

UNIVERSITÀ DEGLI STUDI DI MODENA E  
REGGIO EMILIA

Facoltà di Ingegneria- Sede di Reggio Emilia  
Corso di Laurea in Ingegneria Gestionale

---

---

**APPLICABILITÀ E CRITICITÀ DEL SISTEMA  
A MEDIATORE MOMIS NELL'INTEGRAZIONE DI  
STANDARD DI CLASSIFICAZIONE PER  
L'E-COMMERCE**

Relatore  
Prof.Ing.Domenico Beneventano

Tesi di Laurea di  
Stefania Magnani

Correlatore  
Ing.Maurizio Vincini



*Parole chiave:*

---

E-marketplace

Standard di classificazione dei prodotti

Integrazione Intelligente delle Informazioni

MOMIS

Mapping



*A mamma e papà*



# Indice

Introduzione .....	1
Capitolo 1 .....	5
Il commercio elettronico .....	5
1.1 Generale .....	5
1.2 Quali sono le differenze tra l'e-commerce B2B e B2C? .....	7
1.3 B2B e-commerce .....	8
1.4 eMarketplace .....	9
1.4.1 Cosa fanno gli e-marketplace .....	10
1.4.2 Gli obiettivi e i rischi degli e-marketplace .....	12
1.4.3 I benefici del B2B e-marketplace per le imprese .....	13
1.4.4 Il successo e il fallimento di un e-market .....	14
1.5 Traguardi dell'e-commerce .....	14
Capitolo 2 .....	17
Standard di classificazione dei prodotti .....	17
2.1 Generale .....	17
2.2 Dalle transazioni elettroniche alle decisioni strategiche .....	18
2.3 Perché classificare prodotti e servizi .....	21
2.3.1 Per trovare e acquistare .....	21
2.3.2 Per il marketing e la distribuzione .....	24
2.4 Le differenze tra classificazione e identificazione .....	26
2.5 Le caratteristiche di una convenzione per la denominazione dei beni .....	28
2.5.1 Tassonomia gerarchica per analisi "Drill down" e "Roll up" .....	28
2.5.2 Chiavi di identificazione .....	29
2.5.3 Caratteristiche dello schema .....	30
2.6 Ecl@ss .....	31
2.6.1 La nascita .....	31
2.6.2 L'estensione degli obiettivi .....	32

2.6.3 Verso uno standard globale .....	34
2.6.4 Funzioni e applicazioni .....	34
2.6.5 La struttura .....	35
2.6.6 Le possibilità di ricerca con ecl@ss .....	40
2.6.7 Aggiornamento e mantenimento .....	41
2.6.8 Prospettive future di ecl@ss.....	42
2.6.9 Il mantenimento e la gestione della qualità.....	43
2.7 UNSPSC (Universal Standard Products and Services Classification).....	44
2.7.1 Obiettivo.....	44
2.7.2 Ambito.....	44
2.7.3 Classificazione gerarchica.....	45
2.7.4 Commodity Identifier (EGCI).....	46
2.7.5 Classificazione VS identificazione.....	46
2.7.6 Distribuzione dei files .....	47
2.7.7 ECCMA Membership .....	48
2.7.8 Gestione dei processi di aggiornamento.....	48
2.7.9 Servizi di implementazione.....	50
2.7.10 Electronic Commerce Code Management Association (ECCMA).....	51
2.8 NAICS:dati nuovi per un'economia nuova.....	54
2.8.1 Background .....	54
2.8.2 Classificazione delle nuove industrie.....	55
2.8.3 Più settori economici.....	57
2.8.4 Concetti rivisti della classificazione.....	58
2.8.5 I nuovi sistemi di numerazione .....	59
2.8.6 Le prospettive del Nord America .....	61
2.8.7 Adattamento a NAICS .....	61
Capitolo 3 .....	63
Le ontologie: punti di forza nella gestione del commercio elettronico.....	63
3.1 Generale .....	63
3.2 Extranet : B2B.....	65
3.3 Le ontologie.....	66
3.4 Il commercio elettronico come area di applicazione delle ontologie.....	67
3.4.1 Shopbot.....	68
3.4.2 Gli agenti per il commercio elettronico e l'XML .....	69
3.4.3 Sistema di commercio elettronico .....	70
3.4.4 Le ontologie di prodotto e il commercio elettronico.....	71
3.5 Esempi di ontologia.....	72
3.5.1 WordNet.....	72
Capitolo 4.....	83
I problemi dell' integrazione .....	83
4.1 Generale .....	83
4.2 La gestione delle informazioni.....	83



4.3 L'integrazione delle informazioni .....	84
4.4 I "Content Standard" .....	85
4.5 L'infrastruttura dei documenti per il B2B .....	88
4.6 Standard per i documenti .....	88
4.6.1 Esempio di integrazione di documenti.....	89
4.7 Modello per l'integrazione del catalogo .....	94
4.8 Mapping dei modelli di dati.....	95
Capitolo 5.....	97
L'Integrazione Intelligente con MOMIS .....	97
5.1 L'Integrazione delle Informazioni .....	97
5.1.1 Integrazione Intelligente delle Informazioni.....	98
5.1.2 Il programma I <sup>3</sup> .....	98
5.1.3 Architettura di riferimento per sistemi I <sup>3</sup> .....	100
5.1.4 Il mediatore .....	102
5.2 Il sistema MOMIS .....	104
5.2.1 Il linguaggio ODLI <sup>3</sup> .....	105
5.2.2 Gli strumenti di MOMIS.....	106
5.3 Il processo d'integrazione .....	107
5.4 Generazione del Thesaurus comune .....	108
5.4.1 Estrazione delle relazioni intra-schema .....	109
5.4.2 Estrazione delle relazioni inter-schema .....	109
5.4.3 Arricchimento dell'insieme delle relazioni.....	110
5.4.4 Validazione delle relazioni .....	110
5.4.5 Inferenza di nuove relazioni .....	110
5.5 Il calcolo delle affinità .....	111
5.6 La generazione dei cluster .....	112
5.7 La costruzione delle classi globali .....	112
5.8 Fusione degli attributi .....	113
5.8.1 Fusione degli attributi di relazioni validate .....	113
5.8.2 Fusione di attributi di relazioni non validate .....	114
5.9 La generazione delle mapping table .....	114
5.10 SI-Designer .....	115
5.10.1 L'automa di utilizzo di SI-Designer .....	116
5.10.2 Source .....	117
5.10.3 SIM A .....	118
5.10.4 SLIM.....	119
5.10.5 SIM B.....	122
5.10.6 EXTM Rel.....	122
5.10.7 Cluster .....	123
5.10.8 TUNIM .....	125
Capitolo 6.....	129
Integrazione degli standard di classificazione dei prodotti con MOMIS .....	129

6.1 Generale .....	129
6.2 Lo stato dell'arte.....	130
6.2.1 La ricerca di una Lingua Franca o della Stele di Rosetta.....	131
6.2.2 L'obiettivo: arrivare ad un ontologia comune tra i diversi sistemi di classificazione .....	133
6.2.3 Esempio: Mapping tra un frammento di Ecl@ss e UNSPSC .....	134
6.2.4 La riconciliazione degli standard mediante il sistema MOMIS.....	136
6.3 Analisi dell'approccio MOMIS.....	137
6.3.1 Acquisizione dei codici .....	137
6.3.2 Selezione dei frammenti di codice .....	141
6.3.3 Traduzione dei frammenti .....	150
6.3.4 Wrapper XML .....	152
6.3.5 SI-Designer.....	153
6.3.6 Analisi dei risultati .....	161
Capitolo 7 .....	165
Analisi dei mapping .....	165
7.1 Generale .....	165
7.2 Difficoltà legate al sistema.....	165
7.2.1 WordNet: un dizionario generico .....	166
7.2.2 Problemi di ripetizione .....	167
7.2.3 SLIM: un tool rigido .....	169
7.2.4 Lettura parziale delle gerarchie .....	170
7.3 Difficoltà legate ai codici .....	172
7.3.1 Parziale sfruttamento delle gerarchie .....	172
7.3.2 Diversi obiettivi alla base delle classificazioni .....	173
7.4 Esempi di integrazione .....	174
7.4.1 Integrazione dei frammenti relativi a Book.....	174
7.4.2 Integrazione dei frammenti relativi a Pen .....	179
7.4.3 Integrazione dei frammenti relativi a Food .....	183
7.4.4 Integrazione dei frammenti relativi a Farm.....	187
Conclusioni.....	191
Appendice A.....	193
Appendice B.....	199
Appendice C .....	221
Glossario.....	225
Bibliografia.....	229

# Indice delle figure

Figura 1: Il flusso delle informazioni .....	9
Figura 2: Architettura del marketplace dal punto di vista del compratore .....	11
Figura 3: Architettura del marketplace dal punto di vista del commerciante.....	11
Figura 4: Necessità di integrazione e di standardizzazione .....	20
Figura 5:Incontro tra domanda e offerta .....	22
Figura 6:La funzione di ecl@ss .....	33
Figura 7:La struttura di ecl@ss.....	36
Figura 8: Significato della classificazione .....	38
Figura 9:Procedura di aggiornamento.....	42
Figura 10: L'architettura degli shopbot.....	69
Figura 11: Overview di WordNet .....	73
Figura 12: Relazione di iponimia per "machine" .....	74
Figura 13: Lista dei 9 beginner .....	75
Figura 14: Relazione tra lemmi e significati.....	75
Figura 15: Matrice Lessicale.....	76
Figura 16: Standard orizzontali e verticali.....	86
Figura 17: Integrazione con n*m mapping .....	92
Figura 18: Integrazione con n+m mapping.....	93
Figura 19: Mapping dei modelli di dati .....	95
Figura 20: Diagramma dei servizi I <sup>3</sup> .....	99
Figura 21: Architettura del sistema MOMIS .....	105
Figura 22: Struttura di ODB Tools .....	107
Figura 23: Fasi del processo di integrazione .....	108
Figura 24: Esempio di Mapping table.....	114
Figura 25:L'architettura del sistema MOMIS .....	117
Figura 26: Pannello Source .....	118

