperXMLg* e si attiva il wrapper relativo a uno dei due standard con il comando:

```
java StartWrapper nomefile.xml
```

Nell'esempio preso in considerazione:

```
java StartWrapper uradiotv.xml
```

```
java StartWrapper eradiotv.xml
```

A questo punto è possibile lavorare su MOMIS, per il caricamento delle sorgenti e le successive fasi di integrazione, in particolare si procederà con il tool grafico SI-Designer.

In appendice A si descrivono brevemente i software utilizzati nella fase di Wrapping.

6.3.5 Fasi dell'integrazione su SI-Designer

SI-Designer è la GUI (Interfaccia Utente Grafica) per la parte di creazione del Global Schema sul sistema MOMIS. SI-Designer guida l'utente attraverso le varie fasi dell'integrazione, dall'acquisizione delle sorgenti fino alla messa a punto delle Mapping Table. Si tratta di un contenitore modulare di altri strumenti che permettono la raccolta di informazioni per l'integrazione e permettono di dichiarare relazioni tra gli oggetti degli schemi.

La GUI di SI-Designer è una sequenza di finestre, ognuna delle quali relativa ad una fase del processo di integrazione. Il tool è stato progettato ed implementato in Java e un enorme sforzo è stato fatto per sviluppare un'architettura modulare, in modo da rendere molto semplice l'aggiunta di nuove fasi nel processo di integrazione.

SI-Designer è accessibile on-line da emulatore di terminale nella directory:

```
cd $djp/modules/SIDesigner
```

con il comando:

```
java SI_Designer siDesigner.conf
```

All'inizio del progetto si è scelta questa modalità, tuttavia le prove on-line risultavano troppo lente e la perdita di tempo per comunicare con il server andava a discapito della qualità del lavoro. Si è scelto quindi di adottare la versione software di MOMIS e di limitare l'attività on-line alla creazione dei wrapper, all'inferenza e alla validazione delle relazioni.

Analizziamo i passi che il progettista deve seguire, riprendendo l'esempio del dominio "RadioTV".

6.3.5.1 Acquisizione delle sorgenti

L'acquisizione delle sorgenti avviene nel pannello d'apertura di SI-Designer *Source* in cui l'utente carica i wrapper delle sorgenti di dati interessate all'integrazione e il sistema le visualizza in *Sources acquired* (Figura 41).

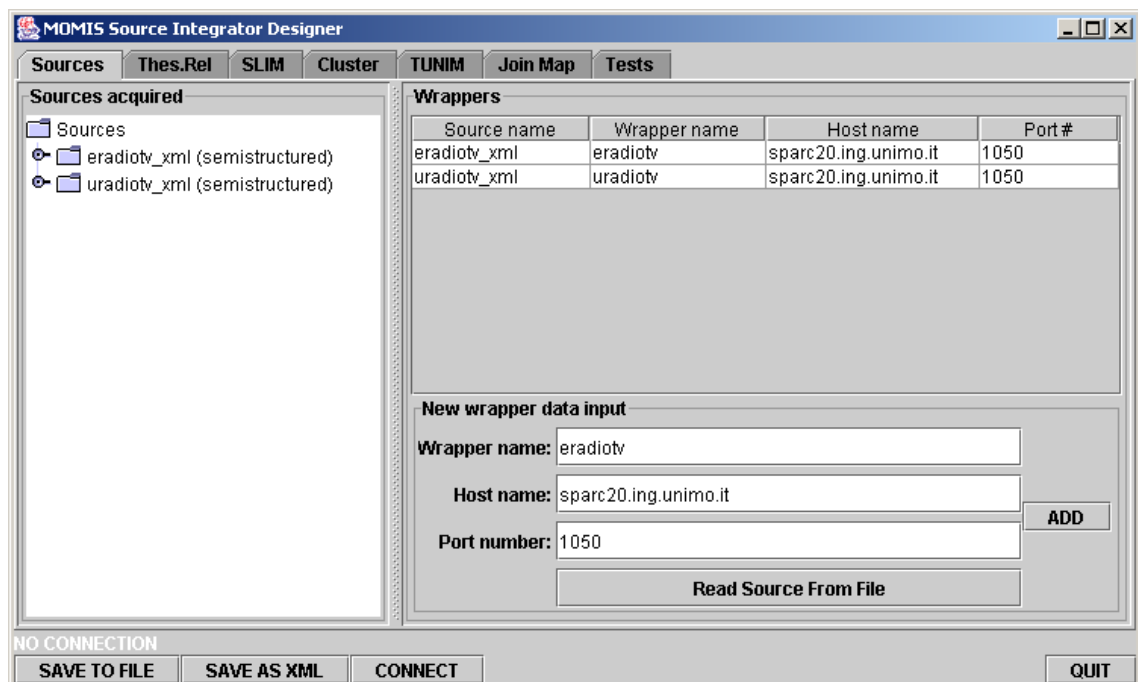


Figura 41 Acquisizione delle sorgenti nell'esempio RadioTV

In questa fase si lavora con la versione CD di MOMIS, ma i wrapper, creati on-line, devono essere attivi, quindi occorre mantenere aperto il collegamento con *sparc20* fino alla conclusione del caricamento.

6.3.5.2 Run SIM A

Una volta lette le sorgenti si accede al pannello *Thes.Rel*, dove avviene il primo passo nella formazione del Thesaurus Comune: l'estrazione delle relazioni intra-schema.

Per fare questo, Figura 42, si seleziona il pulsante *run SIMA* il cui risultato sarà la comparsa nella finestra di destra delle relazioni tra le classi e gli attributi.

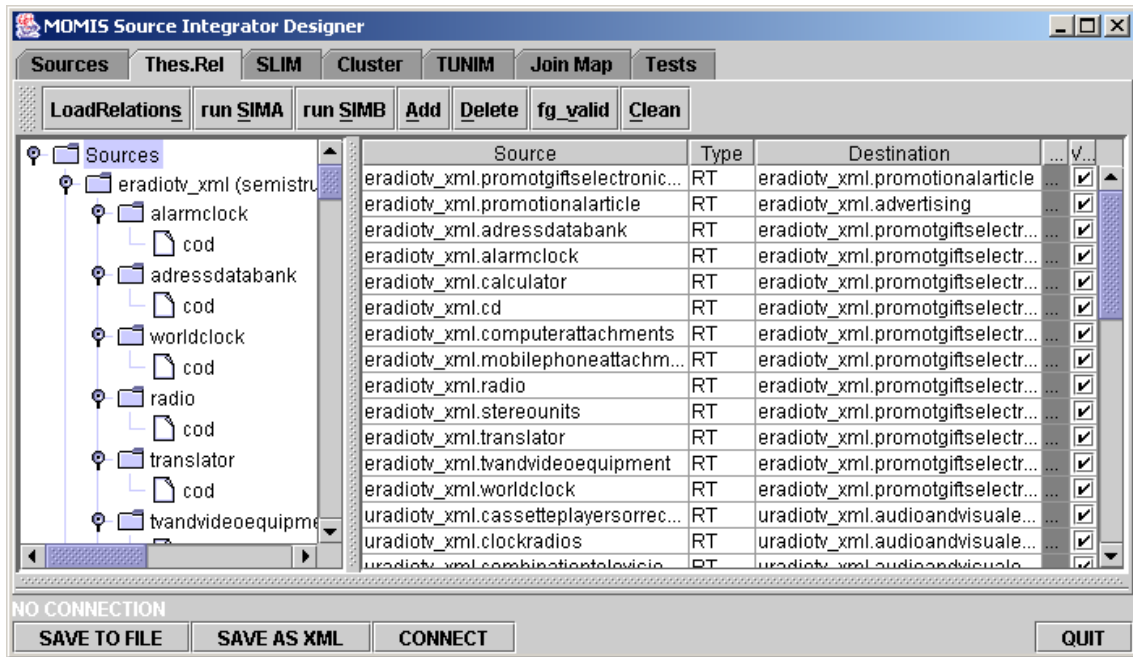


Figura 42 SIM A nell'esempio RadioTV

6.3.5.3 Salvataggio dello stato salvato

A questo punto è bene salvare uno stato di MOMIS con le sorgenti caricate (Figura 43), si genera cioè uno stato salvato da cui ripartire una volta riaperto SI-Designer.

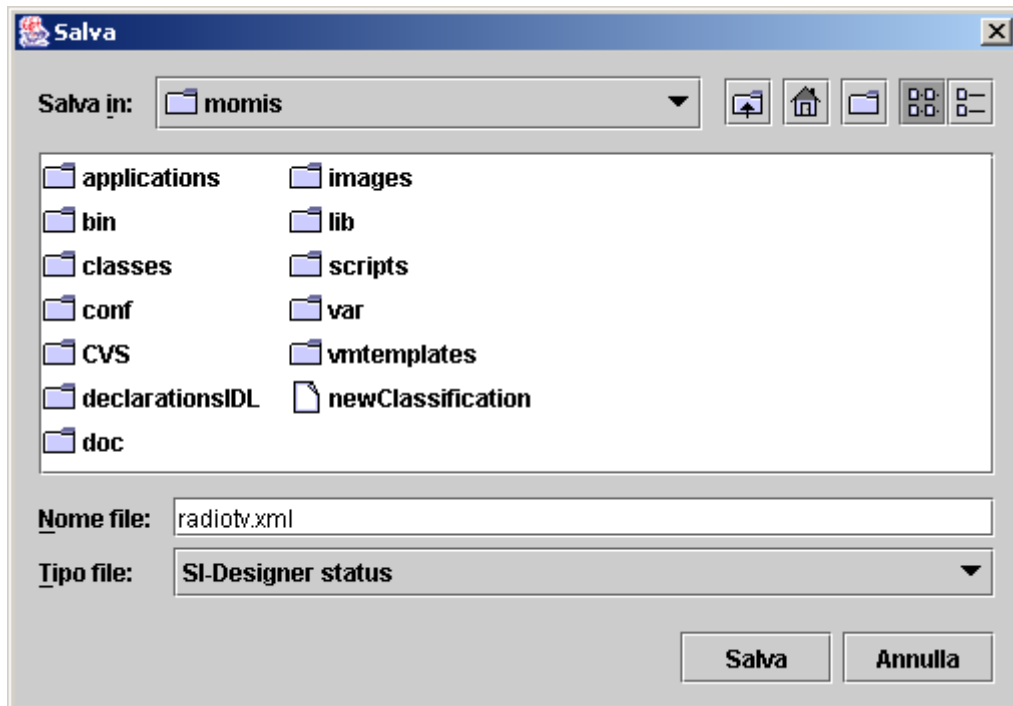


Figura 43 Salvataggio dello stato XML

La particolarità di SI-Designer è che per salvare il lavoro fatto fino a un certo pannello occorre cambiare finestra, il comando SAVE AS XML salva cioè il lavoro fino alla finestra precedente rispetto a quella in cui si effettua il salvataggio.

L'effettuazione del salvataggio in questo momento dell'integrazione è importante perché, nella riapertura dello stato, si controlla che non vi siano errori nella costruzione precedente delle sorgenti: se ad esempio i file DTD di origine contenessero descrizioni di codice uguali, il sistema, riscontrando delle ambiguità, segnalerebbe l'errore sul Prompt di MS-DOS e non consentirebbe l'apertura dello stato.

6.3.5.4 Modulo lessicale

L'operazione successiva all'estrazione delle relazioni intra-schema consiste nella costruzione delle relazioni lessicali tra i termini degli attributi e delle classi dei diversi schemi; quindi, dopo avere riaperto lo stato salvato, si accede al modulo *SLIM*.

Si presenta una sequenza di operazioni standard da seguire:

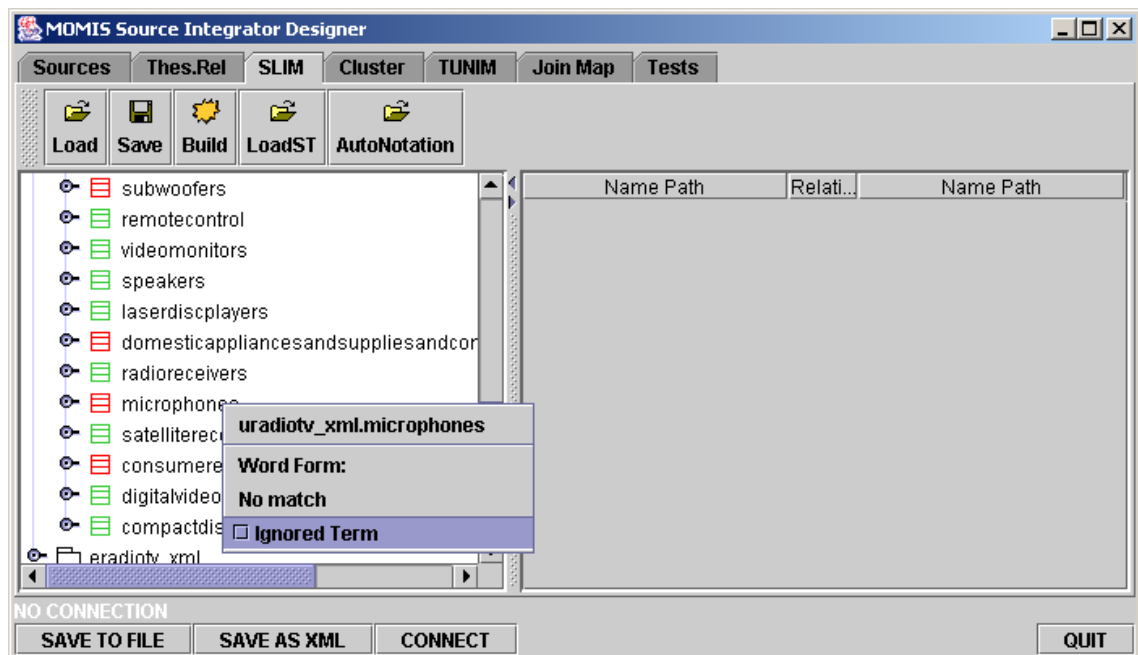


Figura 44 Modulo SLIM: apertura menù a scomparsa

- Apertura del menù a scomparsa per l'assegnazione del significato: l'utente può scegliere di non considerare nell'integrazione successiva il termine in esame (Figura 44), oppure decide di scegliere un significato e segue i passi successivi;

- Assegnazione della WordForm (Figura 45) al prodotto in esame;

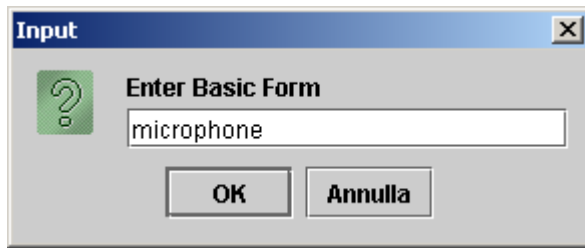


Figura 45 Assegnazione Forma Base

- Ricerca e selezione sul database lessicale WordNet del significato che il progettista sceglie di assegnare alla Word Form (Figura 46);

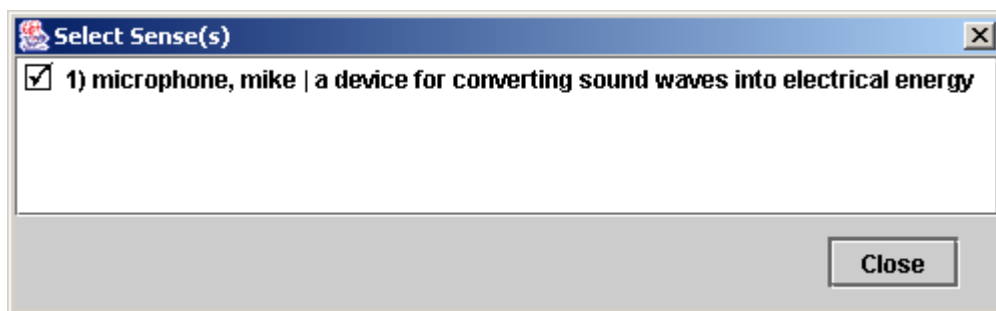


Figura 46 Selezione del significato su WordNet

- Dopo avere preso in considerazione tutte le voci, il progettista può salvare il modulo lessicale in modo da poterlo riaprire (load) in un secondo momento; quindi si costruiscono le relazioni inter-schema cliccando sul pulsante *build* e le relazioni compaiono sulla parte destra della finestra. (Figura 47);

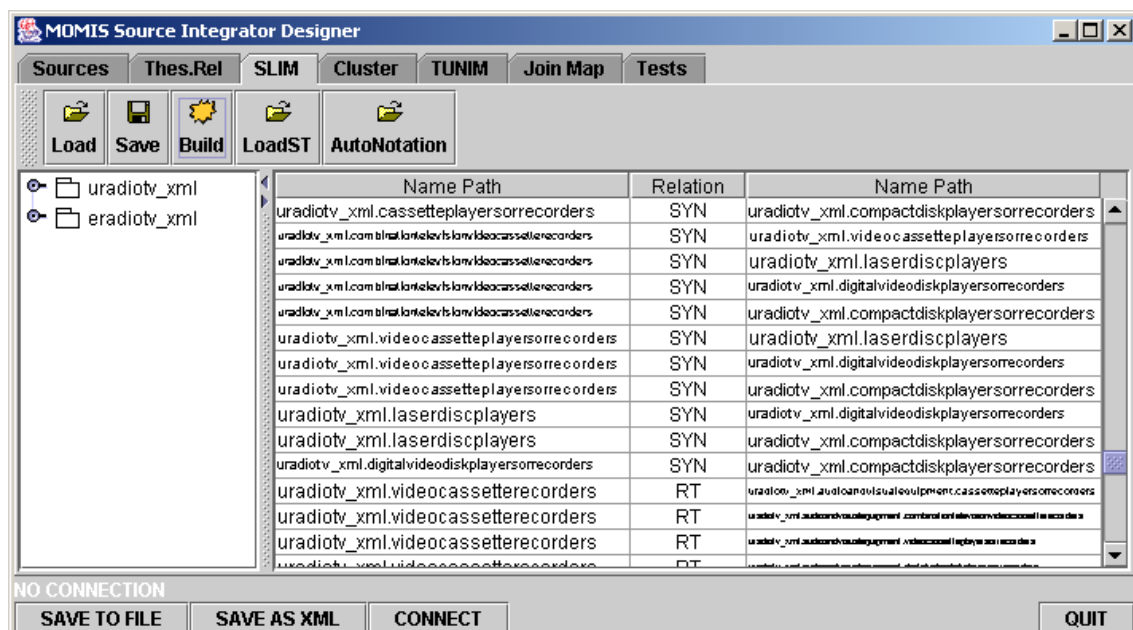


Figura 47 Costruzione delle relazioni inter-schema

6.3.5.5 SIM B

Costruite le relazioni dal modulo lessicale si ritorna al pannello *Thes.Rel*, dove il progettista, sulla base delle scelte effettuate con SLIM, può decidere di introdurre manualmente nuove relazioni, per mezzo del tasto ADD che fa comparire una finestra in cui inserire il termine *source*, quello *destination* e la relazione che li lega. Un caso importante in cui ricorrere a questa possibilità si incontra quando in entrambe le sorgenti vi è una descrizione di codice uguale, ma non rilevata da WordNet; occorre allora impostare una relazione di sinonimia tra le due o più voci simili, affinché nello schema globale rientrino in un'unica classe (Figura 48).

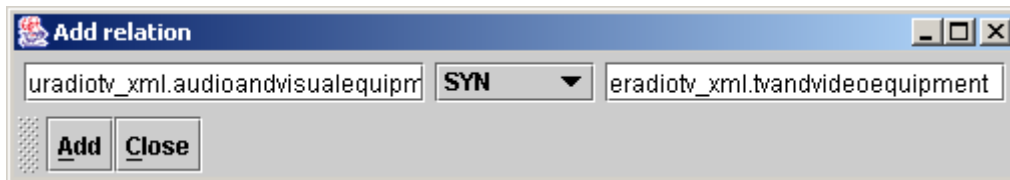


Figura 48 Creazione manuale delle relazioni

Inoltre per mezzo del tasto DELETE il progettista può rimuovere delle relazioni che, in base all'esperienza, ritiene errate.

Concluse queste operazioni di arricchimento, modifica e correzione delle relazioni, premendo il tasto *run SIMB*, si fa compiere al sistema il processo di controllo e validazione di tutte le relazioni contenute nel Thesaurus Comune.

La procedura di validazione ha lo scopo di verificare che i domini degli attributi coinvolti siano compatibili con il tipo di relazione definita tra gli stessi. Per realizzare la suddetta procedura si utilizzano le funzionalità di ODBTools, quindi va precisato che le fasi di inferenza e validazione rientrano fra quelle attuabili non in versione CD di MOMIS, ma direttamente connessi a *sparc20*.

6.3.5.6 Clusterizzazione

Il pannello Cluster è l'interfaccia grafica del modulo ARTEMIS e serve, come già detto, alla costruzione dei cluster, nei quali vengono raggruppate le classi locali dei vari schemi che devono essere integrati.

Una volta settate le variabili di ARTEMIS si procede premendo il pulsante CREATE GLOBAL CLASSES che fa eseguire al sistema il modulo ARTEMIS e fornisce i cluster

costruiti secondo le specifiche.

Il risultato viene presentato nella finestra di sinistra (Figura 49). Su questo risultato il progettista può intervenire con i pulsanti che si trovano al centro, per una modifica manuale.