Figura 27: Pannello Thes.Rel: run SIM .....	119
Figura 28: Pannello di SLIM.....	120
Figura 29:Pannello Hypernym .....	121
Figura 30: Pannello EXTM Rel .....	123
Figura 31:Pannello Cluster: ARTEMIS Configuration .....	124
Figura 32: Pannello Cluster: Cluster Tuning .....	125
Figura 33: Pannello TUNIM .....	126
Figura 34:Mapping tra le codifiche.....	135
Figura 35: Ecl@ss- Paper Product .....	136
Figura 36: UNSPSC- Paper Product .....	136
Figura 37: Integrazione Paper Product con MOMIS .....	137
Figura 38: Download di ecl@ss dal sito Internet.....	138
Figura 39: Sito di UNSPSC.....	138
Figura 40: Tabella delle corrispondenze tra NAICS e SIC.....	139
Figura 41: Motore di ricerca su ecl@ss.....	140
Figura 42: NAICS- Agriculture, forestry, fishing and hunting .....	143
Figura 43: UNSPSC -Live plant, animal material and accessories and supplies.....	143
Figura 44: Individuazione dei frammenti di codice .....	144
Figura 45: UNSPSC -ECL@SS : PEN.....	145
Figura 46: UNSPSC- ECL@SS: HARDWARE .....	145
Figura 47: UNSPSC- ECL@SS: TOY .....	146
Figura 48: UNSPSC- ECL@SS: BOOK.....	148
Figura 49: NAICS- ECL@SS: MINING .....	149
Figura 50: MOMIS- Sources.....	154
Figura 51: MOMIS-SIMA .....	155
Figura 52:MOMIS-Inserimento forma base.....	156
Figura 53:MOMIS-Assegnazione del significato .....	157
Figura 54: MOMIS-inserimento relazioni.....	158
Figura 55: MOMIS-Global Classes.....	159
Figura 56: MOMIS-TUNIM .....	161
Figura 57:Mapping “Toy” .....	163
Figura 58:Gerarchia sul dominio Bottle.....	166
Figura 59: Trasformazione delle gerarchia nelle sorgenti.....	170
Figura 60: Ripetizioni nella gerarchia.....	172
Figura 61: Sfruttamento parziale della gerarchia .....	172
Figura 62: Gerarchia non ottimizzata.....	173
Figura 63: Contesto di classificazione dei prodotti.....	173
Figura 64: UNSPSC-ECLASS :BOOK.....	175
Figura 65: Mapping semplice sul dominio Book.....	176
Figura 66: Mapping sul dominio Book .....	177
Figura 67: Mapping ampio sul dominio Book .....	178
Figura 68:UNSPSC-ECLASS: PEN .....	179

Figura 69: Mapping sul dominio PEN.....	182
Figura 70: NAICS –FOOD .....	185
Figura 71: UNSPSC-FOOD.....	185
Figura 72: ecl@ss -FOOD .....	186
Figura 73: NAICS-FARM .....	189
Figura 74: UNSPSC- FARM .....	189



## Indice delle tabelle

Tabella 1: Pro e contro della classificazione .....	22
Tabella 2: Classificare prodotti e servizi per supportare le vendite e il marketing .....	24
Tabella 3: Esempi di identificazione .....	26
Tabella 4: Differenze tra classificazione ed identificazione.....	27
Tabella 5:La struttura di ecl@ss .....	36
Tabella 6: Le classi di materiali.....	37
Tabella 7:Le parole chiave.....	39
Tabella 8: Struttura di UNSPSC .....	45
Tabella 9: Esempio della struttura di UNSPSC .....	46
Tabella 10: Le nuove aziende di NAICS .....	56
Tabella 11: Differenze tra SIC e NAICS .....	58
Tabella 12:Settori nuovi e settori mantenuti.....	60
Tabella 13: 1997 NAICS matched to 1987 SIC .....	62
Tabella 14: Mapping degli elementi .....	93
Tabella 15: MOMIS-Classi globali.....	162
Tabella 16: Relazioni inserite manualmente.....	177
Tabella 17:Relazioni modificate.....	178





# Sommario

Un marketplace è il luogo dove si incontrano la domanda e l'offerta di compratori e venditori partecipanti ad un processo di business. Perciò, i mercati elettronici sono comunità virtuali nelle quali i compratori possono trovare l'offerta di svariati venditori e fare la scelta migliore. Nel mondo del commercio elettronico il confronto tra i vari prodotti è ostacolato dalla carenza di standard condivisi che li classifichino e li descrivano. La soluzione a questo problema è la creazione di uno standard unico che possa essere adottato da tutte le aziende, ma questa soluzione richiede tempi molto lunghi e comporta una serie di difficoltà implementative. Ne segue la necessità per gli e-marketplace di individuare un mapping tra i diversi sistemi di classificazione.

In questa tesi si affronta il problema dell'integrazione degli standard di classificazione dei prodotti proponendo una metodologia semiautomatica che stabilisce un mapping tra i diversi standard.

Nei capitoli successivi, dopo aver analizzato le problematiche dell'e-commerce e gli standard di classificazione dei prodotti più diffusi, si valuterà la validità del sistema MOMIS nell'ottenimento del mapping tra le codifiche, rilevando le criticità imputabili al sistema e le difficoltà causate dalla struttura dei codici.



# Introduzione

Negli ultimi anni l'e-commerce ha avuto un rapido sviluppo ed ha fornito alle aziende strumenti per essere competitive in un mercato in continua evoluzione. Una delle attività più promettenti tra quelle emergenti dell'e-commerce sono gli e-marketplace, che forniscono un sito virtuale dove possono essere svolte le attività di business. Gli e-marketplace danno la possibilità ai compratori di analizzare una vasta gamma di prodotti e, alla fine, di procurarsi velocemente i prodotti e i servizi, con meno spese e fatica di quelle richieste dal commercio tradizionale. I venditori a loro volta possono presentare una vasta gamma di prodotti, ridurre le spese di vendita e, se vogliono, competere con prezzi e offerte speciali in real time.

Con i marketplace il collegamento non è più tra venditore e compratore e viceversa ma tra i venditori ( $n$ ) ed il marketplace e i compratori ( $m$ ) ed il marketplace. Pertanto un marketplace funziona come un livello di comunicazione intermedio che riduce il numero di collegamenti necessari da  $n*m$  a  $n+m$ . Per fornire questo servizio è necessario affrontare il problema dell'eterogeneità dei cataloghi dei prodotti e dei documenti. Sviluppare un approccio scalabile all'integrazione delle informazioni è diventato un pre-requisito fondamentale. La carenza di standard, o meglio l'inconsistenza dei nuovi standard generati, rende tutte queste attività complicate.

La codifica di prodotti e servizi, in accordo con una convenzione di classificazione standardizzata, è necessaria per semplificare il commercio tra le compagnie. Prodotti e servizi, identificati in modo non ambiguo, permettono la gestione degli acquisti e l'analisi della spesa. Inoltre, nomi di prodotti comprensibili alla macchina assistono le

funzioni di marketing e vendita nella ricerca di clienti e nel fornire migliori servizi di distribuzione.

Pur ipotizzando che anche sorgenti diverse condividano una visione simile del dominio da modellare, e, quindi, un insieme di concetti comuni, niente ci garantisce che i diversi sistemi usino esattamente gli stessi vocaboli per rappresentare questi concetti, né le stesse strutture. Poiché le diverse sorgenti sono state progettate e modellate da persone differenti, è molto improbabile che queste persone condividano la stessa concettualizzazione del mondo esterno, ovvero non esiste nella realtà una semantica univoca a cui chiunque possa riferirsi. Per questo motivo il marketplace deve fornire uno strumento per mediare tra i differenti standard utilizzati dai partecipanti alla transazione, in modo che i diversi attori del processo di business possano scambiare informazioni utilizzando i loro formati di dati.

Negli ultimi anni molti gruppi di ricerca hanno affrontato il problema dell'integrazione delle informazioni. Uno degli approcci seguiti è quello dei sistemi a mediatore il cui obiettivo è quello di fornire uno schema globale virtuale ed integrato di un insieme di sorgenti che devono essere integrate.

Questa tesi ha lo scopo di valutare l'efficacia di un sistema a mediatore per l'integrazione di sorgenti testuali eterogenee nell'ambito degli standard di classificazione dei prodotti. Il sistema in questione è il sistema MOMIS, in corso di sviluppo presso l'Università di Modena e Reggio Emilia. Partendo da alcune idee preliminari che delineano come utilizzare il sistema MOMIS per ottenere il mapping e la riclassificazione dei prodotti tra differenti sistemi di classificazione, si è esteso il dominio di integrazione considerando vari frammenti di codice, selezionati con criteri diversi, e analizzando vari approcci possibili alle codifiche e al sistema.

L'obiettivo della tesi è, quindi, analizzare l'applicabilità e le criticità del sistema a mediatore MOMIS nell'integrazione degli standard di classificazione dei prodotti, attraverso lo svolgimento una serie di prove che comportano un diverso contributo da parte del progettista e del sistema.

Nel capitolo1 si introduce l'e-commerce, prestando una particolare attenzione agli e-marketplace, alle possibilità offerte da queste nuove strutture di business e alle conseguenti prospettive di miglioramento.

Nel secondo capitolo si analizzano gli standard di classificazione dei prodotti, prima in modo generico, poi studiando approfonditamente la struttura di [ecl@ss](#), UNSPSC e NAICS.

Nel terzo capitolo si introduce il legame tra il concetto di ontologia e l'e-commerce, in particolare, come esempio di ontologia, si descrive il dizionario lessicale WordNet.

Nel quarto capitolo si introduce l'integrazione delle informazioni, prendendo in esame sia l'integrazione dei documenti che l'integrazione degli standard di classificazione dei prodotti.

Nel quinto capitolo vengono descritti dettagliatamente la struttura ed il funzionamento del sistema a mediatore MOMIS.

Nel sesto capitolo si ripercorre un esempio di integrazione degli standard di classificazione dei prodotti mediante il sistema MOMIS, illustrando le operazioni preliminari all'inserimento dei dati da integrare nel sistema e le operazioni che il progettista deve svolgere per ottenere il mapping tra le codifiche.

Infine, nel settimo capitolo si prendono in esame le difficoltà incontrate in fase di integrazione, distinguendo quelle che si possono attribuire al sistema da quelle che si possono attribuire alla struttura dei codici. Il capitolo si conclude con la descrizione di alcune delle integrazioni svolte adottando diversi approcci nei confronti delle codifiche e del sistema.



# Capitolo1

## Il commercio elettronico

### 1.1 Generale

L'e-commerce, commercio elettronico, così come definito dal Ministero dell'Industria, del Commercio e dell'Artigianato, *consiste nello svolgimento di attività commerciali e di transazioni per via elettronica e comprende attività diverse quali: la commercializzazione di beni e servizi per via elettronica, la distribuzione on line di contenuti digitali, l'effettuazione per via elettronica di operazioni finanziarie e di borsa,...*

Molte organizzazioni e individui hanno fornito varie definizioni di commercio elettronico, che, a causa della mancanza di una caratterizzazione condivisa, hanno portato all'elaborazione di considerazioni incompatibili su tale fenomeno. Queste differenze hanno significativamente ostacolato l'abilità dei ricercatori nel capire pienamente le ramificazioni del commercio elettronico. È necessario formare un set di definizioni che vengano accettate internazionalmente, che possano essere applicate per lo sviluppo di indicatori statistici. Il contributo dell'OECD (Organization for Economic Co-operation and Development) nella creazione di una definizione di e-commerce si è tradotto nel riconoscimento di tre dimensioni da esplicitare come parti del processo di definizione. Queste dimensioni riguardano:

- 1) Reti sulle quali le attività rilevanti sono portate avanti.
- 2) Processi che dovrebbero essere inclusi nel dominio del commercio elettronico.
- 3) Attori coinvolti nelle transazioni.

## **1) Reti**

Sono state concordate due definizioni da utilizzare:

### *Definizione ampia*

Una transazione elettronica è la vendita o l'acquisto di beni o servizi tra imprese, consumatori o governo e altre istituzioni pubbliche o private, condotta su reti mediate da computer. I beni e i servizi sono ordinati su queste reti ma il pagamento e la spedizione possono essere condotti sia on line sia off line.

### *Definizione stretta*

Una transazione Internet è la vendita o l'acquisto di beni o servizi tra imprese, privati, governo e altre organizzazioni pubbliche o private condotta via Internet. I beni o servizi sono ordinati via Internet ma il pagamento e l'ultima spedizione possono essere svolti sia on line sia off line.

Entrambe le definizioni prendono in considerazione le transazioni tra due o più parti e richiedono l'accordo delle parti sulla vendita o l'acquisto dei beni o dei servizi. Tutti gli altri processi che si svolgono elettronicamente sono stati esclusi. In entrambe le definizioni, la transazione è condotta ma, non necessariamente, completata sulla rete. Ciò significa che gli ordini sono inclusi nella definizione indipendentemente da come si svolgerà il pagamento.

## **2) Processi**

Il secondo aspetto della definizione è legato alle attività o ai processi di business inclusi o esclusi dal dominio del commercio elettronico. Mentre molti stati vogliono restringere la definizione delle attività di vendita e acquisto comprese nella definizione molti altri vogliono includere altri tipi di processi di business, come il marketing e la pubblicità. Indubbiamente, questi ultimi diventeranno molto importanti in futuro, a mano a mano



che il business cresce con sistemi di processamento elettronici sempre più integrati per una vasta gamma di funzioni aziendali e la produttività beneficia del commercio elettronico che diventa più chiaro. Ciò ha portato alla conclusione che ci dovrebbe essere una definizione specifica, legata alle attività di vendita e acquisto, e un'altra definizione che comprende un numero maggiore di processi di business.

### **3) Attori**

Il terzo aspetto della definizione è legato agli attori coinvolti nei processi di commercio elettronico. Abitualmente si divide il commercio elettronico in tre tipologie: Business to Business (B2B), Business to Consumer (B2C) e Business to Government (B2G). Ognuno di questi ha un set di indicatori relativi all'e-commerce. La maggior parte dell'interesse della letteratura si è focalizzato sul B2B e sul B2C, la maggior parte degli indicatori statistici riguarda queste due forme. È generalmente concordato che l'80% del valore totale del commercio elettronico nel mondo è, attualmente, totalizzato dal B2B. Il B2B fornisce i maggiori benefici in termini di produttività. D'altra parte il B2C ha la potenzialità di cambiare sostanzialmente il modo in cui le persone vivono e interagiscono.

## **1.2 Quali sono le differenze tra l'e-commerce B2B e B2C?**

C'è la differenza ovvia data dal fatto che i clienti in un caso sono le aziende e nell'altro i consumatori. Oltre a questa, ci sono due grandi distinzioni:

*Le attività di B2B richiedono negoziazione...*

Vendere ad un'altra azienda significa discutere sui prezzi, le spedizioni e le specifiche di prodotto. Non succede la stessa cosa per la maggior parte delle vendite ai consumatori. Ciò rende più facile l'introduzione di un catalogo on-line per i commercianti al dettaglio, le prime applicazioni di B2B erano, infatti, per i compratori di prodotti finiti o commodities che sono semplici da descrivere e da prezzare.

*...e integrazione.*

I commercianti al dettaglio non si devono integrare con in sistema dei loro consumatori. La maggior parte delle compagnie che vendono alle aziende devono integrare, poiché i loro sistemi devono essere in grado di comunicare con quelli dei clienti senza l'intervento umano.

Nella presente tesi si prenderanno in considerazione le problematiche relative al B2B e-commerce ed in particolare ai marketplace, che rappresentano una realtà molto importante che in questi anni sta vivendo una rapida evoluzione.

### **1.3 B2B e-commerce**

Il Business to Business e-commerce non rappresenta una rivoluzione ma un'evoluzione [Hogan, 2001]. Non cambierà il modo in cui un fornitore svolge le sue attività attualmente, ma arricchirà le attività attuali.

Il B2B e-commerce riguarda principalmente i canali. Gli e-marketplace e, dal punto di vista dei compratori, l'e-procurement<sup>1</sup> sono semplicemente nuovi canali per raggiungere i commercianti. Tradizionalmente le compagnie hanno raggiunto i loro venditori attraverso vendite dirette, canali di vendita indiretti, telemarketing, posta o un'ampia gamma di mezzi pubblicitari. Da quando Internet ha raggiunto il mondo degli affari si è riconosciuto il vantaggio apportato dal disseminare le informazioni su questa rete universale. Ciò ha spinto molte aziende alla creazione di un sito aziendale. Molte compagnie hanno riconosciuto che il passo logico successivo era estendere il loro sito per permettere ai clienti di fare acquisti direttamente via Web. Queste compagnie hanno acquistato sell-side solution, che danno la possibilità di svolgere transazioni di vendita basate sul Web. Quest'approccio è perfetto per il Business to Consumer ma non per il Business to Business. Nel mercato del B2B, gli acquirenti richiedono molta più attenzione. I fornitori devono fare in modo che la loro sell-side solution completi i canali già esistenti, invece di creare un conflitto di canali. Ciò ha portato alla nascita della seconda generazione di sell-side solution.

Ora si sta affrontando una nuova ondata di cambiamenti dovuta all'ubiquità della rete Internet. Molte compagnie, per aver maggior padronanza dei propri acquisti, hanno

---

<sup>1</sup> eProcurement: Programmi di gestione dell'approvvigionamento realizzati per funzionare tramite Web.

scelto di sviluppare un sistema interno di approvvigionamento come Ariba<sup>2</sup> ORMS e Commerce One Buysite<sup>3</sup>, o hanno scelto di far leva su e-marketplace come eSteel<sup>4</sup>.

Questo è un'importante via d'accesso verso i compratori e presenta i venditori con un grande numero di nuovi canali di supporto.