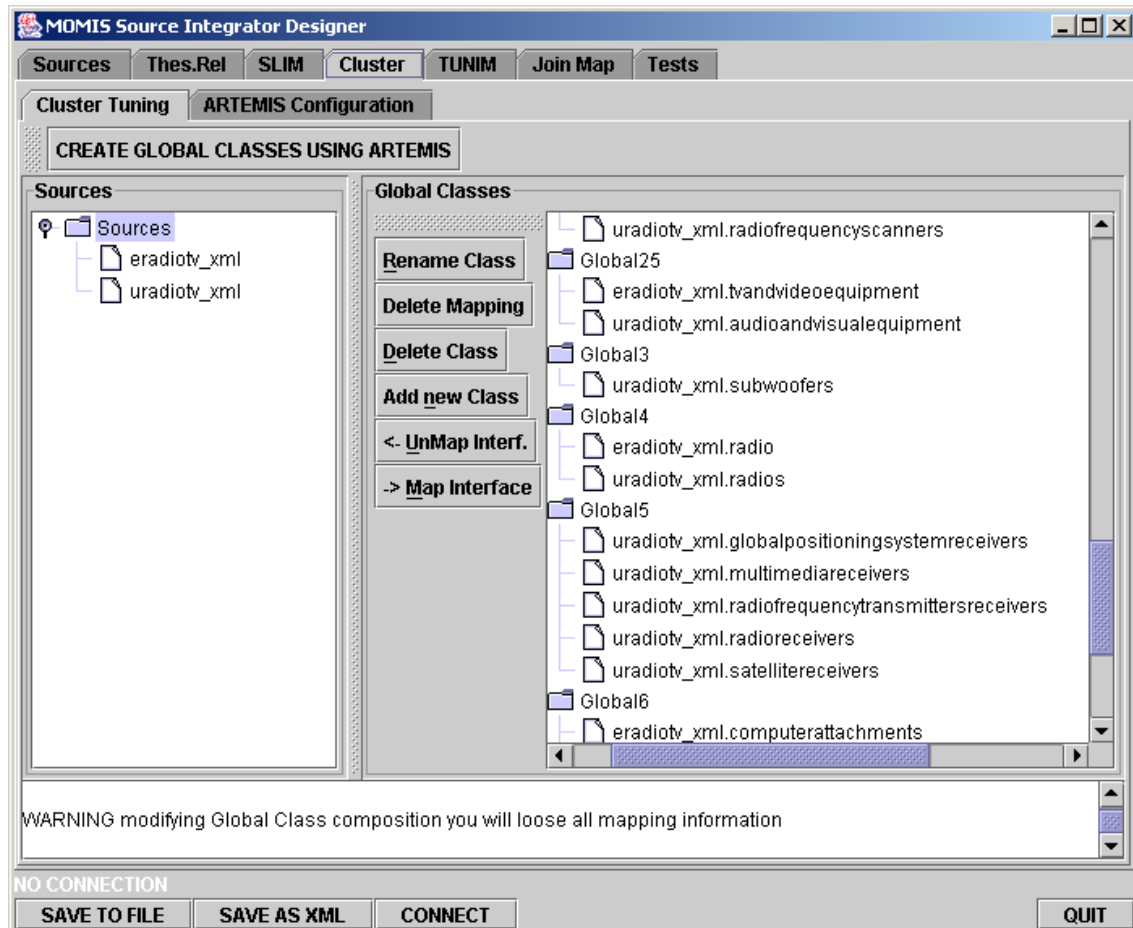


Figura 49 Esempio di visualizzazione del Global Schema

Nell'esempio RadioTV la clusterizzazione ha prodotto 26 classi globali, visualizzate di seguito:

Global Classes:

```
Global0
  uradiotv_xml.equalizers
Global1
  eradiotv_xml.tanslator
Global2
  uradiotv_xml.cassetteplayersorrecorders
  uradiotv_xml.combinationtelevisionvideocassetterecorders
  uradiotv_xml.compactdiskplayersorrecorders
  uradiotv_xml.digitalvideodiskplayersorrecorders
  uradiotv_xml.videocassetteplayersorrecorders
  uradiotv_xml.videocassetterecorders
Global3
  uradiotv_xml.subwoofers
Global4
```

```

    eradiotv_xml.radio
    uradiotv_xml.radios
Global5
    uradiotv_xml.globalpositioningsystemreceivers
    uradiotv_xml.multimediareceivers
    uradiotv_xml.radiofrequencytransmittersreceivers
    uradiotv_xml.radioreceivers
    uradiotv_xml.satellitereceivers
Global6
    eradiotv_xml.computerattachments
    eradiotv_xml.mobilephoneattachments
Global7
    uradiotv_xml.headphones
Global8
    eradiotv_xml.alarmclock
    eradiotv_xml.worldclock
    uradiotv_xml.clockradios
Global9
    uradiotv_xml.televisions
Global10
    eradiotv_xml.adressdatabank
Global11
    uradiotv_xml.speakers
Global12
    eradiotv_xml.calculator
Global13
    uradiotv_xml.speakers
Global14
    uradiotv_xml.microphones
Global15
    uradiotv_xml.remotecontrol
Global16
    uradiotv_xml.promotionarticle
Global17
    eradiotv_xml.advertising
Global18
    uradiotv_xml.homestereosystem
    uradiotv_xml.portablestereosystem
    eradiotv_xml.stereounits
Global19
    eradiotv_xml.promotgiftselectronicsdatabanks
Global20
    eradiotv_xml.cd
Global21
    uradiotv_xml.domesticappliancesandsuppliesandconsumerelectronic
products
Global22
    uradiotv_xml.consumerelectronics
Global23
    uradiotv_xml.compactdiscplayers
    uradiotv_xml.digitalvideodiscplayers
    uradiotv_xml.laserdiscplayers
Global24
    uradiotv_xml.radiofrequencyscanners
Global25
    eradiotv_xml.tvandvideoequipment
    uradiotv_xml.audioandvisualequipment

```

6.3.5.7 Modulo TUNIM

Questo modulo gestisce l'ultima fase del procedimento di integrazione per la creazione dello schema globale. Partendo dalle relazioni del Common Thesaurus, per ciascuno dei cluster individuati da ARTEMIS viene creata una classe globale. Ciascuna classe globale è caratterizzata da un insieme di attributi globali e una mapping table: l'insieme di attributi ne definisce la struttura mentre, la mapping table indica quali informazioni rappresenta ogni attributo globale.

La creazione delle mapping table conclude il processo di integrazione tra i codici sorgente. In appendice B si riporta il listato del file XML generato da MOMIS come risultato dell'integrazione.

6.3.6 Mapping risultato tra i frammenti di codice sorgente

MOMIS, date in input le sorgenti di dati, crea uno schema globale di integrazione che fornisce il matching tra le porzioni di codice degli standard utilizzati.

L'esempio mostra come l'integrazione avvenga in modo semiautomatico e pertanto evidenzia quanto sia importante il contributo del progettista, sia per le relazioni intra-schema che per quelle inter-schema; nella fase di clusterizzazione, oltre alla variazione dei parametri di ARTEMIS (poco significativi nel caso di content standard), è possibile controllare lo schema globale ottenuto e ritornare ai moduli precedenti per variare alcune impostazioni.

La differenza tra gli standard impone molta accuratezza nella preparazione e nella verifica dello schema globale; in questo senso il dominio RadioTV è stato scelto come esempio generale della procedura di integrazione, perché esemplifica un'evidenza sperimentale ritornata di frequente nelle prove effettuate con MOMIS sui codici forniti dagli standard: la numerosità e il dettaglio delle commodity di UNSPSC in confronto all'ultimo livello di ecl@ss.

Nella fase di assegnazione delle forme base si è cercato di uniformare i dettagli che distinguevano fra loro alcuni codici di ultimo livello affinché MOMIS potesse costruire alcune classi globali ampie, piuttosto che classi monocodice. Il risultato è stato ad esempio quello di avere due grandi classi, una che raccoglieva tutti i "recorder" (registratori), l'altra i "player" (lettori) musicali, dividendo pertanto tra funzione di

registrazione del suono e funzione di riproduzione. Si sarebbe potuto distinguere per modalità di riproduzione e registrazione, raggruppando quindi tutti i codici relativi a “compact disc”, quelli relativi a “cassette”,...; oppure per componenti associate a una stessa tipologia di vendita, riunendo prodotti come “equalizers”, “subwoofer”, ...
Segue la visualizzazione del risultato ottenuto dall’integrazione dei frammenti di codice relativi all’esempio RadioTV (Figura 50).

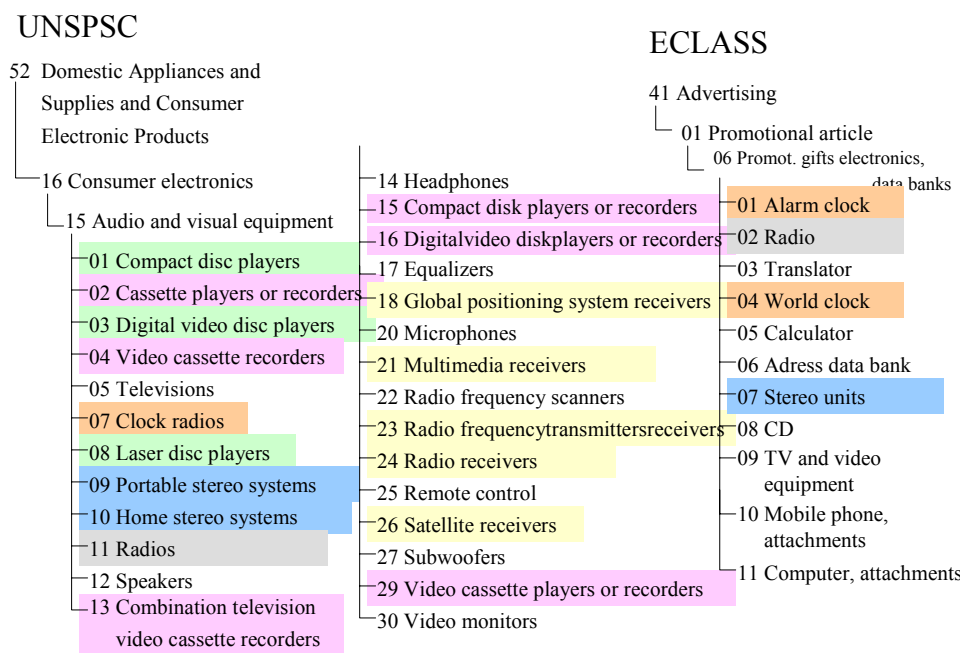


Figura 50 Mapping finale dell’esempio RadioTV

L’obiettivo che deve guidare il progettista nelle decisioni metodologiche è quello di ottenere un mapping che integri nel migliore modo possibile gli standard di classificazione prodotti considerati.

Nel prossimo capitolo entrerà nel dettaglio delle scelte progettuali utilizzate nei test, al fine di proporre alcune risoluzioni a problemi di mismatching (ovvero no matching) evidenziati in fase sperimentale e descrivere una metodologia di integrazione.

Capitolo 7

Individuazione di fattori critici e proposte metodologiche

Dalla letteratura a monte e dall'esperienza acquisita durante le prove sono emerse numerose indicazioni interessanti per individuare alcuni fattori critici di successo del progetto e su questi costruire un apporto originale.

L'ultima parte della trattazione si pone pertanto l'obiettivo di sintetizzare tutti gli elementi raccolti in una formulazione di osservazioni e idee, volte a proporre una nuova metodologia da seguire nell'integrazione dei codici di classificazione prodotto per l'E-Commerce.

Come punto di riferimento per le considerazioni successive occorre ricordare che i marketplace elettronici possono essere arene imparziali in cui fare incontrare domanda e offerta di diversi compratori e venditori solo se si mette a punto un'ontologia di prodotto comune. Per rispondere a tale urgenza OntoWeb, nella proposta lanciata con il Challenge, indica, come soluzione preferibile, l'integrazione tra gli standard di classificazione prodotto esistenti; da questa idea ha mosso i primi passi il tentativo di effettuare l'integrazione con il sistema MOMIS.

La numerosità delle prove effettuate, nonostante alcune fossero molto diverse fra loro, ha messo in luce la ripetitività di certi fattori riconducibili da un lato a caratteristiche del

sistema utilizzato, dall'altro ai codici di classificazione. Nella prima parte del capitolo evidenzierò questi fattori, nella seconda mi occuperò delle soluzioni.

7.1 Fattori critici nell'integrazione con MOMIS

Lo svolgimento dei test di integrazione suggerisce l'individuazione di alcuni aspetti ricorsivi interessanti per descrivere in modo completo le caratteristiche del mapping tra i codici effettuato con il sistema MOMIS.

Si possono schematizzare in due tipologie:

1. Fattori riconducibili al sistema;
2. Fattori riconducibili agli standard.

I primi comprendono i problemi ontologici relativi all'uso del database lessicale WordNet e i problemi metodologici relativi al fondamentale contributo del progettista.

I secondi evidenziano i punti deboli delle codifiche in relazione al loro impiego su MOMIS; in particolare raccolgono da un lato le difficoltà di gestione della gerarchia tra le classificazioni, dall'altro la mancanza di corrispondenza per alcuni rami di codice che non trovano riscontro in entrambi gli standard integrati e rimangono così scoperti.

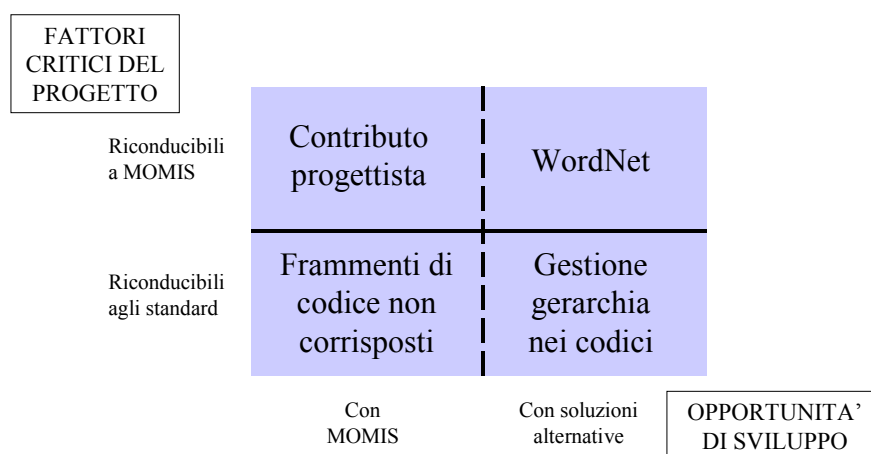


Figura 51 Fattori critici del processo di integrazione

In Figura 51 si presenta una schematizzazione degli aspetti critici appena descritti evidenziando, fra l'altro, due possibili canali di sviluppo verso cui tendere per formulare proposte migliorative:

- Fattori sfruttabili con MOMIS;
- Fattori che necessitano di soluzioni alternative.

Per chiarire questa schematizzazione utilizzerò esempi significativi emersi dalla prove su MOMIS che di volta in volta mettono in luce le diverse caratteristiche ricorsive di questa metodologia di integrazione.

7.2 Fattori riconducibili al sistema

Il sistema MOMIS è un mediatore per l'integrazione e l'estrazione di informazioni da sorgenti di dati eterogenee, del tipo relazionali, ad oggetti, semistrutturate (XML).

Tuttavia nell'utilizzarlo per integrare codici di classificazione prodotto si va oltre le specifiche rispetto alle quali è stato progettato ed è richiesto un impegno supplementare al progettista per adattarlo al nuovo scopo. Quest'esperienza può risultare utile perché da un lato mostra l'efficacia del sistema che si è rivelato adeguato alle prove e capace di adattarsi a sorgenti "non standard" (rispetto alle impostazioni di progetto), dall'altro fornisce utili indicazioni per rendere il sistema flessibile nel rispondere a una molteplicità di problematiche in ambito di Integrazione Intelligente delle Informazioni.

Come spunto per proposte migliorative da apportare all'integrazione degli standard di classificazione prodotto con MOMIS, ho focalizzato l'attenzione su due fattori:

- il lavoro del progettista nella fase di annotazione dei significati o di introduzione delle relazioni manuali, con l'obiettivo di proporre una metodologia di progetto che assicuri di prendere in considerazione, in modo completo, le porzioni di codice interessate al mapping;
- l'ontologia WordNet, utilizzata nel modulo lessicale SLIM per la creazione delle relazioni inter-schema.

7.2.1 Contributo del progettista

Per scegliere la direzione da seguire nell'indirizzare il mapping, il progettista deve tenere presente l'applicazione ultima dell'integrazione tra i frammenti di codice: l'eMarketplace. Infatti, a seconda della tipologia del mercato virtuale, è richiesta una corrispondenza tra i codici, piuttosto che un'altra.

Un marketplace verticale, a livello di prodotto, dovrebbe scendere fino alle commodity, per fare incontrare domanda e offerta di venditori e compratori molto dettagliate; al

contrario un marketplace orizzontale deve coprire un'ampia gamma di tipologie di prodotto senza necessariamente scendere troppo nei particolari.

Queste considerazioni teoriche sono sottoposte al vincolo della struttura degli standard di classificazione prodotto che, a seconda dei settori commerciali, varia in modo consistente. Il ruolo del progettista si inserisce in questa problematica: l'utente MOMIS, sulla base delle richieste generate dal mercato ritenute interessanti per il gestore di un determinato eMarketplace, deve indirizzare il sistema durante l'integrazione dei codici.

Per dimostrare che il mapping tra porzioni di codice non è soddisfacente quando l'utente si limita a seguire il sistema nell'integrazione in modo superficiale, si propongono di seguito un paio di esempi chiarificatori di prove effettuate su MOMIS.

Nel capitolo precedente [6] si è presentato come esempio generale del processo di integrazione con MOMIS il test di tipo bottom-up sul dominio RadioTV.

RadioTv è da considerarsi una prova significativa perché esemplifica un'evidenza sperimentale ritornata di frequente nel lavoro effettuato con MOMIS sui codici forniti dagli standard, vale a dire la numerosità e il dettaglio delle commodity di UNSPSC in confronto all'ultimo livello di ecl@ss. Nel capitolo [6] si forniva un possibile mapping ottenibile tra gli standard, ma si sottolineava l'importanza delle scelte del progettista per condizionare il risultato (si ricordi che l'integrazione su MOMIS si colloca, fra le modalità proposte dal Challenge, nella modalità di tipo semi-automatico).

Di conseguenza l'esempio può essere ripreso per delineare una metodologia che il progettista deve seguire per integrare le classificazioni di prodotto.

In Figura 52 si riportano nuovamente i frammenti di codice di UNSPSC e ecl@ss per il dominio RadioTV.

A colpo d'occhio si intuisce la differenza tra i due frammenti; un dominio che si manifesta in questo modo, peraltro il più comune visto durante le prove, necessita di un contributo importante del progettista per ottenere un mapping significativo. Se invece ci si adegua alle proposte di WordNet nell'assegnare i significati delle Word Form in SLIM e alle relazioni individuate in modo automatico da SIM A, nella fase di clusterizzazione si otterrebbe uno schema globale con troppe classi, poco rilevanti.

Da un'integrazione svolta approssimativamente ne consegue un mapping che esclude numerose categorie di prodotto (Figura 53), mappabili, invece, se il progettista contribuisse in modo originale all'integrazione:

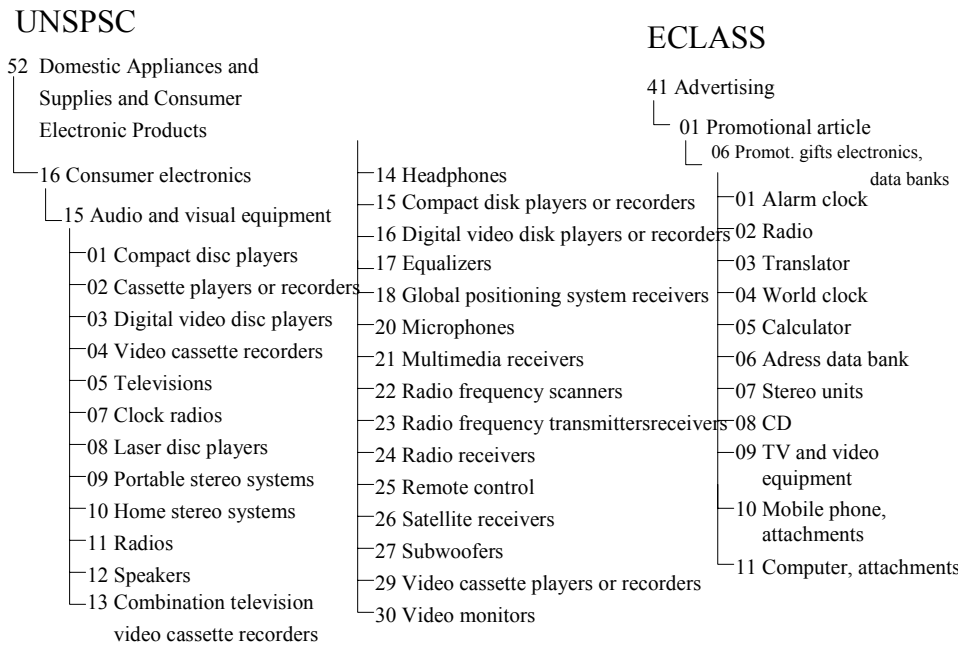


Figura 52 Frammenti dei codici UNSPSC e ecl@ss sul dominio RadioTV

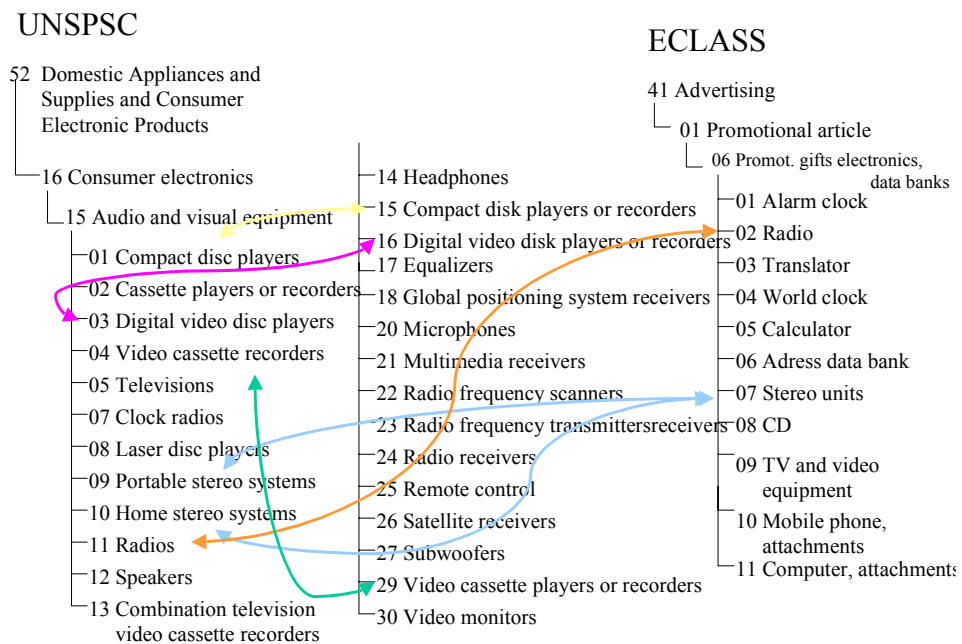


Figura 53 Mapping tra commodity “specializzate” – Esempio RadioTV

Un altro esempio che può avvalorare l'importanza del contributo del progettista è relativo al dominio Bottle.