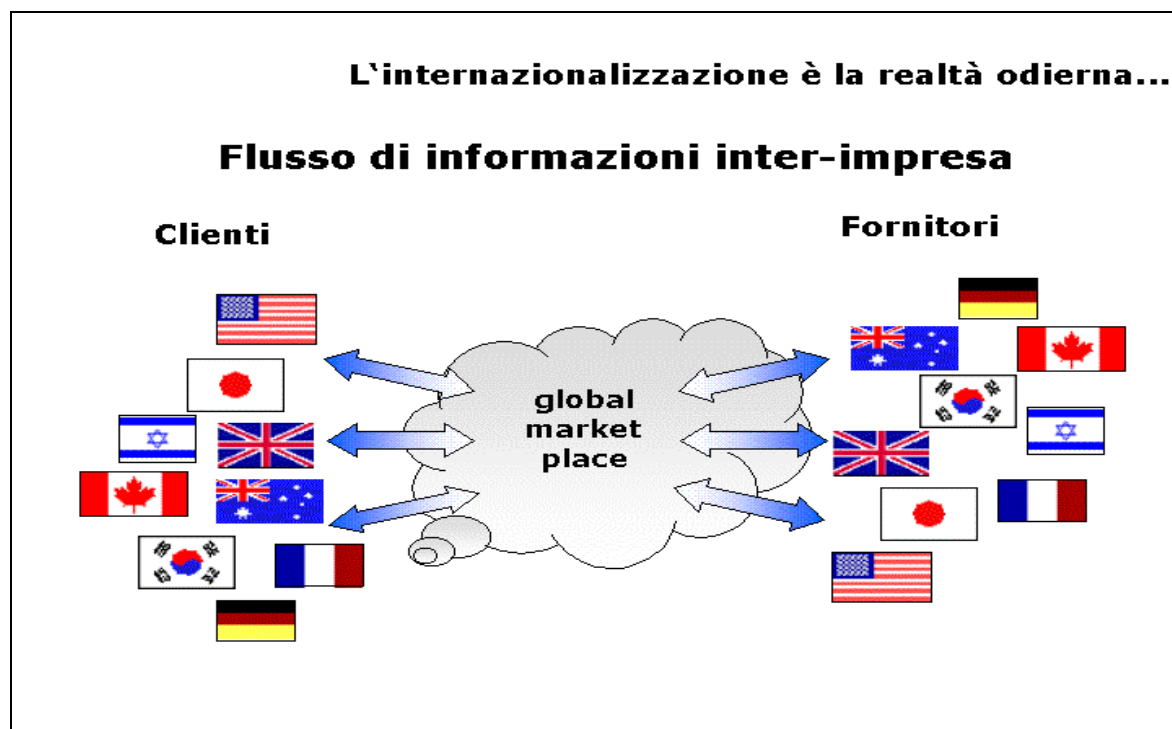


Figura 1: Il flusso delle informazioni

## 1.4 eMarketplace

Una delle attività più promettenti tra quelle emergenti dell'eCommerce sono gli e-marketplace [Acom, 2002]. La complessità dei B2B marketplace svanisce nel momento in cui si capisce che replicano semplicemente attività che sono esistite per lungo periodo - posti dove i beni possono essere comprati e venduti- virtualmente, on line. Nel fornire un sito virtuale dove possono essere svolte le attività di business, gli operatori dei marketplace danno la possibilità ai partecipanti di analizzare una vasta gamma di prodotti e alla fine di procurarsi velocemente i prodotti e i servizi, con meno spese e

<sup>2</sup> [www.ariba.com](http://www.ariba.com) : offre soluzioni per aiutare le compagnie nella gestione delle spese.

<sup>3</sup> [www.commerceone.com](http://www.commerceone.com) : fornisce soluzioni in grado di connettere compratori e venditori e di ottimizzare le loro interazioni.

<sup>4</sup> [www.esteel.com](http://www.esteel.com) : fornisce soluzioni per la gestione di interazioni di business.

fatica di quelle richieste dal commercio tradizionale. I venditori godono del rovescio della medaglia: possono presentare una vasta gamma di prodotti, ridurre le spese di vendita e, se vogliono, competere con prezzi e offerte speciali in real time. Si deve capire che i marketplace non sono identici, ma ricadono in un piccolo numero di categorie. I marketplace orizzontali sono abitualmente indipendenti dalle imprese. I mercati verticali, d'altra parte, si focalizzano sulle singole industrie e sulle loro supply chain. I marketplace possono essere distinti in privati e pubblici, nel primo caso sono riservati ad un ristretto numero di membri, nel secondo sono aperti a tutti i commercianti. Ciò porta ad un'altra distinzione relativa all'ampiezza del servizio. Nei casi più semplici, i mercati offrono solo i servizi base: incontro compratore venditore e gestione dei contenuti. Sviluppandosi su questi servizi di base, alcuni mercati hanno iniziato ad offrire servizi di asta e di vendita a contratto. Altri offrono servizi di vendita completi che includono servizi legali, di banking e di assicurazione.

### **1.4.1 Cosa fanno gli e-marketplace**

Gli e-marketplace sono, fondamentalmente, piazze virtuali che permettono ai venditori di presentare le proprie offerte e danno ai compratori la possibilità di acquistare i prodotti disponibili, ordinare ciò di cui hanno bisogno e, in alcuni casi, svolgere il pagamento in maniera sicura. Potenzialmente offrono alle compagnie un modo conveniente di commerciare con partner di business esistenti e di trovarne dei nuovi.

#### ***Servizi e applicazioni software dal punto di vista del compratore:***

- servizi per il processo di acquisto;
- sicurezza (pagamenti on line, operazioni finanziarie);
- servizi relativi alla supply chain (pianificazione della domanda, pianificazione delle risorse );
- servizi di trasporto;
- applicazioni verticali;
- servizi d'asta.

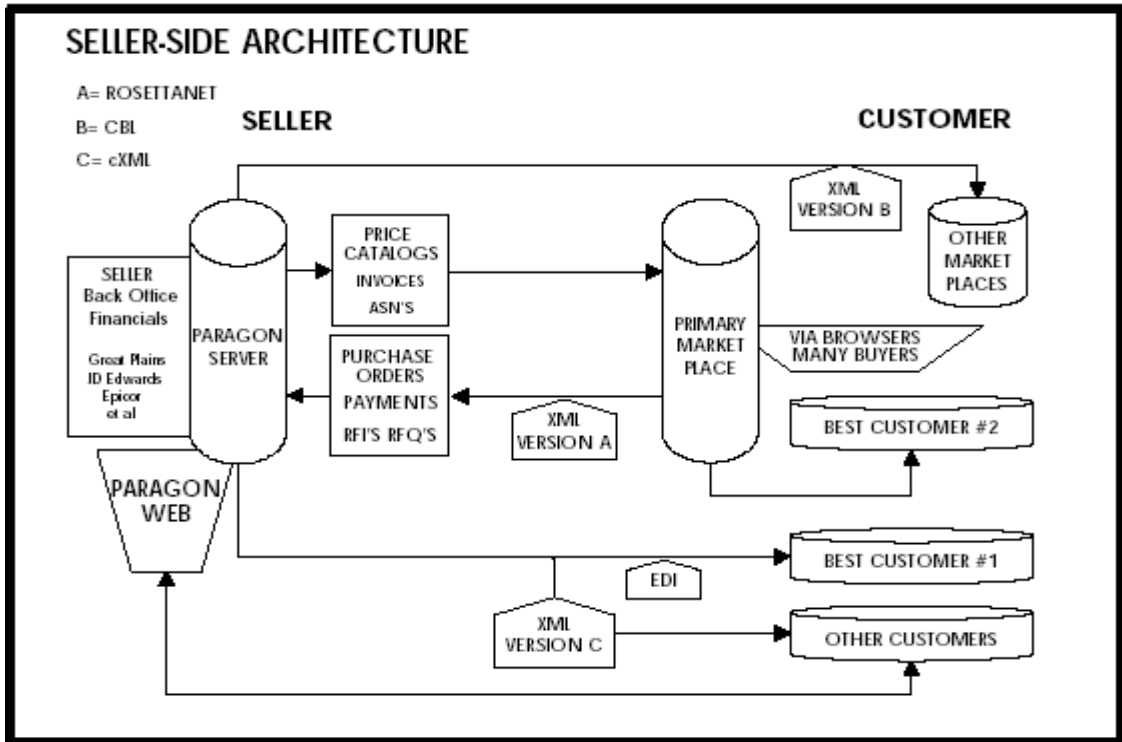


Figura 2: Architettura del marketplace dal punto di vista del compratore

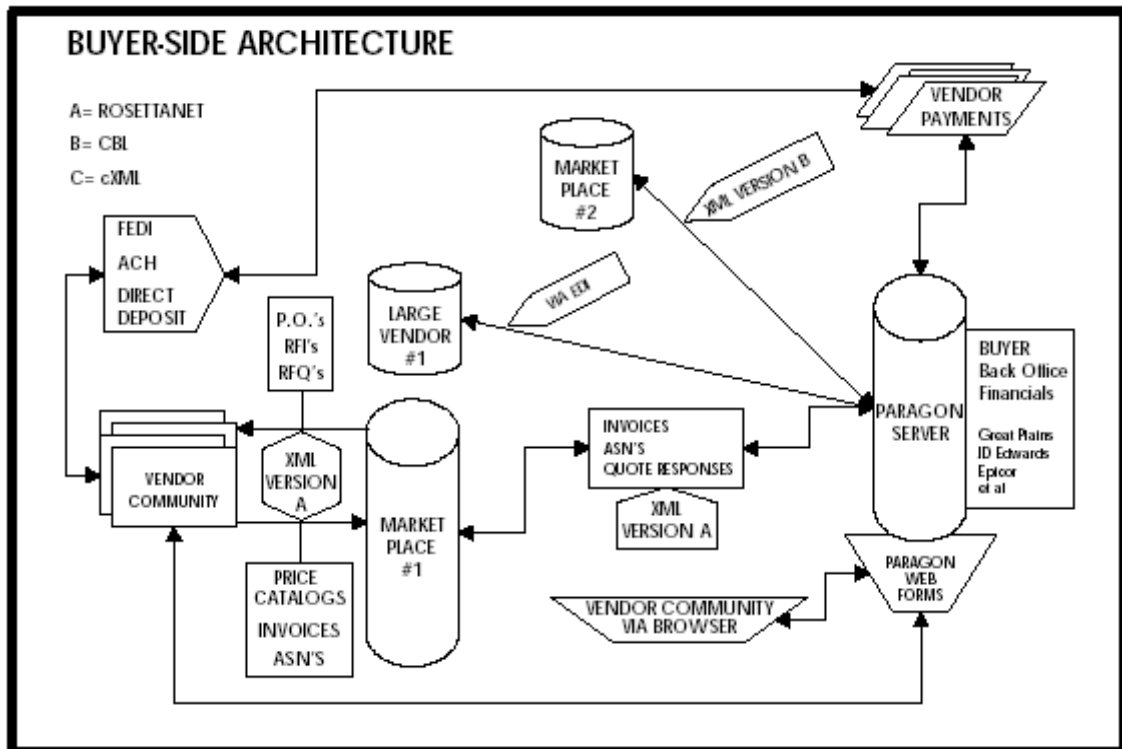


Figura 3: Architettura del marketplace dal punto di vista del commerciante

***Servizi e applicazioni software dal punto di vista del commerciante:***

- interfaccia Web;
- presentazione efficiente del catalogo;
- accesso a nuovi clienti;
- facilitazione nella gestione dell'ASP<sup>5</sup>.