In Figura 54 si riportano i frammenti relativi ai due codici; si può vedere che UNSPSC dettaglia il dominio in modo diverso rispetto a ecl@ss articolando progressivamente il

dettaglio sui quattro livelli; ecl@ss invece utilizza solo tre livelli, mantenendo la stessa definizione eccetto la precisazione tra parentesi relativa al materiale.

Ci si può aspettare un buon mapping tra i due codici, perché all'ultimo livello delle rispettive gerarchie il prodotto è classificato in base al materiale di fabbricazione, quindi con risultato classi che raggruppino bottiglie dello stesso materiale.

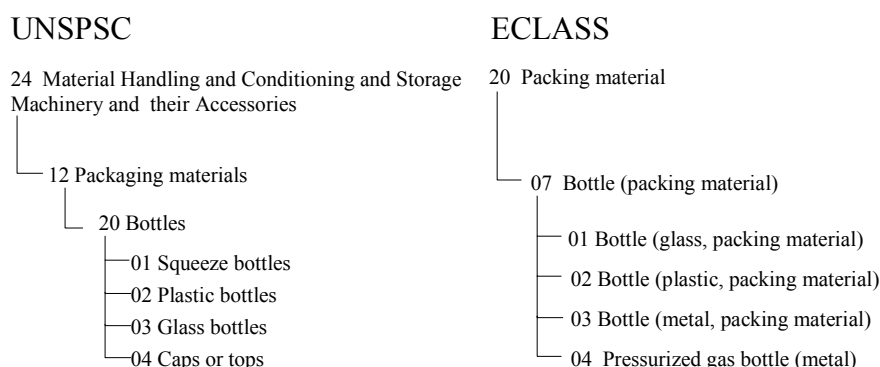


Figura 54 Frammenti di codice di UNSPSC e ecl@ss relativi al dominio Bottle

Anche in questo caso, però, per ottenere il risultato sperato è necessario coadiuvare il sistema nell'integrazione. Nel modulo SLIM, per la costruzione delle relazioni interschema, si incontra infatti qualche difficoltà nell'assegnare significati alle Word Form, poiché WordNet non distingue nel suo database lessicale tra bottiglie di diverso materiale; generalizzando le definizioni al livello superiore si avrebbe un mapping privo di senso. Affinché MOMIS riconosca le relazioni di sinonimia desiderate il progettista deve scegliere di procedere tra alcune possibilità: si impongono manualmente le relazioni nel pannello *Thes.Rel.*, si decide di assegnare come significato alle Forme Base il materiale della bottiglia (ad esempio a “plastic bottle” si assegna “plastic”).

In entrambi i casi si ottiene un risultato positivo perché MOMIS è in grado di rilevare corrispondenze uno a uno tra i codici, coerentemente con la struttura delle codifiche (Figura 55).

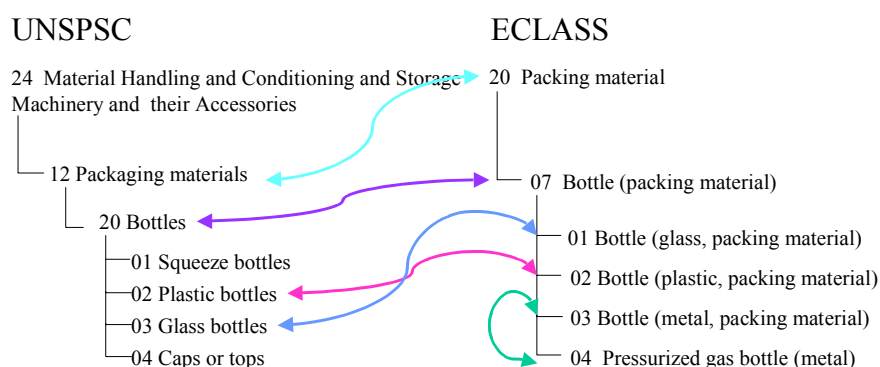


Figura 55 Mapping finale tra UNSPSC e ecl@ss sul dominio Bottle

In un esempio di piccole dimensioni, come *Bottle*, è abbastanza semplice capire quali potrebbero essere gli accorgimenti che l'utente deve adottare in fase di integrazione per avere un risultato soddisfacente. Questo non è così semplice per esempi più articolati, come il dominio *RadioTv*, oppure diventa impossibile per prove complesse fatte con approccio *Top Down*. Di conseguenza, nella fase di proposte migliorative si formalizzerà una metodologia di integrazione a step di tipo generale, adatta anche a prove lunghe e complicate.

Da queste considerazioni emerge inoltre un nuovo scenario di miglioramento: per integrare codici di prodotto, spesso molto specifici e pertanto non rilevati dal database lessicale di *MOMIS*, si potrebbe pensare a un'estensione di *WordNet*, in modo da avere un'ontologia di dominio. Nel paragrafo seguente si chiarirà in che senso *WordNet* rappresenta un fattore critico nel processo di integrazione e successivamente, nel paragrafo [7.7] si illustrerà, anche tramite alcuni esempi, la proposta di estensione di *WordNet*.

7.2.2 *WordNet*

Si è appena mostrato che per porzioni di codice ampie è indispensabile il supporto del progettista nell'indirizzare l'integrazione. Tuttavia le prove fatte hanno messo in luce un altro fattore critico, non sempre superabile con le scelte del progettista.

Spesso i domini da mappare con *MOMIS* dettano codici di prodotto troppo specializzati per il database lessicale di *WordNet* e nel caso il risultato dell'integrazione sia destinato a *eMarketplace* verticali, si producono risultati poco significativi.

Si potrebbe pensare di introdurre relazioni di sinonimia in *Thes-Rel*. per ogni corrispondenza puntuale non riconosciuta da *SLIM*, ma su esempi ampi il processo diventerebbe pressoché manuale e non interessante come proposta per il *Challenge* di *OntoWeb*.

Per presentare questa criticità si utilizza l'esempio di integrazione *Bottom-Up* sul dominio *Cooker*. In *Figura 56* e *Figura 57* si nota che le due porzioni di codice scelte si presentano abbastanza simili: i prodotti esaminati sono quasi gli stessi per un numero di voci da integrare paragonabile.

I termini sono molto specifici, in diversi casi è difficile ricondurli ad una Forma Base comprensibile per WordNet. Entrambi i domini codificano fra le commodity alcuni elettrodomestici non generici, ma relativi alla preparazione o cottura di un unico alimento (evidenziati in Figura 57). Questi termini non sono presenti nel dizionario, mentre sarebbe importante prenderli in considerazione perché caratteristici di tutte le codifiche. Il più delle volte invece il progettista dovrebbe scegliere l'opzione *Ignored Term*, perché vi sono voci sconosciute a WordNet.

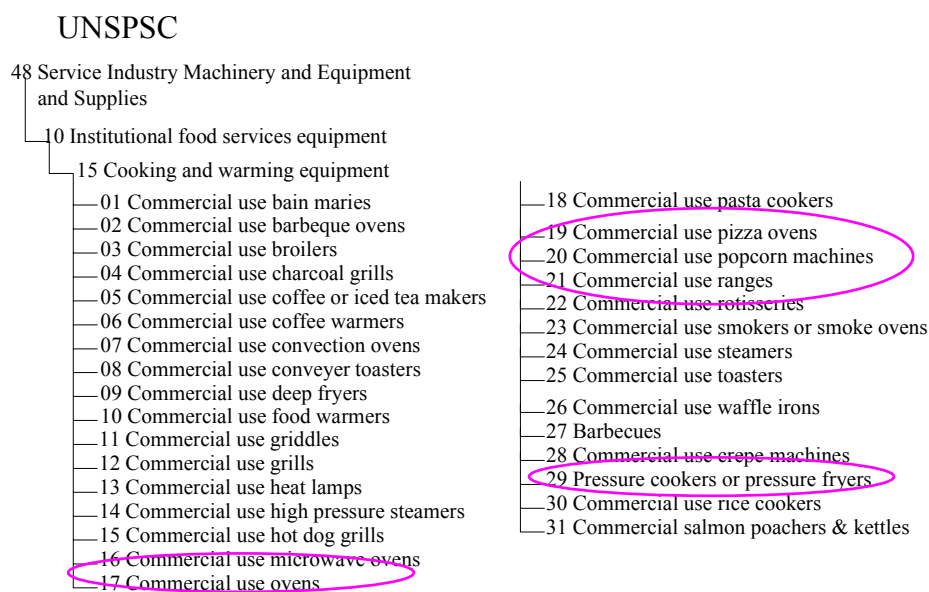


Figura 56 Frammento di codice UNSPSC sul dominio Cooker

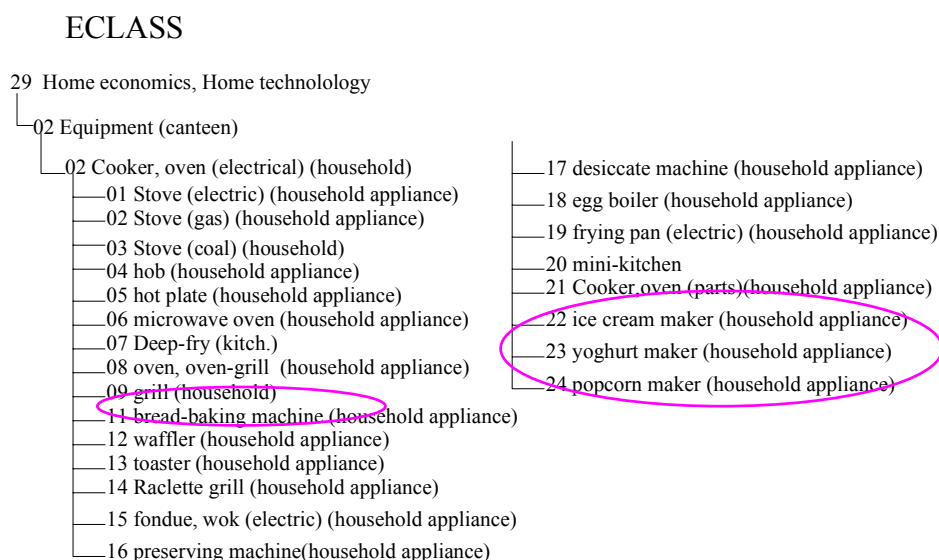


Figura 57 Frammento di codice ecl@ss sul dominio Cooker

Per ottenere comunque un mapping tra i frammenti e per ovviare a questo problema si è scelto, in alcuni casi, di sostituire l'elettrodomestico con il prodotto che esso prepara o

cuoce, per esempio: al posto di “popcorn machines”, che non viene identificato da WordNet si è scritto “popcorn” o al posto di “Ice-cream maker” si è scelta la Word Form “Ice-cream”. Tale soluzione è stata possibile perché il dominio Cooker mappato non prendeva in considerazione i prodotti alimentari stessi, nel qual caso l’operazione non sarebbe stata ammissibile per ovvie ragioni di coerenza. Adottando questa convenzione si sono ottenuti dei buoni risultati, dato che è stato possibile adattare all’uso WordNet, senza ricorrere alla lunga fase di inserimento manuale di relazioni.

Il test di integrazione sul dominio Cooker vuole comunque esemplificare un problema, che, volta per volta, si potrebbe superare con scelte coraggiose del progettista, ma, data la varietà dei codici classificati dagli standard, si rischierebbe di sovrapporre codici introdotti invece con lo scopo di specificare funzioni diverse. Vi potrebbero essere marketplace dedicati alla grande distribuzione e non alla compravendita di materie prime e accessori, o marketplace verticali che trattano tutti i codici relativi a un settore, che troverebbero totalmente inutile l’utilizzo di uno schema di integrazione tra standard non progettato per distinguere tra prodotti e attrezzature con cui fabbricare i prodotti stessi. Pertanto potrebbe essere utile per gli sviluppatori di MOMIS prendere in considerazione un ampliamento dell’ontologia WordNet, o addirittura l’avvicendamento di un’altra ontologia, nel caso si ritenesse importante continuare a testare MOMIS su problematiche relative ai codici di classificazione prodotti per il commercio elettronico. Nella seconda parte del capitolo si prenderanno in considerazione entrambe le possibilità.

7.3 Fattori riconducibili agli standard

I diversi standard di classificazione prodotto considerati sono nati il più delle volte da iniziative di enti o consorzi, sponsorizzati da particolari attori economici, generalmente imprese private che cercano una risposta a proprie esigenze industriali o commerciali. Pertanto le diversità fra i codici individuate nel capitolo [6] sono soprattutto riconducibili alle differenti spinte propulsive che hanno dato vita alle codifiche.

Queste caratteristiche non combacianti tra i codici si scontrano con il processo di integrazione e diventano fattori critici importanti al fine di ottenere un buon mapping tra le porzioni di codice.

Ecl@ss e UNSPSC, gli standard utilizzati effettivamente nel progetto, presentano un diverso grado di dettaglio nello sviluppo della descrizione gerarchica di un segmento di prodotti. Inoltre UNSPSC unisce due elementi importanti per la completezza di uno standard: l'ampiezza degli standard orizzontali e la profondità di dettaglio degli standard verticali. Al contrario ecl@ss pone l'attenzione soprattutto su pochi settori, quelli che hanno sostenuto il progetto di realizzazione della codifica; pertanto assume la struttura di standard verticale, orientato a settori quali l'elettrico e il chimico, ma del tutto assente su altri importanti segmenti.

Queste diversità hanno comportato alcuni problemi nella fase sperimentale di integrazione, in particolare durante le prove su MOMIS si sono individuati due fattori critici relativi agli standard:

- Frammenti di codice individuati in uno standard che non trovano corrispondenza nella porzione di codice dell'altra codifica considerata e che pertanto non sono interessati dal mapping, ma restano completamente esclusi;
- Gestione della gerarchia spesso impiegata in modi differenti dalle codifiche, sia a livello di prodotto (un livello gerarchico definisce una categoria di prodotti allo stesso modo del corrispondente livello dell'altro standard, ma nel dettaglio si declinano codici di prodotto completamente diversi), sia a livello di struttura (si individua una possibile corrispondenza tra frammenti di codice, ma, procedendo, si scopre che presentano una struttura interna inadeguata al matching).

7.3.1 Frammenti di codice non corrisposti

Dallo studio degli standard si individuano frammenti di codice che presentano similarità sia in ecl@ss che in UNSPSC e sembrano prestarsi per una procedura di integrazione con approccio bottom-up [6.3.2.1] o si selezionano segmenti corrispondenti, adatti all'integrazione di tipo top-down [6.3.2.2]. Si sono effettuate numerose prove in entrambe le modalità, ma non sempre il risultato ottenuto è stato interessante.

Il test di tipo bottom-up sul dominio Book esemplifica i problemi di corrispondenza riscontrabili su prove di piccole dimensioni. Nello schema dei due frammenti di codice

selezionati per l'integrazione (Figura 58) si nota la diversa definizione dei codici di prodotto relativi a Book.

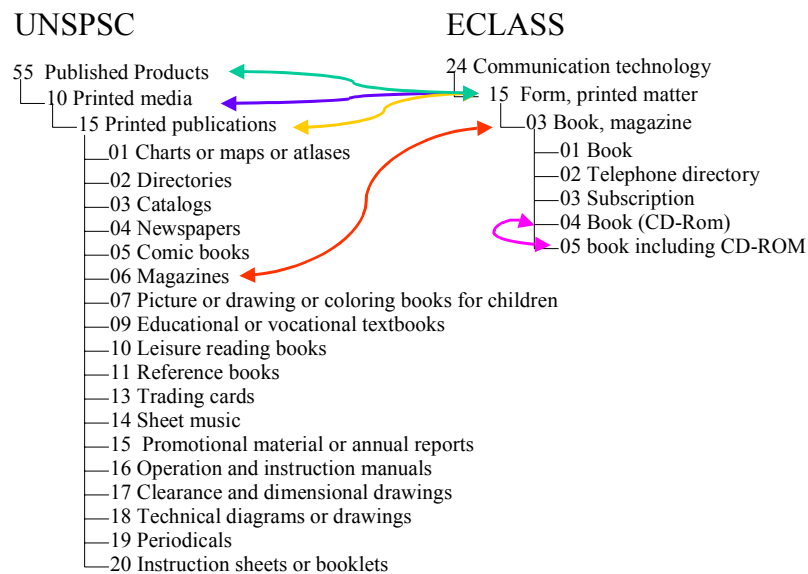


Figura 58 Test di tipo bottom-up sul dominio Book

Limitando l'integrazione con MOMIS alla procedura standard di corretta assegnazione delle Forme Base per i termini riconoscibili da Word Net, si avrebbero numerosi codici di UNSPSC senza corrispondenza in ecl@ss. I risultati dell'integrazione non sono buoni, l'ostacolo al raggiungimento di un mapping di qualità è posto fondamentalmente dalle codifiche. Le poche corrispondenze rilevate si trovano a livelli alti della gerarchia, quindi non sono particolarmente significative.

Il diverso grado di dettaglio nella descrizione delle classi di prodotto fra UNSPSC e ecl@ss è un elemento costante incontrato nel lavoro di integrazione; per raggiungere l'obiettivo di creare una corrispondenza di codici prodotto tra standard, si è superato il problema procedendo a passi successivi, modificando strada facendo le ipotesi di relazione tra i dati caricati su MOMIS, ripercorrendo a ritroso le fasi di integrazione.

Riprendendo l'esempio Book si potrebbero generalizzare alcuni codici prodotto per costruire classi più ampie nello Schema globale, mettendo in un'unica classe tutte le tipologie di libro proposte in UNSPSC e associabili in ecl@ss al generico prodotto "book" o raggruppando "magazine" in ecl@ss con "newspapers", "periodicals" e "magazines" in UNSPSC.

Una tale risposta alle difficoltà sorte nell'integrazione di tipo bottom-up, a livello di content standard, si può ricondurre al contributo del progettista formalizzato nella seconda parte del capitolo nella proposta di una metodologia procedurale.

Di diversa complessità e conseguente trattazione sono i problemi riconducibili agli standard di classificazione quando si integra con approccio top-down.

Nel considerare porzioni di codice adatte all'integrazione, sulla base di corrispondenze fra segmenti, si analizzano vaste parti di codice che potenzialmente possono essere integrate. Tuttavia il vincolo delle dimensioni non permette di conoscere completamente a priori i codici prodotto che rientrano nell'integrazione e può capitare, nel momento in cui si ottiene lo Schema Globale su MOMIS, che vi siano frammenti di codice esclusi dal mapping, perché relativi a categorie di prodotto non considerate da entrambe le codifiche in quel segmento o addirittura trascurate in uno dei due standard.

Si considera a tale proposito l'esempio sul dominio Household tra i segmenti "Domestic Appliances and Supplies and Consumer Electronic Products" di UNSPSC (Figura 59) e "Home economics and Home technology" di ecl@ss (Figura 60).

L'obiettivo dell'esempio era quello di integrare in modo completo, dal primo all'ultimo livello, due parti di codifica fra loro simili e verificare l'adeguatezza di MOMIS nel gestire un numero importante di codici e relazioni. Negli schemi in Figura 59 e 60 si riportano per chiarezza solo i primi tre livelli dei due standard, ma sono sufficienti per notare alcune divergenze strutturali.

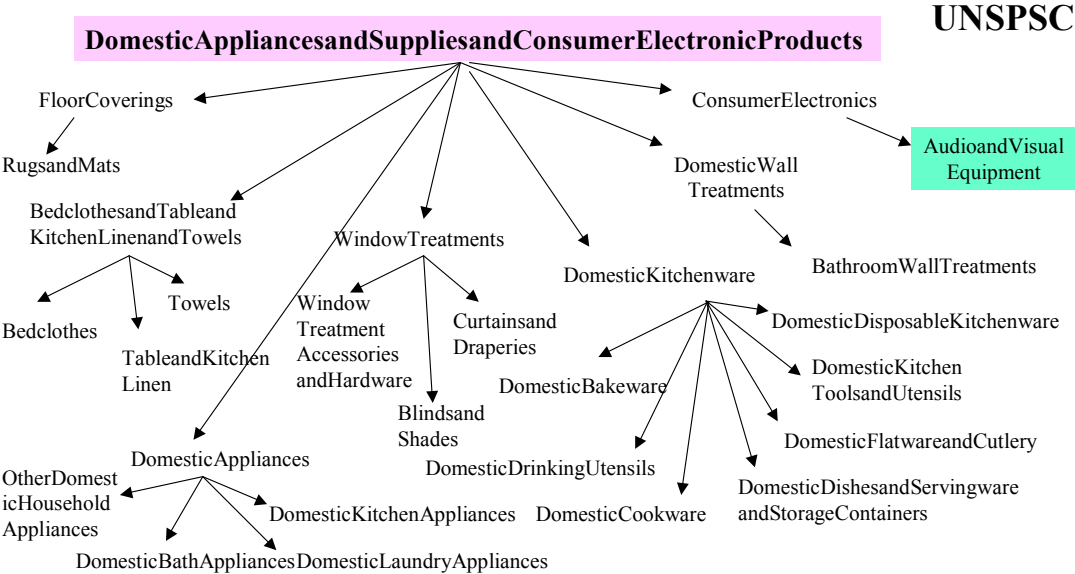


Figura 59 Schema del dominio Household in UNSPSC fino al terzo livello gerarchico

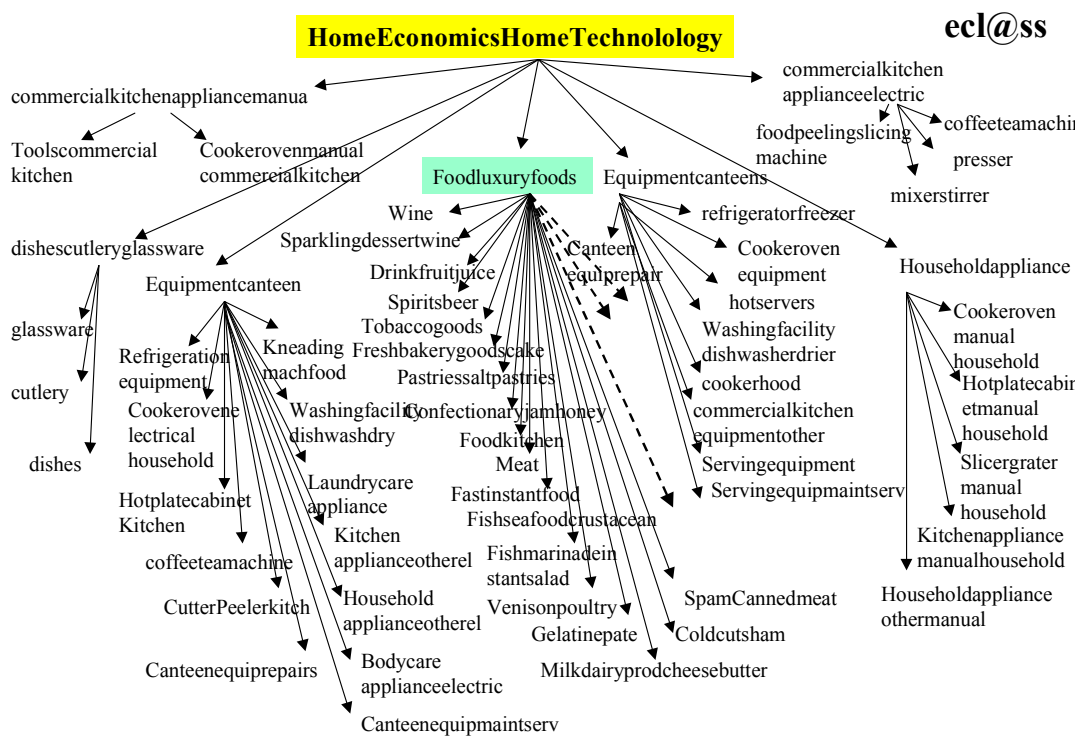


Figura 60 Schema del dominio Household in ecl@ss fino al terzo livello gerarchico

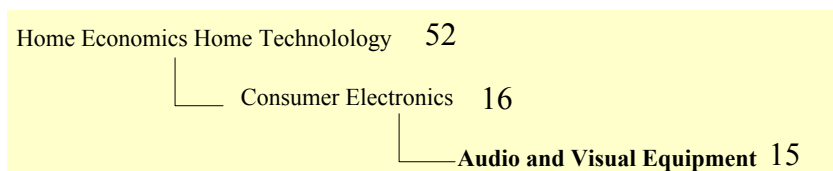
Le due codifiche, dal primo al quarto livello, contengono rispettivamente 185 e 260 voci di codice; pertanto la diversa numerosità di prodotti dovrebbe prospettare problemi di mapping in alcune parti delle ampie porzioni considerate, già riscontrabili nella scrittura dei file DTD, quando cioè si passano in rassegna entrambi i segmenti in esame.