## **1.4.2 Gli obiettivi e i rischi degli e-marketplace**

Col passare del tempo, sempre più business trasferiranno un numero crescente di attività on line, ci si deve aspettare che gli e-marketplace godranno di una percentuale di queste nuove attività. Nel momento in cui le compagnie decidono di utilizzare o costruire un e-marketplace, devono ricordare alcuni fattori fondamentali. Come sottolineato precedentemente, il marketplace replica un tipo di business convenzionale, ma non è una duplicazione precisa. Mentre apporta velocità e vantaggi economici ai processi di compravendita, ci sono elementi del business tradizionale che non sono replicabili quando non c'è la presenza umana. Si deve ricordare:

- I marketplace non danno la possibilità al fornitore di aggiungere valore al processo a causa della mancanza di interazione con il cliente.
- Le informazioni sui prezzi e sulla disponibilità dei beni possono essere visualizzate anche dai *competitor*.

A questo punto, ci si chiede:

- Come cambieranno le relazioni con i partner già esistenti partecipando ad un e-marketplace?
- Chi gestisce il marketplace? Quali organizzazioni si adattano meglio agli obiettivi della compagnia?

---

<sup>5</sup> **ASP:** Application Service Provider è 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.

- Un consorzio industriale.
- Una vasta e indipendente compagnia.
- Una compagnia indipendente priva di affiliazioni industriali.

### **1.4.3 I benefici del B2B e-marketplace per le imprese**

È largamente riconosciuto che l'e-commerce riduce i costi di transazione, aumenta l'efficienza, e genera importanti cambiamenti nella gestione dei business e dei processi produttivi [UNCTAD, 2001]. Per esempio, linkando le aziende e i consumatori attraverso Internet, il B2B e-commerce ha il potenziale di ridurre significativamente i costi delle transazioni. Esso, inoltre, accresce l'accesso alle informazioni da parte dei consumatori, riducendo i costi di ricerca e permettendo ai consumatori di trovare prezzi inferiori per un dato prodotto o servizio. Un gran numero di fornitori aumenta il livello di competitività e riduce i profitti monopolistici delle aziende.

Nel settore del B2B, l'e-commerce contribuisce alla riduzione dei costi collegando le industrie con i fornitori elettronicamente attraverso la supply chain. Riduce i costi di approvvigionamento perché rende possibile il reperimento di fornitori con prezzi inferiori. Aumenta l'efficienza perché la competizione tra i fornitori riduce i profitti monopolistici e il numero d'intermediari. Inoltre, riduce i costi legati alla fornitura di servizi finanziari o d'altri servizi che possono essere resi disponibili elettronicamente attraverso Internet. Un migliore flusso d'informazioni ha anche il vantaggio di ridurre gli stock di magazzino.

Garciano e Kaplan classificano i costi di transazione come costi di coordinazione e di motivazione, e sostengono che il B2B e-commerce abbia il potenziale di toccare entrambi i tipi di costi. I costi di coordinazione sono collegati con la determinazione dei prezzi e dei dettagli della transazione, alla mutua conoscenza di potenziali compratori e venditori, e per portarli a sviluppare la transazione. Il B2B riduce questo tipo di costi aumentando l'efficienza del processo di business, per esempio, quando una transazione, che abitualmente è condotta via telefono o via fax, viene fatta via Internet, vengono ridisegnati i processi di business. I costi di coordinazione sono ridotti quando il B2B e-commerce migliora l'accesso alle informazioni dirette, per esempio riducendo i costi di ricerca dei fornitori e permettendo loro di raggiungere un maggior numero di compratori

a basso costo. Il B2B e-commerce riduce i costi di coordinamento fornendo migliori informazioni sulla disponibilità, sulle caratteristiche e sul prezzo dei beni.

I costi di motivazione sono legati all'incompletezza delle informazioni e agli accordi approssimativi. I costi relativi all'incompletezza delle informazioni si presentano quando i compratori e i fornitori non hanno tutte le informazioni rilevanti, necessarie per capire se un prodotto è conforme alle specifiche. Il B2B e-commerce riduce questi costi attraverso la standardizzazione dei prodotti. I costi relativi ad accordi non chiari insorgono quando compratori e venditori non sono abili nel trattare. Questi costi vengono ridotti attraverso la standardizzazione dei processi.

#### **1.4.4 Il successo e il fallimento di un e-market**

Ci sono un gran numero di possibilità e funzionalità che gli e-marketplace possono offrire ai commercianti. Queste sono legate alla mole d'informazioni che forniscono, al supporto per la collaborazione tra trading partner e alla gestione di transazioni on line, includendo i pagamenti gli scambi on line di documenti e la fornitura di legami con servizi logistici. Gli e-marketplace che riusciranno a fornire maggiori funzionalità ne trarranno vantaggio. Insieme alle funzionalità è importante il raggiungimento di una massa critica di liquidità che rende gli e-market sostenibili. La tecnologia è un fattore critico che accresce le funzionalità del mercato. Per questo motivo molti e-marketplace hanno fondato alleanze strategiche con compagnie nel settore tecnologico o hanno affidato loro in outsourcing la costruzione dell'e-marketplace.

### **1.5 Traguardi dell'e-commerce**

L'e-commerce è una realtà nuova che sta vivendo un lento sviluppo, non privo di ostacoli. Le principali difficoltà sono da attribuire alla complessità della sostituzione delle interazioni umane con il computer. L'e-commerce è, infatti, privo della semantica, della sintassi e della grammatica che costituiscono la base del dialogo umano.



Nel mondo del commercio elettronico il confronto tra i vari prodotti è spesso ostacolato dalla mancanza di *standard* che li classifichino e li descrivano. Numerose proposte di standard di classificazione si sono tradotte nel fatto che ogni fornitore descrive i prodotti in un proprio modo [Bergamaschi, 2002]. Considerando i B2B marketplace, una transazione è resa ancora più difficile dalla necessità del compratore di usare un solo standard per classificare e descrivere prodotti forniti da vari venditori, in modo da assicurare una facile integrazione con il suo sistema ERP (Enterprise Resource Planning).

Il mercato deve fornire degli strumenti mediante i quali si possa mediare tra i differenti standard usati dai vari partecipanti alle transazioni. In questo modo, ogni attore del processo di business potrebbe scambiare informazioni usando il proprio formato di dati. Dunque, la necessità, per i mercati B2C e B2B, è riclassificare prodotti e servizi in accordo con differenti modelli di standardizzazione. Questa richiesta è così importante che UE Thematic Network on Ontology Based Information Exchange for Knowledge Management and Electronic Commerce [Schulten, 2001] ha proposto come *sfida* per tutto il mondo della ricerca e dell'impresa l'ottenimento di questa riclassificazione in modo semiautomatico.

Gli aspetti relativi all'integrazione sono molto critici e richiedono molta attenzione. L'integrazione del marketplace con le applicazioni di back end<sup>6</sup> dell'azienda sta diventando uno dei criteri più importanti nella valutazione del processo. I fornitori hanno bisogno che le transazioni interagiscano con i loro ordini, con gli inventari e con i cataloghi. I compratori vogliono transazioni che possano immettere dati direttamente nel loro sistema di contabilità. Le fatture devono poter passare direttamente al sistema di contabilità aziendale. Molti fornitori hanno sistemi IT (Information Technology) che sono progettati strettamente per uso interno. Sono pieni di acronimi e di abbreviazioni che hanno un senso solo per il personale dell'azienda. Internet, le sell-side solution, i sistemi di e-procurement e i marketplace, hanno creato una domanda di informazioni che siano utilizzabili da un pubblico generico.

Risolvere questo problema relativo alle informazioni è uno delle maggiori sfide che le aziende stanno affrontando in questo momento. Devono immaginarsi il modo degli altri

---

<sup>6</sup> Applicazioni di back-end e front-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 programme). 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.

di impadronirsi di informazioni in loro possesso, informazioni riguardo prodotti, clienti e prezzi, ed esporle selettivamente ai compratori attraverso vari canali elettronici. Necessitano di soluzioni in grado di trasformare i loro dati interni in informazioni elettroniche e in cataloghi che possano essere utilizzati da siti Web, dalle sell-side solution, dai sistemi di e-procurement e dagli e-marketplace.

## **Capitolo 2**

# **Standard di classificazione dei prodotti**

### **2.1 Generale**

La codifica di prodotti e servizi, in accordo con una convenzione di classificazione standardizzata, è necessaria per far fluire il commercio tra le compagnie. Prodotti e servizi che sono identificati in modo non ambiguo permettono la gestione degli acquisti e l'analisi della spesa. Inoltre, nomi di prodotti comprensibili alla macchina assistono le funzioni di marketing e vendita nella ricerca di clienti e nel fornire migliori servizi di distribuzione e ai clienti.

Inserendo i codici in vari documenti elettronici di commercio come i cataloghi, siti Web, ordini d'acquisto, fatture, documenti d'inventario e di vendita, le applicazioni informatiche, lungo un'estesa supply chain, possono processare le transazioni di dati automaticamente, possono svolgere analisi manageriali e prendere decisioni in tempi molto brevi e in un modo efficiente che non sarebbe possibile senza i codici.

Uno schema di classificazione dei prodotti utile deve essere gerarchico, in modo che i singoli beni rappresentino un'unica istanza di una più ampia classe o famiglia. Un'organizzazione gerarchica permette ad una compagnia di focalizzarsi al livello di specificità che si adatta meglio alle sue intenzioni e alla situazione. Attualmente la

maggior parte degli schemi di codifica dei prodotti non sono gerarchici, un esempio sono i codici a barre UCC ed EAN<sup>7</sup>, e, di conseguenza, non sono utilizzabili per fare ricerche ed analisi. Questi codici non possono essere riuniti in più generali categorie di codici individuali e non hanno relazioni con altri codici.