Un esempio di queste proporzioni non evidenzia solo i problemi di integrazione riconducibili agli standard, ma raccoglie tutte le problematiche affrontate nelle prove. D'altro canto, se inserito nelle verifiche finali, mette alla prova il progettista nel considerare i fattori critici precedentemente individuati e testa la validità del sistema MOMIS nell'integrare gli standard di classificazione prodotto gestendo un numero elevato di codici.

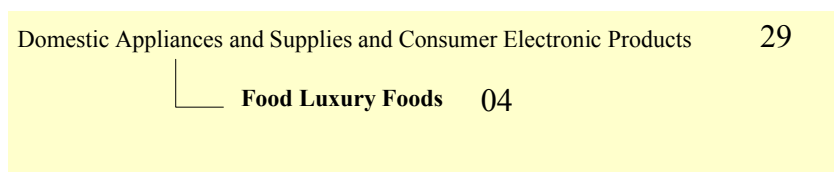
Dalla dimensione della prova ci si aspetta un Global Schema con un numero elevato di classi globali. Durante il processo di integrazione si incontrano alcuni problemi: ecl@ss include nel segmento Home Technology i codici relativi agli alimenti e nel contempo codifica gli strumenti e gli accessori atti alla loro preparazione. Si manifesta così un primo ostacolo nel cammino di mappatura del progettista; infatti il database lessicale WordNet, come descritto nel paragrafo precedente, non comprende fra i suoi significati quelli relativi alla preparazione o cottura degli alimenti. Tuttavia non è possibile

ricorrere all'espedito utilizzato nel test di tipo Bottom-up sul dominio Cooker, in cui si etichettavano gli elettrodomestici da cucina con i cibi preparati da ciascuno, perché si originerebbero delle sovrapposizioni di codici non opportune. D'altro canto anche la costruzione manuale di relazioni di sinonimia prevede un contributo del progettista importante in termini di tempo. Si è scelto comunque di seguire questa seconda strada per continuare l'integrazione e potere studiare il comportamento di MOMIS in una prova così vasta.

Dalla lettura dello Schema Globale così ottenuto si vede che vi sono due frammenti di codice, uno per UNSPSC e l'altro per ecl@ss che non hanno corrispondenza in entrambi gli standard. UNSPSC comprende nel segmento relativo a Household il dominio RadioTv, chiamato "Audio and Visual Equipment":



Poiché in ecl@ss questo dominio è codificato in un altro segmento, MOMIS non è in grado di mappare i codici al quarto livello di UNSPSC relativi alla classe "Audio and Visual Equipment" con alcun codice di ecl@ss. Lo stesso vale, a ruoli rovesciati, per il dominio Food che è descritto da ecl@ss nel segmento HomeTecnology, ma non è presente nel segmento corrispondente di UNSPSC.



In sintesi la prova di tipo top-down sul dominio Household evidenzia da un lato la flessibilità di MOMIS nell'affrontare problemi di integrazione complessi e alternativi agli obiettivi di progetto, nonostante la complessità vada a discapito della velocità del processo di integrazione; dall'altro mette in luce un fattore critico importante al fine di studiare in modo complessivo l'integrazione tra standard di classificazione: la diversità tra le codifiche comporta l'individuazione di frammenti di codice che non hanno corrispondenza in tutti gli standard considerati e si creano di conseguenza dei "buchi" nel mapping finale.

7.3.2 Gestione della gerarchia nei codici

Le diverse origini degli standard di classificazione hanno reso difficile, all'inizio della fase sperimentale, l'interpretazione e la navigazione tra i codici. Le prove di integrazione tra content standard realizzate con MOMIS sono state condizionate dai differenti punti di vista ed esigenze commerciali alla base delle codifiche.

Per scorrere gli standard di classificazione alla ricerca di punti comuni da cui cominciare l'integrazione e superare i problemi iniziali, dopo avere scaricato i codici forniti su Internet dai siti ufficiali, si è scelto di utilizzare da un lato i motori di ricerca messi in rete dai consorzi gestori, dall'altro si sono esportati gli standard su database per creare interrogazioni sui codici e lavorare anche off-line.

Motori di ricerca per codici prodotto

Per "navigare" tra le codifiche e individuare possibili corrispondenze tra porzioni di codice si sono utilizzati in primo luogo i motori di ricerca messi a disposizione on-line dai gestori del codice. Questo servizio è stato molto utile soprattutto per le prove di integrazione con approccio bottom-up.

Ecl@ss nel sito ufficiale³² ha dedicato una pagina alla ricerca (*Search*) dei codici prodotto (Figura 61).

L'utente può inserire la denominazione di un prodotto e il motore di ricerca individua tutti i codici presenti nella codifica associabili al termine desiderato.

Inoltre il servizio permette di splittare il codice ritenuto più interessante per l'utente e visualizzare, per quel codice, il percorso gerarchico relativo.

Un servizio analogo è stato messo a disposizione per UNSPSC da UNDP (United Nations Development Programme)³³, che è il nuovo gestore ufficiale della codifica, sostituendosi a ECCMA, ma fin da prima metteva a disposizione uno strumento di ricerca dei prodotti.

Nella home page di UNDP vi è una sezione "Search for Code" che permette di organizzare la ricerca per nome del prodotto o numero di classificazione.

³² www.eclass.de/

³³ www.un-spsc.net/

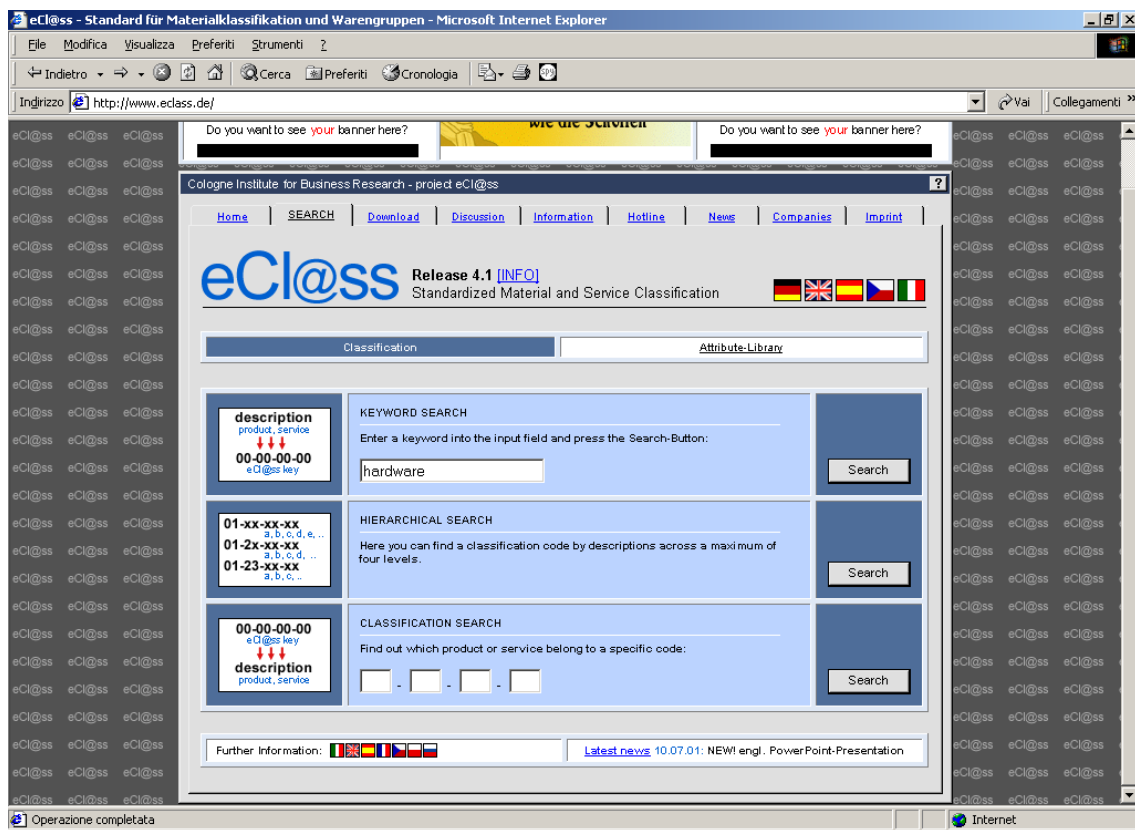


Figura 61 Ricerca on-line su ecl@ss dei codici prodotto

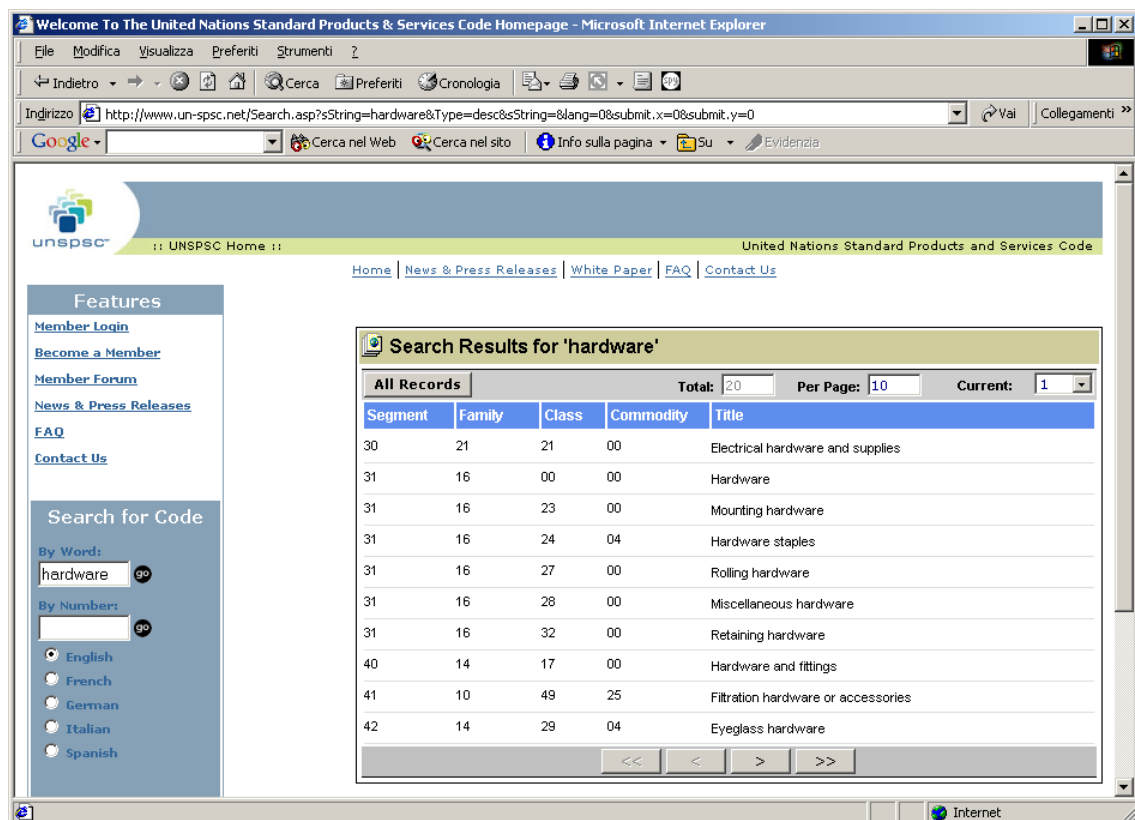


Figura 62 Ricerca on-line su UNSPSC dei codici prodotto

Nel primo caso si ottiene una maschera che riporta tutti i codici prodotto relativi alla parola chiave inserita (Figura 62), nel secondo il nome del prodotto associato al codice prescelto e tutti i figli, se si introduce un codice incompleto, fino al livello di commodity.

I motori di ricerca offerti on-line mettono a disposizione le codifiche sempre costantemente aggiornate. Tuttavia vi sono momenti in cui sono necessarie letture più complessive dei codici, possibili soltanto consultando le versioni complete delle codifiche scaricabili dai siti. Per questa esigenza sono stati raccolti i file relativi alle singole codifiche in un unico database di Microsoft Access.

Database di consultazione dei codici

Le versioni degli standard utilizzate nelle prove sono state scaricate dai siti ufficiali. Ecl@ss fornisce il codice in formato csv (*comma separated value*), UNDP in formato foglio di calcolo di Microsoft Excel.

I due file sono stati importati da Microsoft Access in due tabelle distinte riportando i campi relativi a:

- Chiave identificatrice (identifier)
- Codice prodotto (class)
- Nome prodotto (preferred name)
- Livello gerarchico (level)

Per il lavoro di supporto ai motori di ricerca convenzionali disponibili on-line ci si è limitati a costruire una semplice interfaccia per la consultazione delle tabelle, visualizzabili anche parzialmente mediante dei filtri, e per la ricerca di prodotto con query di selezione, come la seguente:

```
SELECT *  
FROM UNSPSC  
WHERE (([UNSPSC].[preferred name]) Like "**bearing*");
```

Tuttavia vi sono alcuni esempi che mostrano come questi strumenti, comunque fondamentali nello studio delle codifiche, non siano stati sufficienti al superamento dei problemi riconducibili alla diversa costruzione degli standard e legati alla selezione di frammenti adatti all'integrazione

Infatti la ricerca ha permesso di individuare domini potenzialmente interessanti per il mapping, ma il successivo riscontro nelle codifiche ha evidenziato differenze tali da

rendere poco significativa la realizzazione della prova.

Per semplificare la problematica, si può parlare di un unico fattore critico relativo alla gestione della gerarchia, spesso impiegata in modi differenti dalle codifiche, con due tipologie di variabilità:

1. **variabilità di prodotto**, quando un livello gerarchico definisce una categoria di prodotti allo stesso modo del corrispondente livello gerarchico dell'altro standard, ma nel dettaglio si specificano codici di prodotto completamente diversi;
2. **variabilità di struttura**, quando si individua una possibile corrispondenza tra frammenti di codice, ma, procedendo, si scopre che presentano una struttura interna inadeguata al mapping.

L'esempio di potenziale integrazione sul dominio Hardware chiarisce la difficile gestione della gerarchia nel caso di variabilità di prodotto.

Infatti con l'aiuto dei motori di ricerca si scelgono le porzioni di codice che sviluppano questo dominio; quindi, con la consultazione off-line degli standard, attraverso tecniche di tipo *roll-up* e *drill-down* (descritte nella sezione [3.5]), si stabilisce la dimensione del frammento, a partire dal codice selezionato.

Non appena si definisce l'appropriato dominio di integrazione o lo si schematizza ci si accorge dell'assoluta diversità di descrizione del dominio in esame fra le due codifiche (Figura 63).

UNSPSC	ECLASS
31 Manufacturing Components and Supplies	24 Communication technology, office -
└ 16 Hardware	└ 01 Hardware (inform. technology)
├ 15 Screws	├ 01 Hardware (mainframe system)
├ 16 Bolts	├ 02 Hardware (server system)
├ 17 Nuts	├ 03 Hardware (workstation)
├ 18 Washers	├ 04 Printer, plotter, scanner (parts)
├ 19 Springs	├ 05 Network components (parts)
├ 20 Nails	├ 06 Screen
├ 21 Anchors	├ 07 Periph. equip, accessories (PC)
├ 22 Rivets	├ 08 Harddisk
├ 23 Mounting hardware	├ 09 Computer (portable)
├ 24 Miscellaneous fasteners	├ 98 Hardware (maint., service)
├ 25 Brackets and braces	├ 99 Hardware (repair)
├ 26 Hooks	
├ 27 Rolling hardware	
├ 28 Miscellaneous hardware	
├ 29 Clamps	
├ 30 Couplings	
├ 31 Connectors	
├ 32 Retaining hardware	
└ 33 Spacer	

Figura 63 Frammenti di codice relativi al dominio Hardware

Infatti UNSPSC definisce con Hardware una famiglia di prodotti relativi al segmento “Manufacturing Components and Supplies” (componentistica meccanica), ecl@ss si riferisce a una famiglia di prodotti all’interno di “Communication Technology” (componenti per l’ICT). Naturalmente ci si è fermati a questo livello di studio, visto che l’integrazione non avrebbe prodotto alcun mapping.

Per descrivere il problema di gestione della gerarchia nel caso di variabilità di struttura si fa riferimento all’esempio individuato con approccio bottom-up sul dominio Software (Figura 64)

In questo caso entrambi gli standard si riferiscono, con il nome “software”, alla stessa famiglia di prodotti, ma si riscontrano diversità di struttura.

UNSPSC sceglie di affrontare il dominio in modo approfondito, dedicando un codice prodotto per ciascuno dei principali strumenti software; al contrario ecl@ss, scegliendo di non dettagliare il dominio, gestisce in modo del tutto particolare le potenzialità della gerarchia (questa scelta si riscontrata in varie parti della codifica e la si può ricondurre alle ipotesi costruttive che hanno mosso i suoi progettisti, vale a dire di curare solo alcuni segmenti) e a livello di commodity decide di non differenziare il dominio Software, ma di precisare solamente qualche servizio associato.

Anche per questo secondo esempio non si è proceduto nell’integrazione.

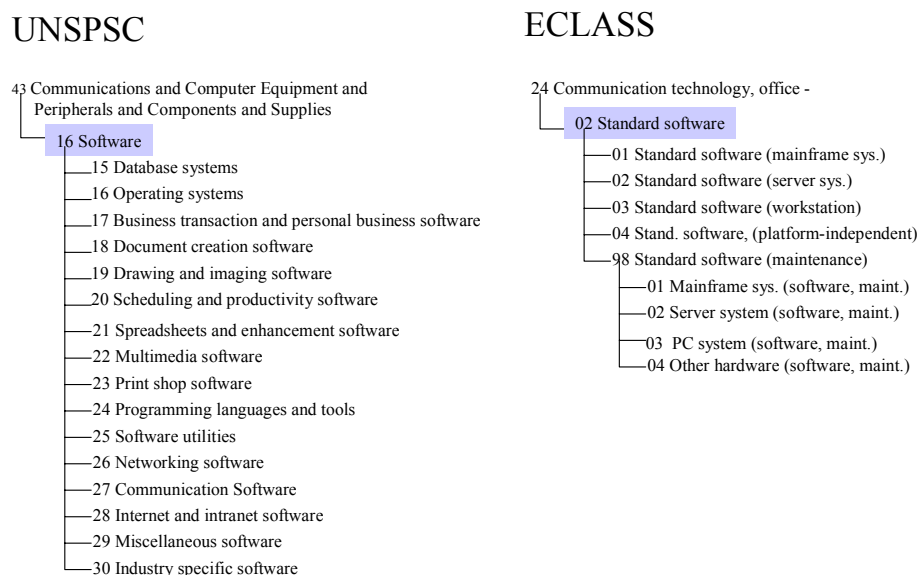


Figura 64 Frammenti di codice relativi al dominio Software

Quest’ultima tipologia di fattori che rendono critico il processo di integrazione tra content standard, riconducibile alle diversità tra UNSPSC e ecl@ss, si aggiunge ad altre problematiche descritte nei paragrafi precedenti.

Si può concludere che dalle numerose prove di integrazione sono emerse parecchie indicazioni per migliorare il processo. In particolare, se ci riferisce ai fattori riconducibili agli standard, vi è materiale a sufficienza per proporre raccomandazioni per la costruzione di futuri standard di classificazione prodotti per l'e-commerce, come suggerito dal Challenge di OntoWeb nelle regole di ricerca formalizzate in [Schulten, 01]. Queste considerazioni saranno argomento dell'ultima parte del capitolo.

7.4 Opportunità di sviluppo nell'integrazione tra codici con MOMIS

La seconda parte del capitolo è dedicata alle soluzioni migliorative che i test su MOMIS hanno suggerito nell'ambito dello studio dell'integrazione di content standard.

MOMIS si è dimostrato un sistema molto versatile nel trattare le sorgenti di dati e pertanto rappresenta la prima risorsa da cui trarre elementi per migliorare il processo di integrazione. Le numerose prove effettuate su questo sistema, hanno permesso, anche senza conoscere nel dettaglio le specifiche di progetto che vi sono alla base, di apprezzare i tool messi a disposizione dell'utente per integrare i dati.

SI-Designer consente al progettista di muoversi all'interno dell'interfaccia grafica in tutte le direzioni, ritornando ai moduli precedenti se occorre modificare o aggiornare i dati e le relazioni su di essi, recuperando i risultati di un'integrazione precedente con l'apertura di MOMIS in uno stato salvato per variare alcuni parametri di integrazione e confrontare gli Schemi Globali ottenuti in momenti diversi.

La struttura flessibile di MOMIS permette di individuare qualche soluzione migliorativa del metodo proposto per l'integrazione, senza ricorrere alla ricerca di nuovi elementi di analisi in letteratura. Questo consente di gestire in modo adeguato la variabile tempo, visto che si possono individuare velocemente soluzioni alternative.

Tra i fattori critici presentati nella prima parte del capitolo, due in particolare possono essere ricondotti a MOMIS per trasformarli in opportunità di successo dell'integrazione:

- 1 Il contributo del progettista in grado di supportare il sistema nell'affrontare le variabili del processo e formalizzato nel prossimo paragrafo all'interno di una nuova metodologia di mapping;
- 2 La presenza di frammenti di codice esclusi dal mapping, perché non trovano

corrispondenza in entrambe le codifiche nei segmenti considerati, ma si potrebbero integrare aggiungendo sorgenti allo stato o addirittura cercando corrispondenza tra diversi stati salvati.

7.5 Metodologia di integrazione per l'interazione del progettista con MOMIS

E' utile definire una nuova metodologia generale nell'integrazione che comprenda i casi particolari. Di seguito si propone una serie di step da seguire in fase di creazione delle relazioni inter-schema, quale possibile procedura ordinata del mapping:

1. **Mapping tra commodity "specializzate"**: inizialmente è preferibile, in fase di assegnamento delle Word Form, cercare con WordNet il significato proprio del codice e non una generalizzazione (ad esempio per il codice "Frying Pan", WordNet riconosce la Word Form "Frying_Pan" e propone un significato);
2. **Mapping tra commodity "generalizzate" o tra classi di prodotto**: se il database lessicale non possiede un significato troppo specifico, occorre generalizzare (ad esempio non è riconosciuto il codice "Cake Pan" e il progettista può scegliere la generalizzazione "Pan" o "Cake" a seconda dell'obiettivo dell'integrazione);
3. **Relazioni manuali introdotte dal progettista**: quando il dettaglio dei codici è troppo elevato per WordNet, ma entrambi gli standard integrati presentano gli stessi codici non rilevabili in SLIM, è possibile ricorrere alle relazioni manuali; in questo modo lo schema globale comprenderà da un lato le relazioni create in modo automatico, dall'altro quelle ottenute con il supporto dell'utente; (ad esempio il già citato "Cake Pan" è codificato nell'ultimo livello della gerarchia sia in UNSPSC che in ecl@ss e si deve mettere il sistema MOMIS in condizione di realizzare il mapping su questo codice prodotto);
4. **Verifica dello schema globale ottenuto**: la visualizzazione dello schema globale risultato permette di valutare le differenze tra i codici e ipotizzare delle

soluzioni per creare classi globali più ampie. Per un lavoro supplementare di questa natura bisognerebbe conoscere il marketplace di destinazione, ma si possono comunque individuare alcune tipologie di semplificazione:

- Raggruppare per materie prime costituenti il prodotto (ad esempio per gli accessori da cucina, raggruppare per accessori di plastica, di legno, ...);
- Raggruppare per uso del prodotto (ad esempio, rimanendo nell'ambito degli accessori da cucina, creare una classe contenente l'uso cutting senza distinguere tra "electric household knife", "household scissors", ...);
- Raggruppare per funzione del prodotto (ad esempio tra prodotti con funzione packaging del tipo "bag" si possono raggruppare "plastic bag", "paper bag");
- Raggruppare per tipologia di vendita (ad esempio si possono raggruppare gli utensili da cucina venduti come elettrodomestici, insieme ai prodotti della categoria "RadioTV").

Si è applicata sperimentalmente questa procedura al caso RadioTv.

La prova condotta in origine, seguendo, senza apporre miglioramenti, il lavoro di MOMIS, dimostrava che dettagliando troppo le Word Form, come suggerito da WordNet, o ignorando quelle voci che non trovavano riscontro nel database lessicale, si perdeva di vista l'obiettivo della prova: l'integrazione fra standard.

Pertanto in un secondo mapping si è deciso di generalizzare alcuni significati al livello di classe di prodotto, individuando alcuni raggruppamenti significativi: "receiver", "recorder", "player", "clock". In Figura 65 si visualizza il risultato del mapping tra classi di prodotto.

A questo punto si verifica lo schema globale ottenuto (già presentato nella sezione [6.3.5.6]); osservando le voci di codice non corrisposte, si può notare che alcune potrebbero essere raggruppate per tipologia di vendita "Audio and Electric Accessories", tra queste consideriamo "Speakers", "Subwoofers", "Mobile phone", "Headphone", "Microphone". Nel realizzare tale accorpamento si ritorna al modulo *Thes.Rel* di MOMIS e l'utente introduce manualmente le relazioni intra e inter-schema, che il sistema non potrebbe essere in grado di riconoscere automaticamente (Figura 66).

UNSPSC

- 52 Domestic Appliances and Supplies and Consumer Electronic Products
 - 16 Consumer electronics
 - 15 Audio and visual equipment
 - 01 Compact disc players
 - 02 Cassette players or recorders
 - 03 Digital video disc players
 - 04 Video cassette recorders
 - 05 Televisions
 - 07 Clock radios
 - 08 Laser disc players
 - 09 Portable stereo systems
 - 10 Home stereo systems
 - 11 Radios
 - 12 Speakers
 - 13 Combination television video cassette recorders

ECLASS

- 41 Advertising
 - 01 Promotional article
 - 06 Promot. gifts electronics, data banks
 - 01 Alarm clock
 - 02 Radio
 - 03 Translator
 - 04 World clock
 - 05 Calculator
 - 06 Address data bank
 - 07 Stereo units
 - 08 CD
 - 09 TV and video equipment
 - 10 Mobile phone, attachments
 - 11 Computer, attachment

Figura 65 Mapping tra classi di prodotto: esempio RadioTV

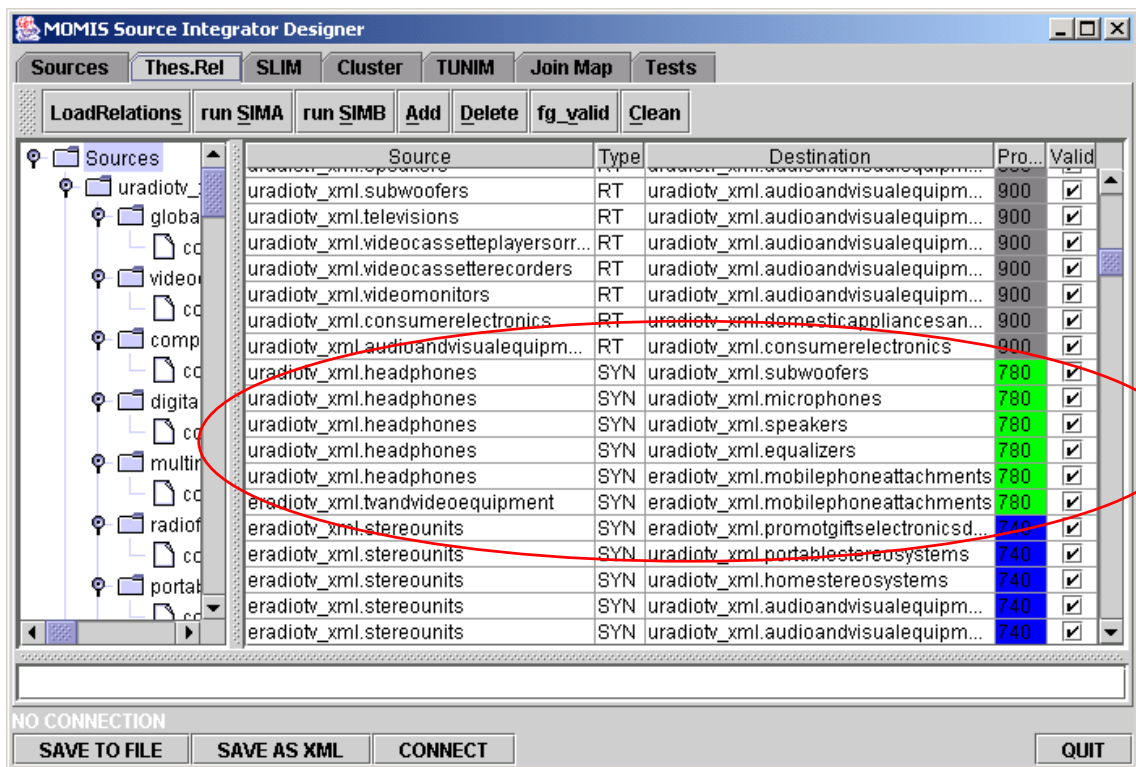


Figura 66 Introduzione di relazioni di sinonimia nel pannello *Thes.Rel*.

Il risultato finale porta a uno schema globale con 21 classi e migliora nettamente il mapping iniziale che determinava, per la stragrande maggioranza, classi monocodice. L'esempio dimostra che, se il progettista conosce le porzioni di codice da integrare e si

costruisce, in base all'esperienza, una metodologia di integrazione completa e modulare, è possibile ottenere un mapping tra standard di classificazione buono, a discapito di importanti diversità iniziali tra i codici.

Nello schema del mapping finale visualizzato in Figura 67 si può notare la ricchezza dell'integrazione ottenuta; numerosi codici di UNSPSC fra loro associati, corrispondono a prodotti codificati in ecl@ss con una sola categoria, ma nel complesso sono prese in considerazione e corrisposte tutte le categorie principali descritte dai due frammenti.

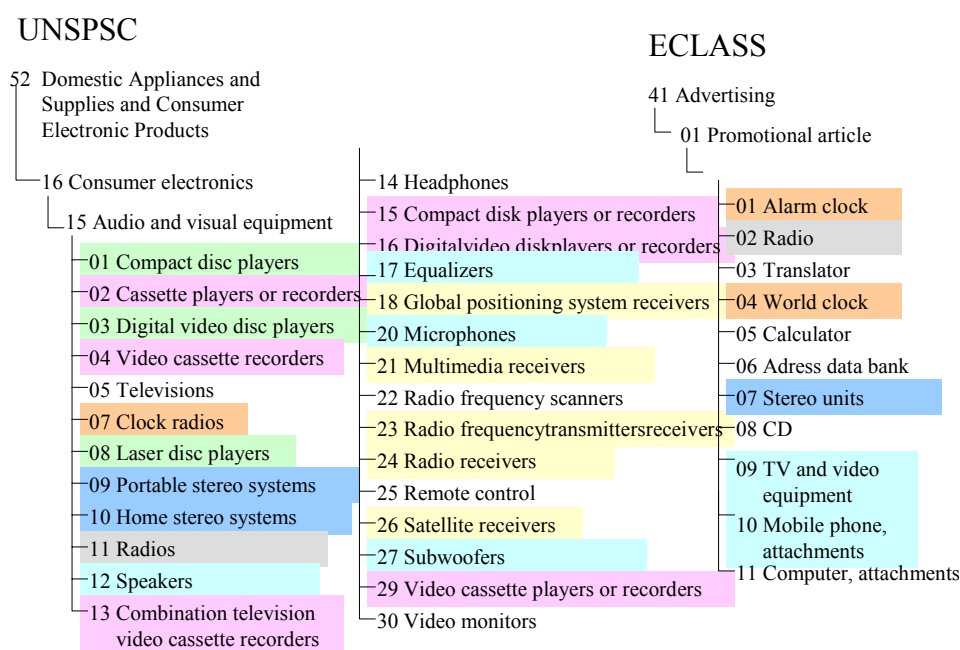


Figura 67 Risultato della nuova metodologia di mapping sul dominio RadioTV

Un esempio di questo tipo testa la validità della metodologia di integrazione individuata. E' chiaro che, se isolato, il contributo del progettista non può bastare all'ottimizzazione del processo su MOMIS, ma migliora notevolmente il punto di partenza da cui riprendere per apportare ulteriori modifiche.

7.6 Aggiornamento di sorgenti

Nella sezione [7.3.1] si è visto che non sempre, trattando esempi di grandi dimensioni, è possibile integrare in modo completo due parti di codifica, ma a volte rimangono

frammenti di codice esclusi dal mapping, perché non esplicitati dallo standard in quella parte; le differenze di base tra i content standard, infatti, sottintendono una diversa gestione delle categorie di prodotto all'interno della codifica.

Nel caso in cui il frammento escluso dal mapping sia codificato anche dall'altro standard in un segmento non considerato, MOMIS consente di riprendere la prova effettuata, aggiornando le sorgenti di dati attraverso alcune operazioni:

1. Creazione dei wrapper su nuovi file XML;
2. Caricamento in MOMIS di altre sorgenti, aggiuntive rispetto alle sorgenti integrate nella prova;
3. Salvataggio di un nuovo stato salvato.

In questo modo MOMIS è in grado di generare un Global Schema con ampie classi globali che raggruppano fra loro i prodotti di tutte le categorie più significative.

Per chiarire il procedimento riprendiamo l'esempio Household.

Dalla lettura dello Schema Globale originale si vedeva che due frammenti di codice, uno per UNSPSC e l'altro per ecl@ss non trovavano corrispondenza in entrambi gli standard. UNSPSC comprende nel segmento relativo a Household il dominio RadioTv, chiamato "Audio and Visual Equipment", ma non è descritto in ecl@ss, ecl@ss, viceversa, inserisce il dominio Food, chiamato "Food Luxury Foods" nel segmento HomeTechnology, mentre UNSPSC non lo comprende.

A fronte di 445 voci di codice si originano 260 classi globali; se si considera che tutti i prodotti contenuti nei domini Food e RadioTV sono privi di match, allora, escludendo questi due frammenti non corrisposti, è possibile dire che la prova di integrazione su Household, tenuto conto delle osservazioni fatte nel paragrafo [7.3.1], ha dato dei buoni risultati, individuando numerosi mapping.

L'esempio si inserisce tra le opportunità di sviluppo dell'integrazione tra content standard con MOMIS, nel caso di aggiornamento delle sorgenti.

Il test è stato effettuato in coda alle prove di tipo bottom-up e gli standard erano già stati ampiamente studiati; pertanto mi è stato facile individuare i segmenti che classificavano i domini Food e RadioTV rispettivamente in UNSPSC e ecl@ss.

I due domini appartenevano a frammenti già considerati nelle prove di tipo Bottom-Up ed è stato possibile recuperare i wrapper creati in precedenza e aggiungere allo stato Household.xml le sorgenti semistrutturate eradiotv_xml e ufood_xml (Figura 68).

Il nuovo schema conterrà un numero di classi globali paragonabili per molteplicità a quelle della prova iniziale, ma molto più ampie per dimensione, perché relative ad un numero di voci di codice più elevato (in effetti si sono aggiunte due sorgenti di dati).

Il risultato della prova, ottenuto utilizzando la flessibilità di MOMIS nel gestire le fasi dell'integrazione, fornisce un buono schema di mapping.

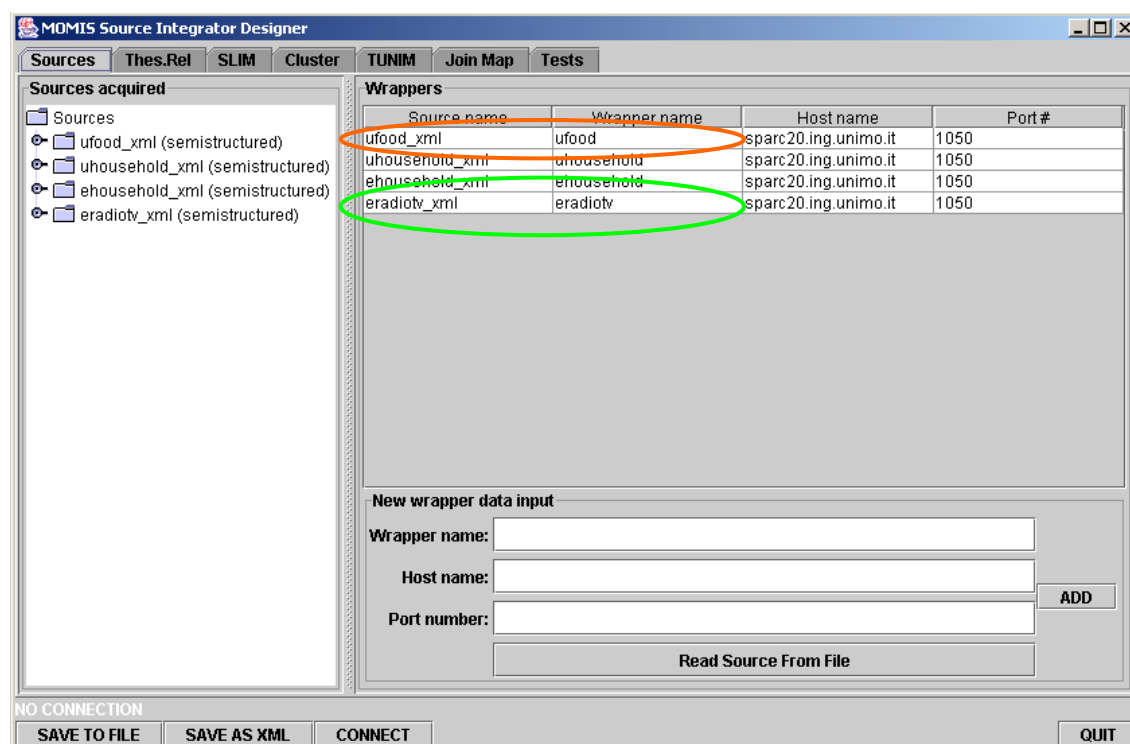


Figura 68 Aggiornamento delle sorgenti nell'esempio Household

La soluzione trovata ha acquisito due valenze: si è realizzata una prova per la quale sono stati selezionati i codici di prodotto, seguendo contemporaneamente i due approcci, bottom-up e top-down, quando in precedenza si erano tenuti distinti i due metodi; si è mostrata una possibile risposta a un problema di mismatch.

In sintesi, in presenza di una prova di tipo top-down che produce uno Schema Globale con frammenti di codice non corrisposti, si amplia il numero di sorgenti, ricercando le porzioni di codice relative alle aree di mismatch nello schema di mapping finale e si aggiornano le sorgenti con dati individuati da una ricerca di tipo bottom-up.

Questo esempio può ritenersi un riferimento sperimentale significativo, perché raccoglie tutti i principali fattori critici nell'integrazione di content standard con MOMIS e mostra l'adeguatezza del mediatore nel favorire soluzioni alternative.

Come ultima considerazione vorrei accennare a una soluzione alternativa che potrebbe

essere approfondita in lavori futuri. Per trarre un'opportunità di sviluppo dal fattore critico relativo a frammenti di codice non corrisposti, si suggerisce di prendere in considerazione le Classi Globali generate da MOMIS con una prova di tipo top-down e integrarle, in una nuova sessione di MOMIS, con le Classi Globali prodotte dalle prove di tipo bottom-up sui frammenti non mappati nella prova top-down.

7.7 WordNet: orizzonti di miglioramento

La disponibilità di database lessicali e ontologie ha reso possibile la formalizzazione degli aspetti semantici dei sistemi informativi (e di fatto dei sistemi software) in nuovi e interessanti modi.

Tra le ontologie lessicali, WordNet è considerata la più importante risorsa a disposizione per i ricercatori [Felbaum, 98]. Tuttavia il progetto sviluppato da questa tesi conferma una problematica presentata diffusamente in letteratura: la necessità di un'ontologia (lessicale e di thesaurus) globale e comune che diventi dominio semantico per tutti i tipi di applicazioni [Meersman, 99].

Partendo dal presupposto che vi sono svariate ontologie ben progettate, la costruzione di una nuova ontologia equivale al problema di assemblare quelle esistenti. La possibile inclusione di un'ontologia in un'altra ha come effetto la nascita di una nuova ontologia che consiste nell'unione delle due precedenti (unione di classi, relazioni, assiomi).

Il progetto di WordNet è stato ispirato dalle attuali teorie psicolinguistiche e computazionali sulla memoria lessicale umana; il suo successo si basa sul fatto che è disponibile on-line in versione freeware e il suo dizionario è basato sui concetti, quindi fornisce molto di più che un semplice elenco di parole in ordine alfabetico. Le sue caratteristiche specifiche sono la grande dimensione (intesa come numero di concetti), l'indipendenza di dominio, il basso livello di formalizzazione; in effetti WordNet non fornisce alcuna definizione di semantica nel linguaggio formale, ma i concetti di semantica sono definiti con termini propri del linguaggio naturale. Questo lascia definizioni vaghe e limita le possibilità di avere un supporto automatico alle elaborazioni; WordNet fonda cioè le sue basi principalmente su motivazioni linguistiche [Fensel, 01a].

Le prove di integrazione effettuate su MOMIS hanno evidenziato, per l'appunto, questo aspetto critico: WordNet, pur avendo un database lessicale ampio, si dimostra un'ontologia inadeguata quando occorrono nuovi concetti semantici e la mancanza di definizioni formali impedisce la loro introduzione. In concreto, nel lavoro su MOMIS, questo ha significato immettere numerose relazioni manuali di sinonimia, quando il modulo lessicale SLIM avrebbe consentito di generarle in automatico, se solo WordNet avesse proposto significati adeguati alle Forme Base.

Per superare queste limitazioni si potrebbe estendere WordNet, aggiungendo al database esistente nuovi synset. Per chiarire la proposta riprendiamo il concetto di matrice lessicale e applichiamo a un esempio tratto dalle prove di integrazione.

Nel test sul dominio "Bottle" si è evidenziato il problema di dovere annotare i prodotti, conservando la distinzione tra i materiali che li costituiscono: è evidente che per un potenziale acquirente che partecipa a un eMarketplace è diverso comprare "glass bottle" da "plastic bottle".