Per mantenere una tassonomia gerarchica, uno schema di classificazione deve essere costantemente aggiornato con l'inserimento di nuovi prodotti e l'adattamento delle strutture esistenti in relazione ai cambiamenti del mercato, deve essere chiaro per l'azienda che l'assegnazione di codici a prodotti e servizi deve essere imparziale.

Sia che le aziende vogliano identificare i loro prodotti usando un sistema di codifica sia che non lo vogliano, questo è argomento di discussione. La rapida crescita del commercio elettronico, particolarmente su Internet, e lo sviluppo di nuovi strumenti tecnologici (come XML, le search engine e i datamining tools<sup>8</sup>) rendono inevitabile l'adozione della codifica dei prodotti [Granada Research, 2002].

## **2.2 Dalle transazioni elettroniche alle decisioni strategiche**

Internet, EDI<sup>9</sup> e gli standard d'identificazione dei prodotti (includendo anche i codici a barre) hanno portato ad una profonda automazione delle transazioni. Fino ad ora 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. Informazioni descrittive relative ai prodotti sono importanti per il marketing automatico, per la presentazione e l'analisi degli acquisti di prodotti, attività che stanno diventando pratica comune per le funzioni di compravendita delle compagnie.

---

<sup>7</sup> UCC: Uniform Commercial Code.  
EAN: European Article Numbering.

<sup>8</sup> Data Mining tools : strumenti per analizzare ingenti moli di dati allo scopo di individuare particolari tendenze.

<sup>9</sup> EDI: Electronic Data Interchange.

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<sup>10</sup> 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 supply chain.

- **Reparto 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. Purchasing group hanno 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 consapevolezza dei loro volumi di spesa per gli approvvigionamenti e beni non di produzione. Gli standard di classificazione dei prodotti permettono ai purchasing group di identificare e selezionare fornitori strategici e, utilizzando search engine 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 inter-company) a vendere i loro prodotti. La disseminazione automatica d’informazioni sui prodotti raggiunge un gran numero di potenziali 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, e permettono la raccolta di informazioni standardizzate relative al market share.

---

<sup>10</sup> Tag: letteralmente “etichetta”, nel contesto si fa riferimento a tag utilizzati per evidenziare, in un format gerarchico, i differenti componenti del documento.

- **Distributori, rivenditori e altri partner di canale.** Chi fornisce servizi di assistenza tecnica pre e post vendita 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 come HTML<sup>11</sup> e TCP/IP<sup>12</sup> hanno portato verso il grande successo di Internet, gli standard di classificazione dei prodotti porteranno ad un altro livello il commercio elettronico.

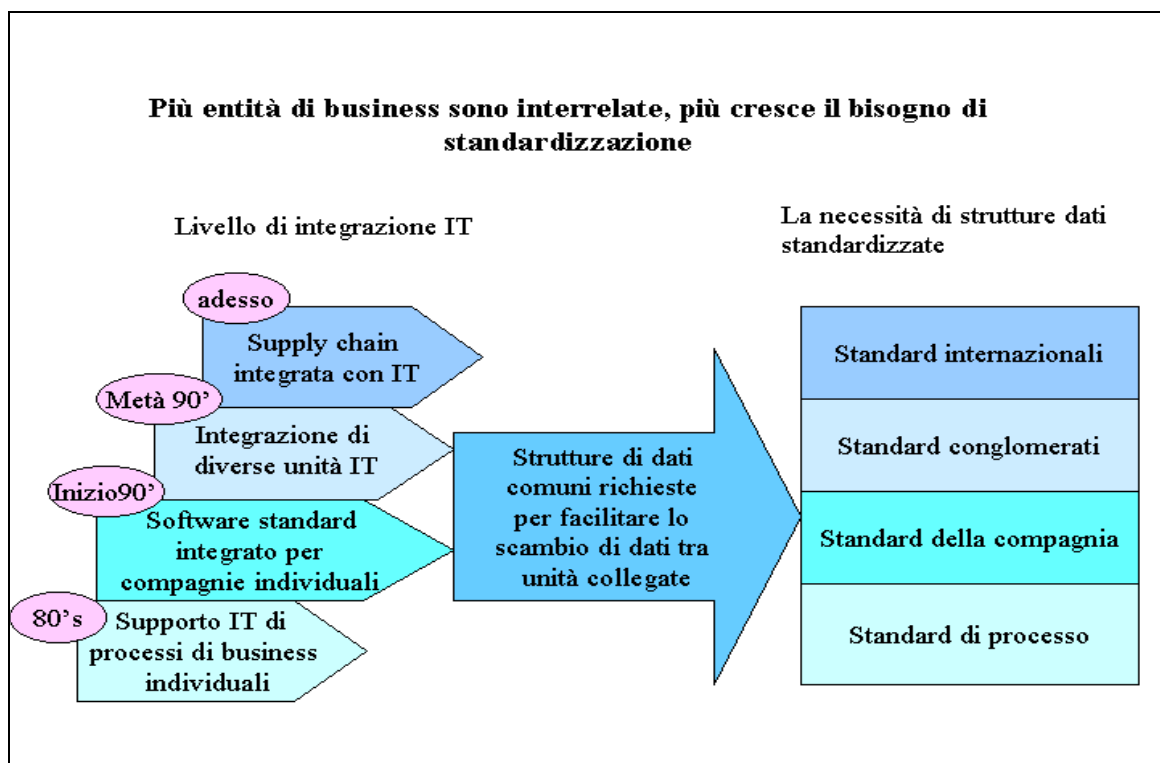


Figura 4: Necessità di integrazione e di standardizzazione

<sup>11</sup> HTML: Hypertext Markup Language 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

<sup>12</sup> TCP/IP: Transmission Control Protocol/Internet Protocol 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.

## 2.3 Perché classificare prodotti e servizi

Classificare prodotti e servizi, con uno schema di codifica comune, facilita il commercio tra venditori e acquirenti e sta diventando obbligatorio nella nuova era del commercio elettronico. Grandi compagnie stanno iniziando a codificare gli acquisti negli ordini per analizzare le loro spese. Una recente ricerca ha mostrato che oltre la metà delle compagnie analizzate utilizza codici per le commodity<sup>13</sup>, alcune hanno codici complessi di dieci o più categorie. La maggior parte delle compagnie con programmi procurement card hanno intrapreso l'onere della codifica.

La maggior parte dei sistemi di codifica per le aziende sono stati molto costosi da sviluppare. Tipicamente occorre un anno a svilupparlo, per ogni item codificato s'impiega circa un'ora e mezza per l'assegnazione di un codice. I fornitori di una compagnia abitualmente non aderiscono agli schemi di codifica dei loro clienti.

C'è una gran duplicazione di spese e di lavoro nel creare codici. Se ci fosse stata una singola universale convenzione di codifica, anche nel caso in cui le imprese avessero desiderato adeguarla alle loro esigenze, ci sarebbe stato un grande risparmio.

### 2.3.1 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.

<b>Pro</b>	<b>Contro</b>
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 strategic sourcing.	
Un sistema di codifica consente attraverso le divisioni della compagnia e i sistemi informativi	

<sup>13</sup> Commodity: per commodity si intendono beni indifferenziati.

di dare un'immagine uniforme delle spese dell'azienda.	
Un unico sistema di numerazione integra l'intero flusso del processo.	
Facilita il controllo e l'adesione ai 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: Pro e contro della classificazione

**L'incontro tra la domanda e l'offerta nell'e-commerce necessita di un linguaggio comune**

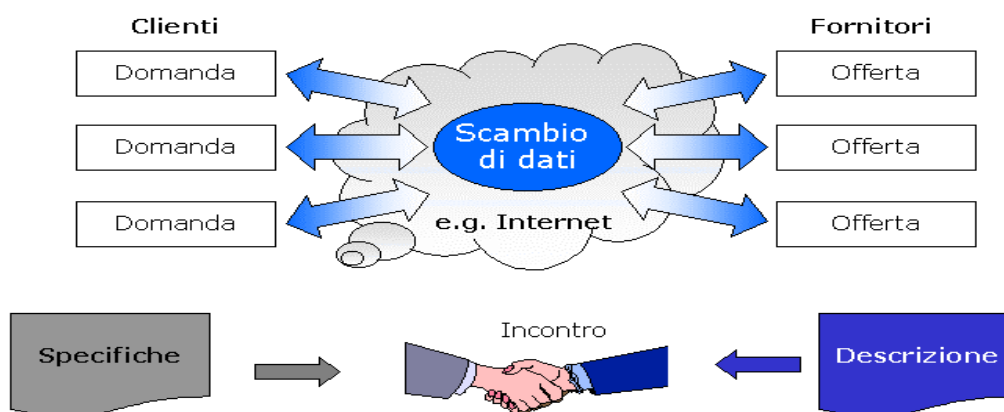


Figura 5: Incontro tra domanda e offerta

**Facilita l'individuazione di prodotti.** Una convenzione di denominazione comune permette ai sistemi informatici 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 sui prodotti dotate di pretagging con un vocabolario di termini pre-settato che è stato sviluppato da partecipanti dell'industria.

**Facilita l'analisi delle spese.** Quando ogni transazione d'acquisto di un'impresa ha il tagging con un set comune di 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 ai purchasing group di:

- Consolidare i fornitori 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 di accordi per l'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.

**Facilita il controllo e l'uniformità attraverso la compagnia.** I codici portano una singola, uniforme visione di tutte le spese di una compagnia. Riuniscono insieme tutti i dipartimenti e le divisioni, includendo le funzioni di business come gli acquisti e i pagamenti. I codici, essendo usati in sistemi informativi, sono strumenti essenziali per un controllo rapido della lista degli item e dei venditori autorizzati, dei flussi di lavoro approvati e del denaro allocato per le spese.

## 2.3.2 Per il marketing e la distribuzione

Quando una compagnia di vendita applica un tag standard ai suoi prodotti e servizi, tale tag può essere automaticamente processato da altre applicazioni informatiche come le Web search engine<sup>14</sup>, sistemi ERP e i *purchasing software at customer site*, software di aggregazione dei cataloghi e agenti software. Con la codifica, il produttore può rapidamente propagare nel mondo la sua offerta commerciale.

<b>Pro</b>	<b>Contro</b>
I propri prodotti vengono automaticamente inseriti nelle Web search e negli aggregatori di cataloghi.	Classificare i prodotti in maniera uniforme con i competitor 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 produttività delle comunicazioni con il canale di distribuzione.	
Abilita al benchmark il marketshare	
Acquisisce microdati chiari e consistenti dal canale di vendita	
Sta 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**

<sup>14</sup> Search engine: Motori di ricerca.

**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 nelle search engine e nei marketplace, nelle directory di prodotto e nei giornali elettronici di commercio;
- Pubblicizzare i propri prodotti e la propria compagnia utilizzando tag XML<sup>15</sup> come metadati;
- Preparare il proprio sito Web a sviluppi futuri, quali le 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 codificate sui prodotti aiutano a coordinare un canale di distribuzione. Quando un venditore vende i suoi prodotti attraverso terze parti come i distributori, rappresentanti indipendenti, commercianti, rivenditori e 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 sta parlando degli 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.

---

<sup>15</sup> XML: Extensible Markup Language (vd. Capitolo 6)

**I microdati del canale rendono possibili analisi migliori.** La classificazione delle transazioni con il canale 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.

**Rende i beni indifferenziati?** Descrivere i prodotti della propria compagnia con lo stesso codice utilizzato dai concorrenti sembra equivalente a ridurre tali prodotti a beni indifferenziati. 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 della scaffalatura, è una scaffalatura virtuale per il commercio elettronico. Come nel mondo fisico i marchi giocano un ruolo di potere, così giocheranno un ruolo differenziatore anche nel mondo non fisico, insieme con altri aspetti come il servizio ai clienti, il packaging e lo sviluppo di nuovi prodotti.

## 2.4 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.

<b>Codice</b>	<b>Cosa identificata</b>
Numero di previdenza sociale	Persona
UCC\EAN(codici a barre)	Prodotti
Duns number	Compagnia

**Tabella 3: Esempi di identificazione**

**I codici di identificazione.** I codici di identificazione sono usati per rendere non ambigua l'identificazione di un oggetto. La corrispondenza uno a uno tra i codici e i beni è molto utile per registrare e linkare records di item e azioni fatte sull'item.

**I codici di classificazione** sono usati per raggruppare oggetti 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 item, 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.

	<b>Codice di classificazione</b>	<b>Codice di identificazione</b>
Principio cognitivo	Indica la relazione tra item e altri item, simili e non	Identifica gli item senza ambiguità
Caratteristica chiave del codice	Gerarchico	Unicità
Cifre del codice	Mostra le classi e le sottoclassi con i loro membri	Crea corrispondenza uno a uno tra i simboli e gli item( 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 classificazione ed identificazione**

**Qual è la differenza?** Per illustrare la differenza tra i codici di identificazione i codici di classificazione, si consideri un codice a barre che identifica un Dell laptop computer. Questo codice identificativo non può essere riunito in una categoria più generica: come accessori per il computer, non ha legami logici con i codici che identificano le varie marche di laptop per il computer. I codici identificativi non permettono la categorizzazione degli acquisti in categorie più generiche e non permettono il confronto tra diversi produttori.

Solo una tassonomia gerarchica può permettere l'aggregazione o la specificazione 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 item specifici e le classi più generiche.

## **2.5 Le caratteristiche di una convenzione per la denominazione dei beni**

La maggior parte dei sistemi commerciali di classificazione dei prodotti sono di uno di questi due tipi: 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 come input.

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

### **2.5.1 Tassonomia gerarchica per analisi “Drill down” e “Roll up”**

La classificazione gerarchica permette di scendere lungo la scala gerarchica attraverso una vasta gamma di prodotti fino ad arrivare esattamente a quello che si desidera. Permette anche ai manager di salire nella gerarchia, riunendo un gran numero di record di acquisto in un unico bacino, 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 compagnia.

Per esempio, analizzando le spese, un compratore di un'azienda manifatturiera troverà la categoria “forniture industriali” troppo generica. Vorrà spezzare in categorie più

specifiche la spesa in detersivi e solventi. Per un compratore, un magazine publisher o per una compagnia di servizio finanziario, la categoria “forniture industriali” sarà sufficientemente dettagliata, perché complessivamente comporta un piccolo ammontare di spesa. Non sono necessarie ulteriori suddivisioni.

D'altra parte, 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 compagnia e dalle sue intenzioni. Una gerarchia di classificazione assicura che una compagnia possa trovare un significativo livello di analisi che le permetterà di agire.

Dato che lo schema di classificazione è gerarchico, una compagnia può organizzare un intero gruppo di transazioni di acquisto per mezzo di categorie che le permettono di agire in modo strategico. Invece di concentrarsi sulla minuzia 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 e/o di fornitori per una famiglia di prodotti cui l'azienda acquirente può approcciarsi per trovare un singolo punto di approvvigionamento. Dato che con un solo contratto si può coprire la famiglia o il venditore, l'impresa può ottenere un trattamento preferenziale.

Generalmente, l'azienda vuole identificatori esattamente al livello di specificità che le permetterà di individuare un corso di azione 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 item.

## **2.5.2 Chiavi di identificazione**

In aggiunta alla gerarchia, ogni specifico nome della classificazione dovrebbe avere associato un numero unico. Ciò assicura che non ci sono 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. Ciò permette di tradurre il codice nelle tassonomie del cliente concepite per scopi personali.

Si considera, per esempio, in una multinazionale che vuole analizzare le spese per area e per divisione. Se usa nomi di prodotti indigeni di ogni stato, ci sarà 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 item si sono trasformati in nuovi item. In effetti, un numero unico permette riferimenti incrociati che assicurano la consistenza.

### **2.5.3 Caratteristiche dello schema**

Lo schema rende le classificazioni affidabili e coerenti con le necessità dell'utente, ciò grazie a delle accortezze che sono state adottate sia in fase di stesura che nelle varie fasi del mantenimento delle codifiche.

**Lo schema è consistente.** La consistenza significa che un singolo item viene identificato in un unico luogo. Significa che 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.

**Lo schema è completo.** Un buono schema identifica tutti i prodotti e i servizi in vendita e mette tutti questi item in una gerarchia. Gli utilizzatori di una classificazione richiedono che tutte le categorie rilevanti e tutti i fornitori vengano rappresentati accuratamente dallo schema di classificazione. Questo permette al management di fare analisi e di ottimizzare.



**Lo schema è sensibile all'e-marketplace.** 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.

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 compagnia i codici 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.

## 2.6 [Ecl@ss](#)

[Ecl@ss](#) è uno degli standard di classificazione dei prodotti più diffusi in Europa, nelle pagine successive se ne prenderà in considerazione la storia e la struttura in vista di una sua più approfondita analisi nei prossimi capitoli.

### 2.6.1 La nascita

L'iniziativa di [ecl@ss](#) è nata nel 1997 come sottogruppo del gruppo di lavoro VCI (Material Management of Technical Goods), nell'ambito dell'industria chimica tedesca. Il sottogruppo fu chiamato Material-Merkmalleisten. Il gruppo 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 notò che creare set di standard d'attributi significa, implicitamente, creare classi di beni tecnici: tutti i beni che erano appropriatamente descritti da un set d'attributi formavano una classe. In ogni modo, si decise di non approfondire gli aspetti della classificazione. Furono coinvolti due imprenditori: uno di una società d'ingegneria chimica e uno di una società d'ingegnerizzazione di controllo del processo. Si era calcolato che il processo avrebbe richiesto due anni.

## **2.6.2 L'estensione degli obiettivi**

Nel 1998 gli obiettivi furono estesi. I reparti d'approvvigionamento delle compagnie dei membri del VCI s'interessarono 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 sono simili al caso di gestione dei materiali. Un catalogo elettronico è un'applicazione centrata attorno ad un'ampia lista d'elementi e alle loro descrizioni. I cataloghi elettronici usano uno schema di categorizzazione gerarchico o un modello noun-modifier 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 di attributi giocano un ruolo chiave. Mentre le applicazioni dei cataloghi elettronici e della gestione dei materiali sono molto simili all'inizio, la gamma dei prodotti nei cataloghi è 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.

I gruppi di materiali sono usati dal reparto d'approvvigionamento per modellare la struttura del mercato. La principale applicazione dei gruppi di materiali è l'analisi del volume d'approvvigionamento della compagnia in base al segmento di mercato. Per effettuare benchmark con imprese multinazionali è necessario avere un sistema di raggruppamento dei materiali accettato in campo internazionale.

Si decise di ampliare gli obiettivi del progetto in due direzioni. La prima direzione era il range dei materiali considerati, che fu esteso a tutta la gamma di materiali che un'industria chimica avrebbe acquistato. La seconda direzione è la reperibilità dei risultati del progetto. Si decise di costruire 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 d'anni internamente. Questo sistema fu scelto come punto di partenza per gli ulteriori progetti. Un altro imprenditore, di un'azienda di consulenza, fu coinvolto per supportare il gruppo nella creazione della gerarchia.

Nell'Agosto 1999 fu 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 sparsamente popolato. L'obiettivo di tale pubblicazione era iniziare un'ampia discussione per guadagnare visibilità e ottenere correzioni, ampliamenti da varie fonti. La versione 2.0 non fu considerata pronta all'uso.