Un suggerimento utile per l'estensione di WordNet potrebbe riguardare l'introduzione di synset di dominio che comprendono oltre al concetto del prodotto quello del materiale di fabbricazione.

Word Form

$F_i = \text{PlasticBottle}$

Word Meaning

$M_j = \text{name: (Plastic Vessel; cylindrical with a narrow neck; no handle)}$

Definizione

$E_{i,j} = (F_i, M_j) \text{ PlasticBottle}$

Word Meaning	Word Form					
	F_1	F_2	F_i	F_n
M_1	$E_{1,1}$	$E_{1,2}$				
M_2						
.....						
M_j				$E_{i,j}$		
.....						
M_n						

Ad esempio la parola "PlasticBottle", dal punto di vista di WordNet, potrebbe essere individuata come associazione tra il significato M_j e la Forma Base F_i . Nella matrice lessicale la cella $E_{i,j}$ rappresenta quindi la definizione di "PlasticBottle" e, se progettata come unico lemma del synset "PlasticBottle", deve essere un'associazione 1 a 1.

I nuovi synset proposti andranno messi in relazione con quelli esistenti, nell'esempio il nome "Plastic Bottle" potrebbe essere posto in relazione di meronimia con il nome "Bottle"

Un'altra categoria di prodotti interessanti per l'estensione di WordNet emersa dalle prove, dalle quali derivano i suggerimenti, è quella degli elettrodomestici per la preparazione di alimenti descritta nella sezione [7.2.2]. La parola "Ice-cream maker" potrebbe essere un nuovo synset di WordNet messo in relazione di meronimia con il Synset "Ice-cream".

In conclusione si può affermare che i test di integrazione effettuati su MOMIS hanno messo in luce vari tipologie di parole da considerare per una proposta di estensione di WordNet; rimane il fatto che un lavoro di ampliamento di questa ontologia deve avere come obiettivo la costruzione di un'ontologia globale e comune che non può prescindere dallo studio delle altre ontologie lessicali esistenti; a tale proposito si descrivono di seguito le ontologie lessicali più importanti.

Se si pensa allo spettro delle ontologie lessicali, in posizione speculare all'estremo opposto di WordNet si colloca CYC³⁴ [Lenat&Guha, 90].

CYC, nell'ambito dell'Intelligenza Artificiale, è stato progettato per rendere accessibile e processabile ai computer la conoscenza di senso comune. In effetti la mancanza di conoscenze di senso comune ha rappresentato un'importante barriera di ingresso alle aree di applicazione dell'Intelligenza Artificiale. Si consideri come esempio la tecnica di machine learning: da un lato l'insegnamento è un prerequisito dell'intelligenza, dall'altro l'intelligenza è un prerequisito per avere un insegnamento di successo. Le persone decidono cosa sia opportuno insegnare in base alle conoscenze di senso comune. CYC ha voluto formalizzare questa conoscenza umana e tradurla in semantica processabile dalle macchine. Centinaia di migliaia di concetti sono stati formalizzati in milioni di regole, assiomi e altre asserzioni che specificano vincoli per la costruzione di significati. L'ontologia CYC è scaricabile dal sito ufficiale in una versione che contiene 3000 concetti e rappresentano solo i concetti generici, situati al livello più alto della tassonomia. I concetti specifici sono tenuti segreti e gestiti da Cycorp, la società che commercializza CYC.

CYC, come WordNet, è un'ontologia piuttosto ampia e a dominio indipendente; al contrario di WordNet fornisce una definizione formale e operativa dei suoi concetti.

Un altro esempio di ontologia è TOVE³⁵ [Fox, 93], [Fox&Grüniger, 97]. TOVE (Toronto Virtual Enterprise) è un'ontologia dedicata a un dominio specifico; nata per supportare l'integrazione tra imprese e fornire una rappresentazione della conoscenza,

³⁴ www.cyc.com

³⁵ <http://www.eil.utoronto.ca/tove/toveont.html>

TOVE si pone l'obiettivo di creare un modello di dati generico e flessibile per offrire una terminologia comune alle imprese che sia comprensibile a tutti, per definire il significato di ogni termine in un modo preciso e non ambiguo, per implementare la semantica in un set di assiomi, affinché si possa rispondere alle domande di senso comune delle imprese. In sintesi TOVE fornisce un'ontologia di concetti industriali.

Un altro esempio interessante di ontologia lessicale è KA³⁶ (Knowledge Annotation Initiative of Knowledge Acquisition Community), nata come case study [Benjamins, 99] per approfondire il processo di sviluppo di un'ontologia adeguata ad una comunità di ricerca eterogenea e internazionale e per valutare l'uso dell'ontologia come accesso semantico alle sorgenti di informazione on-line. L'idea che ha mosso il progetto prevede che, se questa iniziativa fosse in grado di rappresentare una conoscenza condivisa, molti ricercatori collaborerebbero insieme – da luoghi diversi – per costruire in KA un'ontologia; vi sarebbe così la certezza di dare luogo a un'ontologia accettata dalla maggioranza dei ricercatori. I criteri fondamentali individuati per il progetto sono stati la specializzazione dei concetti generali in più concetti specifici e la modularità in modo da avere flessibilità d'uso.

Ciascun modello di ontologia lessicale proposto contiene alcuni elementi potenzialmente importanti per rispondere ai problemi emersi con l'uso di WordNet su MOMIS. Tuttavia a vantaggio di WordNet rimangono l'ampiezza di termini, non paragonabile con le altre, e il grado di diffusione e notorietà raggiunto negli ambienti di ricerca. Di conseguenza, per riprendere l'idea originale, che sottolineava la necessità di dotarsi di un'ontologia lessicale globale e comune, si può pensare che questa ontologia non solo sia un'integrazione di iniziative già esistenti, ma abbia come punto di partenza imprescindibile WordNet.

I test effettuati su MOMIS confermano queste considerazioni e possono rappresentare per gli sviluppatori del sistema un riferimento interessante per valutare un ampliamento dell'ontologia lessicale utilizzata.

³⁶ <http://www.aifb.uni-karlsruhe.de/WBS/broker/KA2.html>

7.8 Raccomandazioni per la costruzione di un nuovo standard globale

Uno dei temi di ricerca lanciati dal Challenge di OntoWeb riguardava la proposta di una nuova Lingua Franca che classifichi i prodotti per il commercio elettronico B2B.

Le Classi Globali di MOMIS, per come sono state progettate e ottenute nelle prove di integrazione, rappresentano una possibile Lingua Franca. Tuttavia vi sono almeno due vincoli che attualmente ne limitano uno studio più approfondito.

In primo luogo i codici prodotto commercializzabili e richiesti dal mercato sono in continua evoluzione, quindi i sistemi di classificazione dovrebbero essere aggiornati e sviluppati in continua. In secondo luogo le imprese, per essere competitive e innovative, devono essere capaci di dare risposte veloci ai clienti e non possono aspettare decenni per vedere crescere uno standard globale. Quindi molte imprese hanno popolato il Web in varie comunità virtuali, sviluppando e implementando numerosi sistemi di classificazione dei prodotti.

Gli attuali content standard si dividono in orizzontali e verticali, alcuni seguono il punto di vista del prodotto, mentre altri sono orientati alla produzione; queste caratteristiche divergenti rispecchiano le esigenze dei loro gestori e allo stesso tempo permettono di soddisfare le molteplici richieste dei marketplace elettronici, che a seconda della loro tipologia (presentati in sezione [1.3]) gestiscono lo scambio di informazioni sui prodotti con caratteristiche diverse.

Si riconferma quindi più interessante l'approccio della Stele di Rosetta, che vuole fornire una codifica generale a partire dalle ontologie esistenti.

Lo studio effettuato in questa tesi sugli standard ecl@ss, UNSPSC e in un primo momento NAICS, suggerisce alcune raccomandazioni da indicare per futuri standard di classificazione.

UNSPSC e ecl@ss sono stati confrontabili e integrabili perché orientati alla classificazione dei prodotti; da entrambi si possono ricavare spunti di miglioramento:

- Una prima raccomandazione può riguardare la gestione della gerarchia; per sfruttarne le potenzialità lo standard deve distribuire i codici prodotto sui livelli gerarchici, dal significato più ampio a quello più specifico. Nella realtà questo non è

stato sempre riscontrabile, in particolare in ecl@ss;

- Nomi di prodotto uguali non devono essere ripetuti all'interno dello stesso segmento, ma al limite si possono ripresentare in un'altra parte della codifica, se descrivono una funzione diversa;
- Il dominio relativo a una determinata categoria di prodotto va dettagliato con cura all'ultimo livello gerarchico, se si vuole orientare la classificazione di prodotto a codici scambiati su eMarketplace verticali;
- I codici dei servizi associati alle famiglie di prodotto possono essere consultati con maggiore facilità e rapidità se vengono declinati in corrispondenza della famiglia di prodotto interessata dal servizio, piuttosto che in segmenti esclusivamente dedicati ai servizi. UNSPSC dedica ben 18 segmenti ai servizi e l'utente, in fase di consultazione delle codifiche alla ricerca di un prodotto, può dimenticare di prendere in considerazione i servizi relativi, perché non codificati nello stesso ramo in cui si presenta il prodotto;
- I codici di prodotto standardizzati devono favorire l'incontro in tempi brevi fra domanda o offerta di numerosi compratori e venditori che partecipano a un eMarketplace B2B; di conseguenza, in ambito e-commerce, risultano più importanti le ontologie orientate ai prodotti (UNSPSC, ecl@ss), piuttosto che quelle orientate al processo produttivo (NAICS).

Questi elementi, riscontrati sul campo, soddisfano i criteri di generalità e condivisione suggeriti nella definizione di ontologia in [Gruber, 93] e possono contribuire a migliorare gli standard esistenti; in alternativa si potrebbero considerare come consigli e raccomandazioni per un nuovo standard globale.

Le ontologie sono state sviluppate per fornire informazioni semantiche processabili dai computer a partire dal linguaggio umano; devono rappresentare la conoscenza condivisa di un dominio e quindi il loro sviluppo è spesso un processo di cooperazione che coinvolge molti ricercatori, da luoghi diversi. La comunanza della semantica diventa fondamentale nello scambio di informazioni, in particolare se si ragiona di content standard nell'area B2B. Nel settore dell'e-commerce le ontologie corrispondono agli standard di classificazione prodotto.

La nuova Stele di Rosetta dell'era digitale deve essere in grado di mediare tra ampi e condivisi standard, al fine di rispondere in modo veloce e concreto alla domanda di

generalità e universalità semantica richiesta dagli attori del commercio elettronico.
Un'ontologia globale metterebbe a disposizione informazioni sempre più accessibili agli utenti e leggibili dalle macchine - informazioni *human&machine understandable* – e garantirebbe la condizione di imparzialità necessaria in un mercato elettronico.

Conclusioni

Questa tesi ha affrontato lo studio del problema di integrazione di alcuni standard per la classificazione dei prodotti nell'e-commerce, sulla base di una proposta di ricerca lanciata da OntoWeb, un network dell'Unione Europea dedicato agli studi ontologici.

In particolare l'obiettivo è stato quello di studiare e proporre una nuova metodologia di integrazione dei principali standard di classificazione esistenti (content standard), attraverso il sistema a mediatore MOMIS, progettato per integrare sorgenti di dati eterogenee e distribuite su Web.

In base alle analisi effettuate, l'importanza di una metodologia di integrazione è risultata evidente soprattutto considerando che il sistema MOMIS è un sistema di integrazione semiautomatico, ovvero effettua parte delle integrazioni in modo automatico (in base agli schemi delle sorgenti) e in parte richiede l'interazione con il progettista per effettuare delle inevitabili scelte durante l'integrazione. Inoltre la proposta metodologica è diventata necessaria per considerare il contesto specifico dell'integrazione di standard di classificazione.

Durante le analisi e per la proposta si è fatto riferimento a parametri e temi di progetto descritti da OntoWeb che indica l'integrazione tra gli standard esistenti come la risposta migliore all'urgenza di avere uno standard globale e universale per la classificazione dei prodotti, da preferire alla creazione di una nuova codifica, che richiederebbe elevati investimenti in termini finanziari e temporali.

In sostanza si è prospettato l'obiettivo di realizzare dei mapping tra i codici prodotto degli standard di classificazione già esistenti, piuttosto che proporre un nuovo standard.

L'obiettivo è stato illustrato con un esempio su un piccolo frammento delle due codifiche, quindi si è mostrata la possibilità di ampliare lo studio di integrazione delle classificazioni a tutte le voci di codice, utilizzando MOMIS.

La fase sperimentale della tesi si è incentrata su numerose prove di integrazione effettuate con MOMIS per testare l'adeguatezza del sistema a generare i mapping tra gli standard.

Un'analisi finale dei risultati ottenuti ha permesso di individuare dei fattori critici nel processo di integrazione, alcuni imputabili al sistema MOMIS:

- Contributo del progettista;
- Adeguatezza dell'ontologia lessicale WordNet.

Altri riconducibili agli standard:

- Frammenti di codice non corrisposti;
- Gestione della gerarchia nei codici.

Da questi elementi di debolezza sono state tratte delle opportunità di sviluppo: si è formulata una nuova metodologia di integrazione che il progettista può seguire nell'utilizzare MOMIS con sorgenti di dati del tipo dei content standard proposti; sono stati raccolti esempi sufficienti a sostegno di un'estensione di WordNet con nuovi synset, per costruire un'ontologia di dominio.

Infine lo studio effettuato nella tesi sugli standard ecl@ss e UNSPSC e in particolare le difficoltà riscontrate per le diverse impostazioni strutturali delle codifiche hanno suggerito alcune raccomandazioni da indicare per futuri standard di classificazione, riferite soprattutto al buon impiego commerciale dei codici da parte delle imprese.

Appendice A

Software di supporto alla fase di testing

XML Spy, Versione 4.4, 2002

XML Spy Suite è un prodotto completo che permette di utilizzare tutti gli strumenti necessari allo sviluppo e applicazione di file XML; in particolare XML Spy IDE è un software leader nelle soluzioni industriali per lo sviluppo di applicazioni basate su XML, permettendo una semplice creazione e gestione di documenti XML, schemi XML e fogli XSLT.

Per questa tesi si è utilizzato l'editor di file DTD che permette di verificare i vincoli di file "valido" e file "ben-formato" e, a partire da DTD validi, l'editor di file XML.

Tera Term Pro, Versione 2.3, 1998

È un software freeware per emulazione di terminale; si è utilizzato in particolare il collegamento TTSSH (TeraTerm Secure Shell extension) versione 1.5.1, compatibile con il protocollo SSH, versione 1.5, per il collegamento al server *sparc20* e l'esecuzione di SI-Designer on-line.

Sec Ex, Versione 1.4, 2001

Security Express for Windows è un software per la gestione di file in rete e offre un emulatore di terminale che permetterà di trasferire dati tra un PC e un server con

modalità Secure Shell, versione 2, ma che recupera la versione 1 quando necessario.

Il trasferimento dei file da e per i server SSH avviene con protocollo SFTP (Secure FTP) quando disponibile, oppure con il più vecchio SCP (Secure Copy Program).

Nella tesi si è impiegato questo software, in versione *evaluation*, per spostare e recuperare file dalla cartella di account personale bonini@sparc20.ing.unimo.it su *sparc20*, trasferimento dati piuttosto utile durante il caricamento dei file XML per la creazione dei wrapper.

Micro X-Win 32, versione 1.30.166.0, 1992

E' un Xserver per la connessione a un Terminale X, una rete distribuita che usa X Windows System, necessario per ridirigere l'output grafico sulla macchina corrente, quando si eseguono operazioni in connessione a un server. Veniva attivato quando si utilizzava un'emulazione da terminale per interagire in remoto con MOMIS.

Appendice B

Test di riferimento in XML

Di seguito è riportata la descrizione in XML dello schema finale di integrazione, ottenuto con MOMIS, relativo al test di riferimento impiegato nella trattazione della tesi (effettuato sul dominio RadioTV).

schemaXML Description:

```
<?xml version="1.0"?>
<!-- Build file for application -->
<Schema name="" >
  <Source
    name="uradiotv_xml"
    type="semistructured"
    wrapperName="uradiotv"
    hostName="sparc20.ing.unimo.it"
    portNumber="1050" >
    <Interface
      name="globalpositioningsystemreceivers"
      persistent="false" >
      <extent name="globalpositioningsystemreceivers" />
      <Attribute
        name="cod"
        type="string">
      <AttributeAdditionalInfo>
        <SlimNode formaBase="cod" >
          <SlimNodeSenso numero="-2" />
        </SlimNode>
      </AttributeAdditionalInfo>
      </Attribute>
      <InterfaceAdditionalInfo>
      <SlimNode
        formaBase="globalpositioningsystemreceivers" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
      </InterfaceAdditionalInfo>
      </Interface>
      <Interface
        name="videocassetterecorders"
        persistent="false">
      <extent name="videocassetterecorders" />
      <Attribute
        name="cod"
        type="string">
```

```
      <AttributeAdditionalInfo>
        <SlimNode formaBase="cod" >
          <SlimNodeSenso numero="-2" />
        </SlimNode>
      </AttributeAdditionalInfo>
    </Attribute>
    <InterfaceAdditionalInfo>
    <SlimNode formaBase="videocassetterecorders" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
    </InterfaceAdditionalInfo>
  </Interface>
  <Interface
    name="compactdiscplayers"
    persistent="false" >
    <extent name="compactdiscplayers" />
    <Attribute
      name="cod"
      type="string">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="cod" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
    </Attribute>
    <InterfaceAdditionalInfo>
    <SlimNode formaBase="compactdiscplayers" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
    </InterfaceAdditionalInfo>
  </Interface>
  <Interface
    name="digitalvideodiscplayers"
    persistent="false"
    >
    <extent name="digitalvideodiscplayers" />
    <Attribute
      name="cod"
      type="string">
```

```

<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="digitalvideodiscplayers" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="multimediareceivers"
  persistent="false">
  <extent name="multimediareceivers" />
  <Attribute
    name="cod"
    type="string">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="cod" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <InterfaceAdditionalInfo>
    <SlimNode formaBase="multimediareceivers" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
<Interface
  name="radiofrequencyscanners"
  persistent="false">
  <extent name="radiofrequencyscanners" />
  <Attribute
    name="cod"
    type="string">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="cod" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <InterfaceAdditionalInfo>
    <SlimNode formaBase="radiofrequencyscanners" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
<Interface
  name="portablestereosystems"
  persistent="false">
  <extent name="portablestereosystems" />
  <Attribute
    name="cod"
    type="string">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="cod" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <InterfaceAdditionalInfo>
    <SlimNode formaBase="portablestereosystems" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
<Interface
  name="audioandvisualequipment"
  persistent="false">
  <extent name="audioandvisualequipment" />
  <Attribute
    name="globalpositioningsystemreceivers"
    type="globalpositioningsystemreceivers">
    <AttributeAdditionalInfo>
      <SlimNode
        formaBase="globalpositioningsystemreceivers" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="videocassetterecorders"
    type="videocassetterecorders">
  <AttributeAdditionalInfo>
    <SlimNode
      formaBase="videocassetterecorders" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="digitalvideodiscplayers"
    type="digitalvideodiscplayers">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="digitalvideodiscplayers">
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="compactdiscplayers"
    type="compactdiscplayers">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="compactdiscplayers" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="multimediareceivers"
    type="multimediareceivers">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="multimediareceivers" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="radiofrequencyscanners"
    type="radiofrequencyscanners">
  <AttributeAdditionalInfo>
    <SlimNode
      formaBase="radiofrequencyscanners" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="portablestereosystems"
    type="portablestereosystems">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="portablestereosystems" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="homestereosystems"
    type="homestereosystems">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="homestereosystems" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="clockradios"
    type="clockradios">
  <AttributeAdditionalInfo>

```



```

        <SlimNode formaBase="clockradios" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="headphones"
    type="headphones">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="headphones" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="televisions"
    type="televisions">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="televisions" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="radiofrequencytransmittersreceivers"
    type="radiofrequencytransmittersreceivers">
    <AttributeAdditionalInfo>
        <SlimNode
formaBase="radiofrequencytransmittersreceivers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="equalizers"
    type="equalizers">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="equalizers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="cassetteplayersorrecorders"
    type="cassetteplayersorrecorders">
    <AttributeAdditionalInfo>
        <SlimNode
formaBase="cassetteplayersorrecorders" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="combinationtelevisionvideocassetterecorders"
    type="combinationtelevisionvideocassetterecorders">
    <AttributeAdditionalInfo>
        <SlimNode
formaBase="combinationtelevisionvideocassetterecorders"
">
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="videocassetteplayersorrecorders"
    type="videocassetteplayersorrecorders">
    <AttributeAdditionalInfo>
        <SlimNode
formaBase="videocassetteplayersorrecorders" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="radios"
    type="radios">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="radios" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="remotecontrol"
    type="remotecontrol">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="remotecontrol" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="subwoofers"
    type="subwoofers">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="subwoofers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="speakers"
    type="speakers">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="speakers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="videomonitors"
    type="videomonitors">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="videomonitors" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="laserdiscplayers"
    type="laserdiscplayers">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="laserdiscplayers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="radioreceivers"
    type="radioreceivers">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="radioreceivers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="satellitereceivers"
    type="satellitereceivers">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="satellitereceivers" >
        <SlimNodeSenso numero="-2" />
        </SlimNode>
    </AttributeAdditionalInfo>
</Attribute>
<Attribute
    name="microphones"
    type="microphones">
    <AttributeAdditionalInfo>
        <SlimNode formaBase="microphones" >

```



```

<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode
formaBase="cassetteplayersorrecorders" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="combinationtelevisionvideocassetterecorders"
persistent="false">
<extent
name="combinationtelevisionvideocassetterecorders" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode
formaBase="combinationtelevisionvideocassetterecorders"
">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="videocassetteplayersorrecorders"
persistent="false">
<extent name="videocassetteplayersorrecorders" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode
formaBase="videocassetteplayersorrecorders" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="radios"
persistent="false">
<extent name="radios" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="radios" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="subwoofers"
persistent="false">
<extent name="subwoofers" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="subwoofers" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="remotecontrol"
persistent="false">
<extent name="remotecontrol" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="remotecontrol" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="videomonitors"
persistent="false">
<extent name="videomonitors" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="videomonitors" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="speakers"
persistent="false">
<extent name="speakers" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="speakers" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="laserdiscplayers"

```

```

persistent="false">
<extent name="laserdiscplayers" />
<Attribute
  name="cod"
  type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="laserdiscplayers" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="domesticappliancesandsuppliesandconsumerelec
tronicproducts"
  persistent="false">
<extent
  name="domesticappliancesandsuppliesandconsumerelec
tronicproducts" />
<Attribute
  name="consumerelectronics"
  type="consumerelectronics">
<AttributeAdditionalInfo>
  <SlimNode formaBase="consumerelectronics" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode
formaBase="domesticappliancesandsuppliesandconsume
relectronicproducts" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="radioreceivers"
  persistent="false">
<extent name="radioreceivers" />
<Attribute
  name="cod"
  type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="radioreceivers" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="microphones"
  persistent="false">
<extent name="microphones" />
<Attribute
  name="cod"
  type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="microphones" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="satellitereceivers"
  persistent="false">
<extent name="satellitereceivers" />
<Attribute
  name="cod"
  type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="satellitereceivers" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="consumerelectronics"
  persistent="false">
<extent name="consumerelectronics" />
<Attribute
  name="audioandvisualequipment"
  type="audioandvisualequipment">
<AttributeAdditionalInfo>
  <SlimNode
formaBase="audioandvisualequipment" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="consumerelectronics" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="digitalvideodiskplayersorrecorders"
  persistent="false">
<extent
  name="digitalvideodiskplayersorrecorders"/>
<Attribute
  name="cod"
  type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode
formaBase="digitalvideodiskplayersorrecorders" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="compactdiskplayersorrecorders"
  persistent="false">
<extent name="compactdiskplayersorrecorders" />
<Attribute
  name="cod"
  type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="compactdiskplayersorrecorders" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>

```

```

</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode
formaBase="compactdiskplayersorrecorders" >
    <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
</Source>
<Source
name="eradiotv_xml"
type="semistructured"
wrapperName="eradiotv"
hostName="sparc20.ing.unimo.it"
portNumber="1050">
<Interface
name="alarmclock"
persistent="false">
<extent name="alarmclock" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="alarmclock" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="worldclock"
persistent="false">
<extent name="worldclock" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="worldclock" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="adressedatabank"
persistent="false">
<extent name="adressedatabank" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="adressedatabank" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="radio"
persistent="false">
<extent name="radio" />
</Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="radio" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="translator"
persistent="false">
<extent name="translator" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="translator" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="promotionalarticle"
persistent="false">
<extent name="promotionalarticle" />
<Attribute
name="promotgiftselectronicssdatabanks"
type="promotgiftselectronicssdatabanks">
<AttributeAdditionalInfo>
  <SlimNode
formaBase="promotgiftselectronicssdatabanks" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="promotionalarticle" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="tvandvideoequipment"
persistent="false">
<extent name="tvandvideoequipment" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="tvandvideoequipment" >
  <SlimNodeSensio numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="stereounits"
persistent="false">

```

```

<extent name="stereounits" />
<Attribute
  name="cod"
  type="string">
  <AttributeAdditionalInfo>
    <SlimNode formaBase="cod" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode formaBase="stereounits" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
  name="computerattachments"
  persistent="false">
  <extent name="computerattachments" />
  <Attribute
    name="cod"
    type="string">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="cod" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <InterfaceAdditionalInfo>
    <SlimNode formaBase="computerattachments" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
<Interface
  name="advertising"
  persistent="false">
  <extent name="advertising" />
  <Attribute
    name="promotionalarticle"
    type="promotionalarticle">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="promotionalarticle" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <InterfaceAdditionalInfo>
    <SlimNode formaBase="advertising" >
      <SlimNodeSenso numero="-2" />
    </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
<Interface
  name="promotgiftselectronicsdatabanks"
  persistent="false">
  <extent name="promotgiftselectronicsdatabanks" />
  <Attribute
    name="alarmclock"
    type="alarmclock">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="alarmclock" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="adressdatabank"
    type="adressdatabank">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="adressdatabank" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="worldclock"
    type="worldclock">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="worldclock" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="radio"
    type="radio">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="radio" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="translator"
    type="translator">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="translator" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="tvandvideoequipment"
    type="tvandvideoequipment">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="tvandvideoequipment" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="stereounits"
    type="stereounits">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="stereounits" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="computerattachments"
    type="computerattachments">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="computerattachments" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="mobilephoneattachments"
    type="mobilephoneattachments">
    <AttributeAdditionalInfo>
      <SlimNode
formaBase="mobilephoneattachments" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="calculator"
    type="calculator">
    <AttributeAdditionalInfo>
      <SlimNode formaBase="calculator" >
        <SlimNodeSenso numero="-2" />
      </SlimNode>
    </AttributeAdditionalInfo>
  </Attribute>
  <Attribute
    name="cd"
    type="cd">

```

```

<AttributeAdditionalInfo>
  <SlimNode formaBase="cd" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
  <SlimNode
formaBase="promotgiftselectronicstabanks" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="mobilephoneattachments"
persistent="false">
<extent name="mobilephoneattachments" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="mobilephoneattachments">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="calculator"
persistent="false">
<extent name="calculator" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="calculator" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
<Interface
name="cd"
persistent="false">
<extent name="cd" />
<Attribute
name="cod"
type="string">
<AttributeAdditionalInfo>
  <SlimNode formaBase="cod" >
    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
<InterfaceAdditionalInfo>
<SlimNode formaBase="cd" >
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
</Source>

<ThesRelations>
  <TRInterfaceRel
attribute1="eradiotv_xml.promotgiftselectronicstabanks"
relType="rt"
attribute2="eradiotv_xml.promotionalarticle"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel
attribute1="eradiotv_xml.promotionalarticle"
relType="rt"
attribute2="eradiotv_xml.advertising"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel
attribute1="eradiotv_xml.adresstabank"
relType="rt"
attribute2="eradiotv_xml.promotgiftselectronicstabanks"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel
attribute1="eradiotv_xml.alarmclock"
relType="rt"
attribute2="eradiotv_xml.promotgiftselectronicstabanks"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel attribute1="eradiotv_xml.calculator"
relType="rt"
attribute2="eradiotv_xml.promotgiftselectronicstabanks"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel attribute1="eradiotv_xml.cd"
relType="rt"
attribute2="eradiotv_xml.promotgiftselectronicstabanks"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel
attribute1="eradiotv_xml.computerattachments"
relType="rt"
attribute2="eradiotv_xml.promotgiftselectronicstabanks"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel
attribute1="eradiotv_xml.mobilephoneattachments"
relType="rt"
attribute2="eradiotv_xml.promotgiftselectronicstabanks"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
  <TRInterfaceRel attribute1="eradiotv_xml.radio"
relType="rt"

```

```

attribute2="eradiotv_xml.promotgiftselectronicsdatabanks
"
    validated="true"
    extensional="false"
    class="InterfaceRel"
    producerId="900" />
<TRInterfaceRel attribute1="eradiotv_xml.stereounits"
relType="rt"

attribute2="eradiotv_xml.promotgiftselectronicsdatabanks
"
    validated="true"
    extensional="false"
    class="InterfaceRel"
    producerId="900" />
<TRInterfaceRel attribute1="eradiotv_xml.translator"
relType="rt"

attribute2="eradiotv_xml.promotgiftselectronicsdatabanks
"
    validated="true"
    extensional="false"
    class="InterfaceRel"
    producerId="900" />
<TRInterfaceRel
attribute1="eradiotv_xml.tvandvideoequipment"
relType="rt"

attribute2="eradiotv_xml.promotgiftselectronicsdatabanks
"
    validated="true"
    extensional="false"
    class="InterfaceRel"
    producerId="900" />
<TRInterfaceRel attribute1="eradiotv_xml.worldclock"
relType="rt"

attribute2="eradiotv_xml.promotgiftselectronicsdatabanks
"
    validated="true"
    extensional="false"
    class="InterfaceRel"
    producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.cassetteplayersorreorders"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel attribute1="uradiotv_xml.clockradios"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.combinationtelevisionvideocasse
ttereorders"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.compactdiscplayers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.compactdiskplayersorreorders"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.digitalvideodiscplayers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.digitalvideodiskplayersorreord
ers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel attribute1="uradiotv_xml.equalizers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.globalpositioningsystemreceiver
s"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.headphones"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.homestereosystems"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.laserdiscplayers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"

```



```

        validated="true"
        extensional="false"
        class="InterfaceRel"
        producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.microphones"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.multimediareceivers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.portablestereosystems"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.radiofrequencytransmittersreceiv
ers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.radiofrequencyscanners"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.radioreceivers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel attribute1="uradiotv_xml.radios"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.remotecontrol"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.satellitereceivers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel attribute1="uradiotv_xml.speakers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.subwoofers"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.televisions"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.videocassetteplayersorreorder
s"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.videocassetterecorders"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.videomonitors"
relType="rt"

attribute2="uradiotv_xml.audioandvisualequipment"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.consumerelectronics"
relType="rt"

attribute2="uradiotv_xml.domesticappliancesandsupplies
andconsumerelectronicproducts"
validated="true"
extensional="false"
class="InterfaceRel"
producerId="900" />
    <TRInterfaceRel
attribute1="uradiotv_xml.audioandvisualequipment"
relType="rt"

```

```

attribute2="uradiotv_xml.consumerelectronics"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="900" />
<TRInterfaceRel
attribute1="uradiotv_xml.headphones"
  relType="syn"
  attribute2="uradiotv_xml.subwoofers"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="780" />
<TRInterfaceRel
attribute1="uradiotv_xml.headphones"
  relType="syn"
  attribute2="uradiotv_xml.microphones"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="780" />
<TRInterfaceRel
attribute1="uradiotv_xml.headphones"
  relType="syn"
  attribute2="uradiotv_xml.speakers"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="780" />
<TRInterfaceRel
attribute1="uradiotv_xml.headphones"
  relType="syn"
  attribute2="uradiotv_xml.equalizers"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="780" />
<TRInterfaceRel
attribute1="uradiotv_xml.headphones"
  relType="syn"
attribute2="eradiotv_xml.mobilephoneattachments"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="780" />
<TRInterfaceRel
attribute1="eradiotv_xml.tvandvideoequipment"
  relType="syn"
attribute2="eradiotv_xml.mobilephoneattachments"
  validated="true"
  extensional="false"
  class="InterfaceRel"
  producerId="780" />
</ThesRelations>

<GlobalClass name="Global0" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="eradiotv_xml.computerattachments" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="eradiotv_xml.mobilephoneattachments" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="eradiotv_xml.tvandvideoequipment" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  <LocalAttribute localClass="uradiotv_xml.equalizers" >
    <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  <LocalAttribute
localClass="uradiotv_xml.headphones" >
    <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  <LocalAttribute
localClass="uradiotv_xml.microphones" >
    <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  <LocalAttribute localClass="uradiotv_xml.speakers"
>
    <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  <LocalAttribute
localClass="uradiotv_xml.subwoofers" >
    <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global1" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute localClass="eradiotv_xml.cd" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global10" iffused="true">
  <GlobalAttribute name="promotionalarticle"
domain="complex" complex="false">
    <LocalAttribute
localClass="eradiotv_xml.advertising" >
      <SimpleMapping
attributeName="promotionalarticle" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global11" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="eradiotv_xml.adressdatabank" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global12" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.videomonitors" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global13" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="eradiotv_xml.stereounits" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.homestereosystems" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.portablestereosystems" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

```

```

<GlobalClass name="Global14" iffused="true">
<GlobalAttribute name="cod" domain="string"
complex="false">
<LocalAttribute
localClass="uradiotv_xml.cassetteplayersorrecorders" >
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute
localClass="uradiotv_xml.combinationtelevisionvideocass
etterecorders" >
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute
localClass="uradiotv_xml.compactdiskplayersorrecorders"
>
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute
localClass="uradiotv_xml.digitalvideodiskplayersorrecord
ers" >
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute
localClass="uradiotv_xml.videocassetteplayersorrecorder
s" >
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute
localClass="uradiotv_xml.videocassetterecorders" >
<SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalClass>

```

```

<GlobalClass name="Global15" iffused="true">
<GlobalAttribute name="cod" domain="string"
complex="false">
<LocalAttribute
localClass="eradiotv_xml.alarmclock">
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute localClass="eradiotv_xml.worldclock">
<SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute
localClass="uradiotv_xml.clockradios">
<SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>

```

```

<GlobalClass name="Global16" iffused="true">
<GlobalAttribute name="cod" domain="string"
complex="false">
<LocalAttribute localClass="uradiotv_xml.televisions">
<SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>

```

```

<GlobalClass name="Global17" iffused="true">
<GlobalAttribute name="cod" domain="string"
complex="false">
<LocalAttribute localClass="eradiotv_xml.translator" >
<SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>

```

```

<GlobalClass name="Global18" iffused="true">
<GlobalAttribute name="adressedatabank"
domain="complex" complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >

```

```

<SimpleMapping
attributeName="adressedatabank" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="alarmclock"
domain="complex" complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping attributeName="alarmclock" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="calculator" domain="complex"
complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping attributeName="calculator" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="cd" domain="complex"
complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping attributeName="cd" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="computerattachments"
domain="complex" complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping
attributeName="computerattachments" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="mobilephoneattachments"
domain="complex" complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping
attributeName="mobilephoneattachments" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="radio" domain="complex"
complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping attributeName="radio" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="stereounits"
domain="complex" complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping attributeName="stereounits" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="translator" domain="complex"
complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >
<SimpleMapping attributeName="translator" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="tvandvideoequipment"
domain="complex" complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdataban
ks" >

```

```

    <SimpleMapping
attributeName="tvandvideoequipment" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="worldclock" domain="complex"
complex="false">
<LocalAttribute
localClass="eradiotv_xml.promotgiftselectronicsdatabank
s" >
    <SimpleMapping attributeName="worldclock" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global19" iffused="true">
<GlobalAttribute name="cod" domain="string"
complex="false">
<LocalAttribute
localClass="uradiotv_xml.radiofrequencyscanners" >
    <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global2" iffused="true">
<GlobalAttribute name="cod" domain="string"
complex="false">
<LocalAttribute localClass="eradiotv_xml.radio" >
    <SimpleMapping attributeName="cod" />
</LocalAttribute>
<LocalAttribute localClass="uradiotv_xml.radios" >
    <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global20" iffused="true">
<GlobalAttribute name="cassetteplayersorrecorders"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="cassetteplayersorrecorders" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="clockradios"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="clockradios" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute
name="combinationtelevisionvideocassetterecorders"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="combinationtelevisionvideocassetterecor
ders" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="compactdiscplayers"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="compactdiscplayers" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute
name="compactdiskplayersorrecorders"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >

```

```

    <SimpleMapping
attributeName="compactdiskplayersorrecorders" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="digitalvideodiscplayers"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="digitalvideodiscplayers" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute
name="digitalvideodiskplayersorrecorders"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="digitalvideodiskplayersorrecorders" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="equalizers"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="equalizers" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute
name="globalpositioningsystemreceivers"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="globalpositioningsystemreceivers" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="headphones"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="headphones" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="homestereosystems"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="homestereosystems" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="laserdiscplayers"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="laserdiscplayers" />
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="microphones"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="microphones"
/>
</LocalAttribute>
</GlobalAttribute>
<GlobalAttribute name="multimediareceivers"
domain="complex" complex="false">
<LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="multimediareceivers" />
</LocalAttribute>
</GlobalAttribute>

```

```

    <GlobalAttribute name="portablestereosystems"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="portablestereosystems" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="radiofrequencyscanners"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="radiofrequencyscanners" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute
name="radiofrequencytransmittersreceivers"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="radiofrequencytransmittersreceivers" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="radioreceivers"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="radioreceivers" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="radios" domain="complex"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="radios" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="remotecontrol"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="remotecontrol"/>
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="satellitereceivers"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="satellitereceivers" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="speakers" domain="complex"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="speakers" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="subwoofers"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="subwoofers" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="televisions" domain="complex"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="televisions" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute
name="videocassetteplayersorrecorders"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="videocassetteplayersorrecorders" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="videocassetterecorders"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping
attributeName="videocassetterecorders" />
    </LocalAttribute>
    </GlobalAttribute>
    <GlobalAttribute name="videomonitors"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.audioandvisualequipment" >
    <SimpleMapping attributeName="videomonitors" />
    </LocalAttribute>
    </GlobalAttribute>
    </GlobalClass>
    <GlobalClass name="Global3" iffUsed="true">
    <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.globalpositioningsystemreceiv
ers" >
    <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.multimediareceivers" >
    <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.radiofrequencytransmittersrece
ivers" >
    <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.radioreceivers" >
    <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.satellitereceivers" >
    <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    </GlobalAttribute>
    </GlobalClass>
    <GlobalClass name="Global4" iffUsed="true">
    <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute localClass="eradiotv_xml.calculator"
>
    <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    </GlobalAttribute>
    </GlobalClass>
    <GlobalClass name="Global5" iffUsed="true">
    <GlobalAttribute name="consumerelectronics"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.domesticappliancesandsupplie
sandconsumerelectronicproducts" >
    <SimpleMapping
attributeName="consumerelectronics" />
    </LocalAttribute>
    </GlobalAttribute>
    </GlobalClass>

```

```

<GlobalClass name="Global6" iffused="true">
  <GlobalAttribute name="audioandvisualequipment"
domain="complex" complex="false">
    <LocalAttribute
localClass="uradiotv_xml.consumerelectronics" >
      <SimpleMapping
attributeName="audioandvisualequipment" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global7" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.compactdiscplayers" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.digitalvideodiscplayers" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
    <LocalAttribute
localClass="uradiotv_xml.laserdiscplayers" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global8" iffused="true">
  <GlobalAttribute name="cod" domain="string"
complex="false">
    <LocalAttribute
localClass="uradiotv_xml.remotecontrol" >
      <SimpleMapping attributeName="cod" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>

<GlobalClass name="Global9" iffused="true">
  <GlobalAttribute
name="promotgiftselectronicsdatabanks"
domain="complex" complex="false">
    <LocalAttribute
localClass="eradiotv_xml.promotionalarticle" >
      <SimpleMapping
attributeName="promotgiftselectronicsdatabanks" />
    </LocalAttribute>
  </GlobalAttribute>
</GlobalClass>
</Schema>

```

Glossario

Agenti software sul Web gli agenti software (software agent) sono programmi che raccolgono informazioni o migliorano altri servizi, operando senza la costante presenza umana o una programmazione precisa delle attività.

Application Program Interface (API) è un metodo specifico prescritto dal sistema operativo o da un programma applicativo, con il quale un programmatore mentre scrive un programma può interrogare il sistema operativo o un'altra applicazione.

Applicazione configurazione persistente o transitoria dei componenti, rivolta a risolvere un problema del cliente, e che può coprire diversi domini.

Applicazioni di front-end e back-end front-end e back-end sono termini utilizzati per caratterizzare interfacce di programmi e servizi relativi all'utente iniziale di queste interfacce e servizi. (l'utente può essere un essere umano o un programma). Un'applicazione front-end si ha quando gli utenti interagiscono in modo diretto, un'applicazione back-end supporta in modo indiretto i servizi di front-end.

Application Service Provider (ASP) una compagnia che offre l'accesso Internet ad applicativi e servizi correlati che altrimenti dovrebbero essere collocati sul proprio computer o su quello della compagnia.

Architettura insieme di componenti.

Client/Server descrive la relazione tra due programmi di cui uno, il client, fornisce un servizio richiesto da un altro programma, il server, che risponde alla richiesta. Sebbene il modello di client/server possa essere utilizzato da programmi all'interno di un singolo computer, esso è molto più importante in una rete in cui fornisce una modalità utile per connettere programmi distribuiti in modo efficiente su varie postazioni.

Common Object Request Broker Architecture (CORBA) è un'architettura per la creazione, distribuzione e gestione di programmi ad oggetti su una rete, quando i programmi sono sviluppati da diversi venditori, in diversi luoghi.

Data Mining Tools strumenti per analizzare ingenti moli di dati allo scopo di individuare particolari tendenze.

Data Warehouse in ambito IT è inteso come sede centrale per le parti significative di dati che i vari sistemi di un'impresa vogliono immagazzinare.

Dominio area, argomento, caratterizzato da una semantica interna, per esempio la finanza, o i componenti elettronici.

Electronic Data Interchange (EDI) è un formato standard per lo scambio di dati nelle attività di business.

Eterogeneità incompatibilità trovate tra risorse e servizi sviluppati autonomamente, che vanno dalla piattaforma utilizzata, sistema operativo, modello dei dati, alla semantica, ontologia.

File Transfer Protocol (FTP) è il metodo più comune per effettuare il download e l'upload di file attraverso una connessione Internet. Il File Transfer Protocol è un metodo standardizzato per connettere computer in modo che i file possano essere condivisi.

HyperText Markup Language (HTML) è considerato la base del World Wide Web. Questo linguaggio consente infatti di creare in maniera standardizzata pagine di informazioni formattate in grado di raggiungere, tramite Internet, un numero di utenti in costante aumento.

Gerarchia struttura di un modello che assegna ogni oggetto ad un livello e definisce per ogni oggetto l'oggetto da cui deriva.

Graphical User Interface (GUI) - pronunciata GOO-ee - è un'interfaccia utente grafica di un computer.

Host gli Internet Host permettono agli utenti su una macchina Client la connessione, la condivisione di file e il trasferimento di informazioni.

Interoperare combinare sorgenti e domini multipli.

Mediatore componente che fornisce i servizi di mediazione e che provvede a dare valore aggiunto alle informazioni che sono trasmesse al cliente in risposta ad un'interrogazione.

Metadata: informazione descrittiva relativa ai dati di una risorsa, compresi il dominio, la proprietà, le restrizioni, il modello di dati.

Object Linking and Embedding (OLE) è un'interfaccia API per accedere a diverse sorgenti di dati.

Open Database Connectivity (ODBC) è un'interfaccia per la programmazione di applicazioni API dalle quali accedere ai database. ODBC è uno standard di tipo open source.

Resource Description Framework (RDF) fornisce uno strumento per aggiungere semantica ad un documento senza fare alcuna assunzione relativa alla sua struttura. RDF è un'infrastruttura che rende possibile la codifica, lo scambio ed il riutilizzo di metadati strutturati.

Schema lista delle relazioni, degli attributi e, quando possibile, degli oggetti, delle regole, e dei metadati di un database. Costituisce la base dell'ontologia della risorsa.

Semantico che si riferisce al significato di un termine, espresso come un insieme di relazioni.

Sintattico che si riferisce al formato di un termine, espresso come un insieme di limitazioni.

Structured Query Language (SQL) è un linguaggio interattivo per ottenere informazioni da un database e aggiornarlo. SQL è uno standard ANSI e ISO.

Secure Shell (SSH) è un programma per connettersi in rete ad un altro computer per eseguire comandi su una macchina remota e per spostare file da un computer all'altro. Fornisce comunicazioni sicure grazie all'autenticazione utente.

Stato istanza o versione di una base di dati o informazioni.

Strumento (tool) programma software che realizza un servizio, tipicamente indipendentemente dal dominio.

Tag letteralmente "etichetta", nel contesto si fa riferimento a tag utilizzati per evidenziare, in un formato gerarchico, i differenti componenti del documento.

Transmission Control Protocol/Internet Protocol (TCP/IP) è il protocollo standard di comunicazione richiesto dai computer per Internet. Per comunicare utilizzando TCP/IP, i PC necessitano di componenti software chiamati TCP/IP stack. I sistemi UNIX sono progettati con le funzionalità TCP/IP.

X-server è un server di connessione a un Terminale X, una rete distribuita che usa X Window System.

X Window System è un sistema aperto client/server per la gestione delle interfacce grafiche a finestra in una rete.

XSL-T è un linguaggio per trasformare documenti XML in altri documenti XML. XSL-T è progettato per essere utilizzato come parte di XSL. In più di XSL-T, XSL include un vocabolario XML per specificare la formattazione. XSL specifica lo stile di un documento, XML utilizza XSL-T per descrivere come il documento è trasformato in un altro documento XML che usa il vocabolario di formattazione.

Warehouse database che contiene o dà accesso a dati selezionati, astratti e integrati da una molteplicità di sorgenti. Tipicamente ridondante rispetto alle sorgenti di dati.

Wrapper strumento utilizzato per accedere alle risorse conosciute e per tradurre i suoi oggetti.

Bibliografia

[BCG, 02] The Boston Consulting Group, U.S. Business-to-Business e-commerce to reach \$4.8 trillion in 2004, Research Bulletin, settembre 2000

[Beneventano, 97a] D. Beneventano, S. Bergamaschi, C. Sartori, and M. Vincini, Odb-tools: a description logics based tool for schema validation and semantic query optimization in object oriented databases. In *Sesto Convegno AIIA - Roma*, 1997

[Beneventano, 97b] D. Beneventano, Sonia Bergamaschi, Claudio Sartori, Maurizio Vincini, ODB-QOPTIMIZER: A tool for semantic query optimisation in oodb. In *Int. Conference on Data Engineering - ICDE97*, 1997

[Beneventano, 00] D. Beneventano, S. Bergamaschi, A. Corni, R. Guidetti, G. Malvezzi, SIDesigner un tool di ausilio all'integrazione intelligente di sorgenti di informazione. In *SEBD: Sistemi Evoluti per Basi di Dati*, L'Aquila, Giugno 2000

[Benjamins, 99] V.R. Benjamins, D. Fensel, S. Decker, A. Gomez Perez, (KA): Building Ontologies for the Internet: a Mid Term Report, in *International Journal of Human-Computer Studies*, 51:687-712, 1999

[Bergamaschi, 98] S. Bergamaschi, D. Beneventano, S. Castano and M. Vincini, Integrazione di informazione: il linguaggio ODLI3 e la logica descrittiva OLCD, *Tecnical report T3-R03*, 16 Luglio 1998

[Bergamaschi, 99] S. Bergamaschi, C. Sartori, M. Vincini and D. Beneventano, Semantic Integration of Heterogeneous Information Sources, in *Journal of Data and Knowledge Engineering*, 1999

[Bergamaschi, 02] S. Bergamaschi, F. Guerra, M. Vincini, A data integration framework for E-commerce product classification, *1st International Semantic Web Conference (ISWC2002)*, Sardegna, 9-12 Giugno 2002

[Berners-Lee, 01] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web: a new form of Web content that is meaningful to computers will unleash a revolution of new possibilities, in *Scientific American*, maggio 2001

[Bernstein, 01] P.A. Bernstein, E. Rahm, A survey of approaches to automatic schema matching, *Springer-Verlang*, 2001

[Bertelè, 02] U Bertelè, A, Rangone, a cura di, eMarketplace: quale ruolo nel B2B italiano?, Associazione Impresa Politecnico, Milano, marzo 2002

[Bharat, 98] Bharat & Broder, A technique for measuring the relative size and overlap of public Web search engine, in *Proceedings of the 7th International WWW Conference*, Brisbane, Aprile 1998

[Buneman, 97] P. Buneman, Semistructured Data, in *Proceedings of Symposium on Principles of Database systems*, 1997

[Catarci, 93] T. Catarci and M. Lenzerini, Representing and using interschema knowledge in cooperative information systems. *Journal of Intelligent and Cooperative Information Systems*, volume 2/4, pages 375-398, 1993

[Cattell, 94] R.G. Cattell, The Object Database Standard: ODMG93. Morgan Kaufmann Publishers, San Mateo, CA, 1994

[Cattell, 97] R.G.G. Cattell e altri, The Object Data Standard: ODMG 2.0, Morgan Kaufmann Publishers, San Francisco, CA, 1997

[Census, 02] AA.VV., A Nation online, How Americans Are Expanding Their Use of the Internet, U.S. Department of Commerce, Economics and Statistics Administration, National Telecommunications and Information Administration, Febbraio 2002

[Clark, 99a] J. Clark, XSL Transformation (XSLT) Version 1.0, W3C Working Draft, Agosto 1999

[Clark&DeRose, 99] J. Clark, S. DeRose, XML Path Language (XPath) Version 1.0 W3C Recommendation, Novembre 1999

[CommerceOne, 00] Commerce One, Inc.: Commerce One XML Common Business Library (xCBL) 3.0, Press releas made at the eLink Conference, Hong Kong, Novembre 2000

[CNR, 00] AA.VV., Ontologia, azienda integrata e commercio elettronico, CNR Istituto di Sistemica e Bioingegneria, Report 2000

[Deloitte, 01] 2001 Technology Trends Annual Report, in *Technology & Communications practices of Deloitte & Touche LLP and Deloitte Consulting LLC*, San Jose (CA) , Maggio 2001

[Ding, 02] Y. Ding, D. Fensel, M. Klein, B. Omelayenko, The Semantic Web: Yet Another Hip, in *Data and Knowledge Engineering*, 2002

- [Duschka, 96] Oliver M. Duschka and Micheal R. Genesereth. Infomaster - An Information Integration Toolkit. Technical Report, Department of Computer Science, Stanford University, 1996
- [ECCMA, 02] UNSPSC, Technical Manual Working, 06-19-2002
- [Ecl@ss, 00] Ecl@ss, White Paper, Version 0.6, 2000-01-07
- [Felbaum, 98] Felbaum, C.(ed), WordNet: An Electronic Lexical Database, with a preface by George Miller, The MIT Press, 1998
- [Fensel, 01a] D. Fensel, *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*, Springer-Verlang, Berlin 2001
- [Fensel, 01b] D. Fensel, Y. Ding, B. Omelayenko, E. Schulten, G. Botquin, M. Brown, A. Flett, Product Data Integration in B2B E-Commerce, *IEEE Intelligent System*, 16(4), 2001
- [Fensel, 01c] D. Fensel, I. Horrocks, F. Van Harmelen, D. McGuinness, P.F. Patel-Schneider, OIL: Ontology Infrastructure to Enable Semantic Web, *IEEE Intelligent Systems*, 2001
- [Fensel&Groenboom, 97] D. Fensel, R. Groenboom, Specifying Knowledge-based Systems with Reusable Components, in *Proceedings 9th International Conference on Software Engineering and Knowledge Engineering (SEKE 97)*, Madrid 1997
- [Fensel&Musen, 01] D. Fensel, M. Musen, Special Issue on Semantic Web Technology, *IEEE Intelligent Systems (IEEE IS)*, 2001
- [Ferrari, 00] M. Ferrari, Progetto e realizzazione di tecniche di Object Fusion nel sistema MOMIS, Tesi di Laurea, Università di Modena e Reggio Emilia, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 2000
- [Fox, 93] M.S. Fox, J.F. Chionglo, F.G. Fadel, A common Sense Model of the Enterprise, in *Proceedings of the 2nd Industrial Engineering Research Conference*, Norcross GA, USA, 1993
- [Fox&Gruninger, 97] M.S. Fox and M.Gruninger, On Ontologies and Entrprise Modelling, in *Proceedings of the International Conference on Enterprise Integration Modelling Technology 97*, Springer-Verlang, 1997
- [Franceschi, 00] M. Franceschi. Il componente query manager di MOMIS: utilizzo della conoscenza estensionale. Tesi di Laurea, Università di Modena e Reggio Emilia, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 2000
- [Fridman-Noy & Hafner, 97] N. Fridman-Noy, C.D. Hafner, The State of the Art in Ontologies Design, in *AI Magazine*, 1997
- [Granada, 02] Granada Research, White Paper, Why Coding and Classifying Products is Critical to Success in Electronic Commerce, 2002 www.un-spsc.net

- [Gruber, 93] T. R. Gruber, A Translation Approach to Portable Ontology Specifications, in *Knowledge Acquisition*, 1993
- [Guarino, 1998] N. Guarino (ed), Formal Ontology in Information System, IOS Press, Amsterdam, 1998
- [Guerra, 00] F. Guerra, MOMIS: Il wrapper per sorgenti di dati XML. Tesi di Laurea, Università di Modena e Reggio Emilia, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, A.A.1999-2000
- [Guidetti, 00] R. Guidetti. SI-Designer: un tool per l'integrazione di sorgenti distribuite ed eterogenee Tesi di Laurea, Università di Modena, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 2000
- [Hull, 95] R. Hull and R. King et al. Arpa I^{3a} reference architecture, 1995. Available at <http://www.isse.gmu.edu/I3 Arch/index.html>
- [ICE, 02] ICE (Istituto Nazionale per il Commercio Estero), a cura di, Speciale E-Procurement in Italia, in Dossier eMarketServices, maggio 2002
- [IOTP, 99] D. Burdett, Internet Open Trading Protocol – IOTP, Version 1.0. *Proceedings of the 46-th Internet Engineering Task Force*, Hosted by Nortel Networks, Inc., Washington, DC, USA, November 7-12, 1999
- [Karp, 99] P. D. Karp, V. K. Chaudhri, and J. Thomere, XOL: An XML-Based Ontology Exchange Language, Version 0.3, July 3, 1999
- [Keenan, 01] Keenan Vision, E-Merchant 2001: Accelerating Free Trade, in Keenan Report, Luglio 2001
- [Kent, 99] R. E. Kent, Conceptual Knowledge Markup Language. In *Proceedings of the Twelfth Workshop on Knowledge Acquisition, Modeling and Management (KAW99)*, Banff, Ottobre 1999
- [Lenat&Guha, 90].D. B. Lenat, R.V. Guha, Building large knowledge-based systems. Representation and inference in the Cyc project, Addison-Wesley, Reading, Massachusetts, 1990
- [Lief, 99] V. Lief with B.D. Temkin, K. McCarthy, J. Sharrard, T.O. Brown, Net Marketplaces Grow Up, in *TechStrategy Report*, December 1999
- [Luke, 97] S. Luke, L. Spector, D. Rager, and J. Hendler, Ontology-based Web agents In *First International Conference on Autonomous Agents (AA'97)*, 1997
- [Meersman, 99] R.A. Meersman, Semantic Ontology Tools in IS Design, in ISMIS 99 Conference in Warsaw, June, 1999
- [Melnik&Decker, 00] S. Melnik, S. Decker, A layered approach to information modeling and interoperability on the Web, in *ECDL 2000 Workshop on the Semantic Web*

- [Miller, 95] A.G. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995
- [Miller, 98] E. Miller, An Introduction to the Resource Description Framework, *Dlib Magazine*, May 1998
- [Montanari 98] S. Montanari. “Un approccio intelligente all’Integrazione di Sorgenti Eterogenee di Informazione”. Tesi di Laurea, Università di Modena, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 1998
- [Neri, 99] S. Neri, Progetto e realizzazione di un catalogo virtuale per e-commerce con il sistema MOMIS. Tesi di laurea, Università di Modena e Reggio Emilia, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, A:A 1999-2000
- [OAGIS, 00] Open Applications Group Inc. Open Applications Group Integration Specification. OAGIS Release 7.0.2, 2000
- [Omelayenko, 01a] B. Omelayenko, D. Fensel, An analysis of B2B catalogue integration problems, in *Proceedings of the International Conference on Enterprise Information System (ICEIS-2001)*, Luglio 2001
- [Omelayenko, 01b] B. Omelayenko, D. Fensel, A two-Layered Integration Approach For Product Information in B2B E-commerce, *Second International Conference on Electronic Commerce and Web Technologies, EC-Web*, Settembre 2001
- [Omelayenko, 01c] B. Omelayenko, D. Fensel, A Layered Integration Approach for Product Descriptions in B2B e-commerce, 2001
- [Prevete, 01] B. Prevete, K. Barzan, M Del Giaccio, a cura di, Survey2001: eMarketplace in Italia, e ruolo degli eMarketplace nel commercio internazionale, Istituto nazionale per il Commercio Estero, ottobre 2001
- [Ricciardi, 99] S. Ricciardi, Tesi di Laurea, Rimozione dell’ambiguità nell’interazione tra *WordNet* e il sistema MOMIS Università di Modena e Reggio Emilia, Facoltà di Ingegneria sede di MO, Corso di laurea in Ingegneria Informatica, A:A 1999-2000
- [Roth, 97] M.T. Roth, M.P. Schwartz, Don’t Scrap It, Wrap It! A wrapper Architecture for Legacy Data Sources, in *Proceedings of 26th Int. Conf. On Very Large DataBases*, 266-275, 1997
- [Schulten, 01] E. Schulten, H Akkermans, N. Guarino, G. Botquin, N. Lopes, M. Dörr, N. Sadeh, The E-Commerce Product Classification Challenge, in *IEEE Intelligent Systems Magazine, special issue on Intelligent E-business*, Luglio/Agosto 2001
- [Seybold, 98] P. B. Seybold, R. T. Marshak, Customer.com, *Time Business*, 1998
- [Studer, 96], R. Studer, H. Eriksson, J.H. Gennari, S.W. Tu, D. Fensel, M. Musen, Ontologies and the Configuration of Problem-Solving Method, in B.R. Gaines and M. Musen, (eds), *Proceedings of the 10th Banff Knowledge Acquisition for Knowledge-based Systems Workshop*, Banff, Canada, 1996

[UN Report, 01] AA.VV., United Nations Conference on trade and development: E-commerce and development, United Nations, New York and Geneva, Report 2001

[Van Heijst, 97] G. Van Heijst, A. Schreiber, B. Wielinga, Using Explicit Ontologies in KBS development, in *International Journal of Human Computer Studies*, 1997

[Weibel, 95] S. Weibel, J. Gridby, E. Miller, OCLC/NCSA Metadata Workshop Report, Dublin, EUA, 1995

[Wiederhold, 96] G. Wiederhold et al. Integrating Artificial Intelligence and Database Technology, volume 2/3. *Journal of Intelligent Information Systems*, Giugno 1996

[Zaccaria, 98] A. Zaccaria, MOMIS: il componente Query Manager. Tesi di laurea, Università di Modena e Reggio Emilia, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, A:A 1997-1998

Siti Internet di riferimento

www.dbgroup.unimo.it sito gestito dal Dipartimento di Ingegneria dell'Informazione dell'Università di Modena e Reggio, si possono reperire tutte le informazioni relative al sistema MOMIS.

www.w3c.org il World Wide Web Consortium (W3C) sviluppa tecnologie interoperabili con lo scopo di portare Internet a piene potenzialità. In particolare si trovano informazioni sull'XML.

www.ontoweb.org portale gestito dall'Università di Amsterdam ed in particolare dal professor Dieter Fensel. Lo scopo del sito è rendere possibile lo scambio di informazioni per la gestione della conoscenza e del commercio elettronico.

www.emarketservices.it osservatorio sull'e-Business Internazionale, nasce dalla collaborazione tra l'Istituto nazionale per il Commercio Estero e le principali Agenzie europee per la promozione del commercio estero.

www.bcg.com sito ufficiale del Boston Consulting Group, una società di consulenza gestionale, leader mondiale nell'individuazione e sostegno di un vantaggio competitivo.

www.unctad.org/ecommerce sito dell'UNCTAD Electronic Commerce Branch, offre politiche di sostegno per Paesi in via di sviluppo nell'introduzione delle Information Technology.

www.keenanvision.com sito ufficiale della società Keenan Vision, specializzata nell'analisi e ricerca di soluzioni basate su Internet.

www.forrester.com sito ufficiale della Forrester Research, società di individuazione e analisi dei trend emergenti nello sviluppo tecnologico e del loro impatto sull'economia.

www.dc.com sito ufficiale di Deloitte Consulting, una delle società di consulenza più affermate nel campo manageriale, nota per la forza con cui affronta le sfide più complesse.

www.census.gov sito ufficiale del U.S.Census Bureau, ente di statistica del Dipartimento per il Commercio degli Stati Uniti.

www.eclass.de sito ufficiale dell'omonima classificazione.

www.un-spssc.net oppure www.unspssc.com sito gestito da UNDP che attualmente gestisce e distribuisce la codifica UNSPSC.

www.eccma.org sito ufficiale di Electronic Commerce Code Management Association (ECCMA), associazione che fino all'ottobre 2002 ha gestito e distribuito la codifica UNSPSC.

www.ntis.gov/product/naics.htm il National Technical Information Service è la maggior fonte di informazioni relative alle attività tecniche scientifiche ed ingegneristiche del governo statunitense. Nel sito si possono reperire informazioni relative alla codifica NAICS e si possono acquistare libri e CD che la riguardano.

www.census.gov/epcd/www/naics.html sito gestito dal Census Bureau che fornisce informazioni relative alla codifica NAICS.

www.naics.com/search.htm di proprietà della NAICS Association che fornisce una vasta gamma di informazioni relative a NAICS e SIC e permette di ricercare codici all'interno delle codifiche.

www.statcan.ca/english/Subjects/Standard sito ufficiale di NAICS Canada.

www.siccode.com fornisce informazioni sulle codifiche SIC e NAICS e permette di cercare codici al loro interno.

www.bls.gov/bls/naics.htm fornisce informazioni relative alle codifiche NAICS e SIC, gestito da U.S.Department of Labor.

www.rosettanel.org sito ufficiale di Rosettanet, consorzio delle maggiori compagnie di Information Technology, Electronic Components e Semiconductor Manufacturing, si occupa della creazione e dell'implementazione di standard per processi di business.

www.gradaresearch.com sito ufficiale del Granada Research che è una società di consulenza con base a Half Moon Bay, in California, specializzata in sistemi di B2B.

www.ietf.org sito del consorzio Internet Engineering Task Force (IETF), fornisce IOTP (un formato standard per le operazioni di pagamento per le transazioni svolte via Internet).

www.rets-wg.org sito del Real Estate Transaction Standard (RETS) nuovo standard aperto per lo scambio di informazioni per le transazioni di business. Dal sito si può scaricare la versione aggiornata di Real Estate Transaction Markup Language (RETML)

www.openapplications.org sito di Open Applications Group che è un consorzio industriale pubblica specificazioni per i contenuti di business nel settore delle applicazioni specifiche per messaggi di business e scenari di integrazione.

<http://www.cogsci.princeton.edu/~wn/> sito gestito dal Cognitive Science Laboratory dell'Università di Princeton, dal quale è possibile scaricare la versione aggiornata di WordNet, i manuali e le pubblicazioni che riguardano tale dizionario lessicale.

<http://www.let.uva.nl/~ewn> sito che cura la versione multi-lingue europea di WordNet.

www.cyc.com sito ufficiale del Cycorp, Inc. organizzazione che gestisce e distribuisce l'ontologia lessicale CYC.

<http://www.eil.utoronto.ca/tove/toveont.html> sito dedicato allo sviluppo del progetto ontologico TOVE.

<http://ka2portal.aifb.uni-karlsruhe.de/> sito curato dall'Università di Karlsruhe (Germania) per lo sviluppo del progetto KA².

www.buyusa.com BuyUSA permette l'incontro di compratori e venditori in un ambiente on-line supportato dal U.S. Department of Commerce.

www.bravobuild.com creato nel luglio 2000 da BravoSolution, società leader nella fornitura di soluzioni e strumenti di procurement basati su tecnologia Internet, BravoBuild è un marketplace sviluppato per rispondere al meglio alle esigenze specifiche di approvvigionamento dell'industria delle costruzioni.

www.italbiz.com Italbiz aggrega la più grande comunità di aziende straniere e italiane che operano nel mercato internazionale e aiuta ad aumentare le esportazioni, a trovare nuovi fornitori e ad aprire nuovi mercati esteri.

www.ceramicandmore.com eMarketplace verticale realizzato in italiano e inglese, contiene un database di 2000 aziende del comparto (materia prima, produzione industriale, lavorazione artigianale, macchine per lavorazione) e vuole essere il punto di riferimento per chi apprezza, produce, ricerca ceramiche di qualità.

RINGRAZIAMENTI

Ringrazio il Professore Domenico Beneventano per l'aiuto e la disponibilità forniti durante la realizzazione della presente tesi; un ringraziamento va inoltre all'Ing. Francesco Guerra e all'Ing. Maurizio Vincini per la preziosa collaborazione.

Ringrazio con affetto tutti gli amici e amiche del Corso di Laurea che mi sono stati vicini in questi anni di studio e hanno reso indimenticabili tanti momenti vissuti insieme.

Ringrazio Stefania con la quale ho lavorato al progetto sviluppato nella tesi e condiviso giornate e trovate incredibili (?freezanti?).

Un ringraziamento speciale va alla mia famiglia, che mi ha sostenuto durante i circa vent'anni dedicati allo studio e ha riposto fiducia in me, lasciandomi indipendenza nelle decisioni di vita.

Infine non posso dimenticare gli amici e amiche extra-universitari che mi hanno sopportato ogniqualevolta ero stressata o irreperibile per gli esami e quelli che mi hanno coinvolto in esperienze importanti al punto da mettere in secondo piano gli studi.