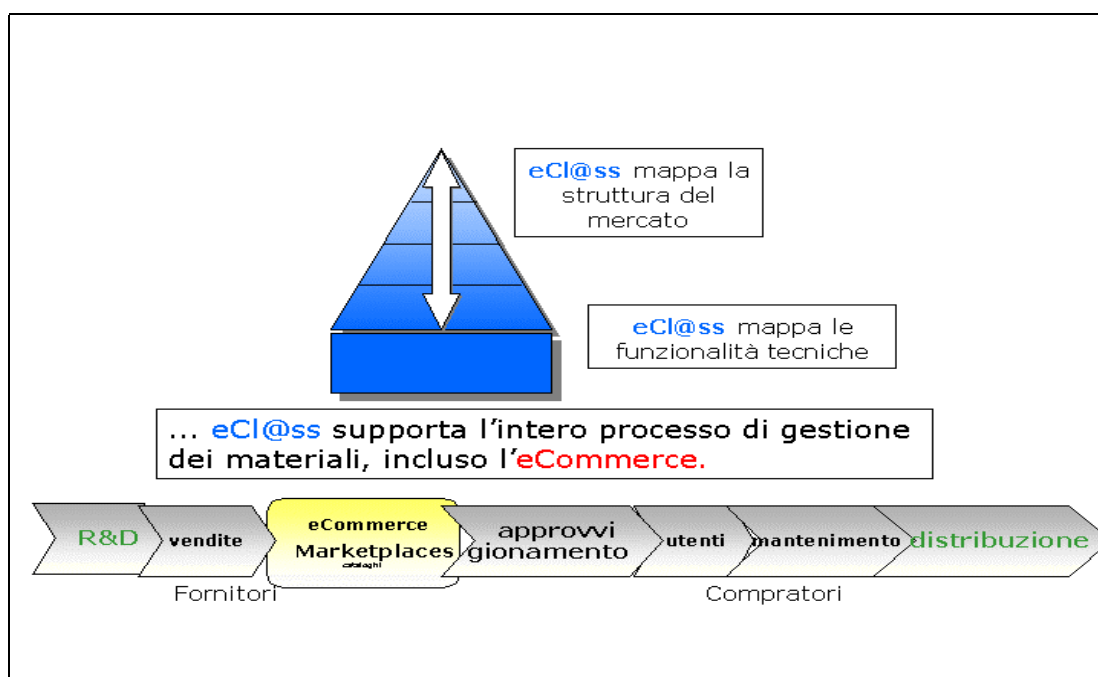


Figura 6: La funzione di [ecl@ss](#)

### 2.6.3 Verso uno standard globale

Ben presto si capì che si doveva andare oltre in entrambe le direzioni, sia quella geografica sia nell'estensione del codice a tutti i segmenti. Tutti i membri del gruppo di lavoro appartenevano a imprese multinazionali, sapevano fin dall'inizio che uno standard locale non poteva funzionare. Il gruppo di [ecl@ss](#) capì subito che i beni e i servizi che stavano descrivendo sarebbero stati venduti a tante altre industrie. L'industria chimica ha fornitori comuni con queste altre industrie, e questi fornitori non possono gestire numerosi standard specifici di segmento.

L'internazionalizzazione è, quindi, un'attività centrale. I primi passi sono stati fatti in America e in Asia. L'Unione Europea sta seguendo. La versione 2.0 è bilingue, l'originale è in tedesco ed è accompagnata da una traduzione in inglese. Si è anticipato che ciò potrà cambiare in futuro. Ci potrà essere bisogno di altre traduzioni.

Il gruppo di [ecl@ss](#) si apre ad altre industrie. Molti dei membri originali hanno attività in altri settori industriali. L'estensione ad altri segmenti è naturale ma non sufficiente.

### 2.6.4 Funzioni e applicazioni

[Ecl@ss](#) supporta e integra le seguenti funzioni e applicazioni:

- Ingegnerizzazione (sistema CAE<sup>16</sup>).
- Gestione dei magazzini.
- Gestione dello stabilimento.
- Approvvigionamento:
  - Richiesta di offerte;
  - Ordini;
  - Cataloghi elettronici;
  - Analisi.
- Vendite:
  - Informazioni sui prodotti;
  - Offerte;
  - Processamento degli ordini;
  - Cataloghi elettronici.
- Mercati virtuali.

---

<sup>16</sup> CAE: Computer Aided Engineering

## 2.6.5 La struttura

[Ecl@ss](#) si compone di quattro livelli di classificazione gerarchica dei materiali con un indice di parole chiave contenente 18000 termini. I livelli gerarchici sono: segmento, gruppo principale, gruppo e classe di beni. La classificazione è strutturata in 22 segmenti. Sono disponibili due posizioni per ognuno dei quattro livelli. Ne segue che 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. 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). [Ecl@ss](#) può essere applicato lungo l'intera supply chain: la struttura del mercato di approvvigionamento è mappata a 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 per supportare i tecnici 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

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

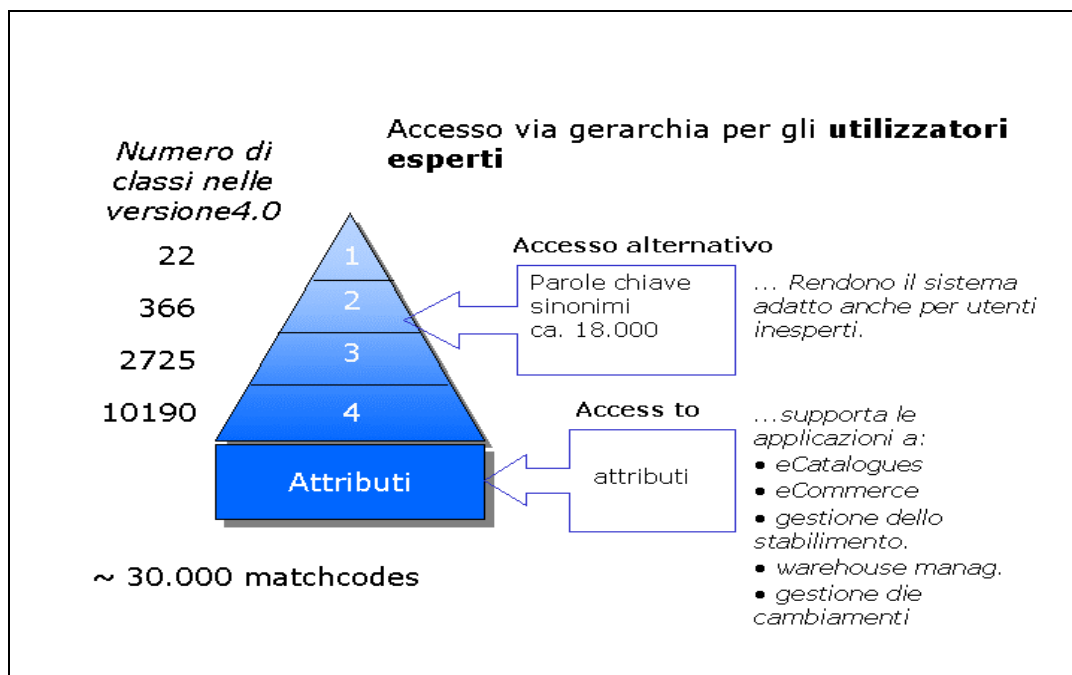


Figura 7: La struttura di ecl@ss

## 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 attaccati 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:

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

Tabella 5: La struttura di [ecl@ss](#)

Ogni classe di materiali è associata con i seguenti campi:

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

	<p>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.</p>
Nome identificativo	<p>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 <a href="#">ecl@ss</a> . Ciò implica che sono richiesti ulteriori qualificatori se un termine è usato con differenti significati in diversi contesti.</p>
Nome breve	<p>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 <a href="#">ecl@ss</a> .</p>
Definizione	<p>Una precisa descrizione testuale dei beni e dei servizi che appartengono alla corrispondente classe dei materiali.</p>
Versione di <a href="#">Ecl@ss</a>	<p>È la versione di <a href="#">ecl@ss</a> per cui questi valori e le caratteristiche dei vari campi sono valide. Le chiavi identificativi non saranno mai cambiate nelle successive versioni. Tutti gli altri campi possono cambiare. Le chiavi identificativi potrebbero scomparire nelle versioni successive se venisse rimossa la corrispondente chiave di identificazione dei materiali.</p>

**Tabella 6: Le classi di materiali**

## Cosa si intende per classificazione

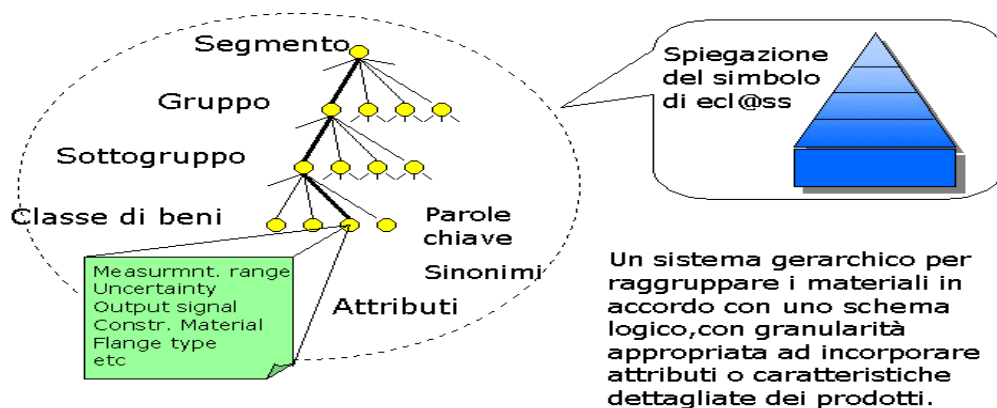


Figura 8: Significato della classificazione

## 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 diversi 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.

Appena un set di attributi è stato realizzato e controllato è disponibile nel sistema di classificazione di [ecl@ss](#).

Gli utilizzatori di [ecl@ss](#) 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 e si devono usare regole accettate internazionalmente. [Ecl@ss](#) è basato su questi standard internazionali e quindi fornisce una piattaforma standard. Le regole sono basate sulla



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 aggiuntiva può essere legata alle classi di materiali. Le parole chiave standard non hanno bisogno di essere definite.

Ogni parola chiave è associata ai seguenti campi caratteristici:

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 <a href="#">ecl@ss</a> .
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 <a href="#">ecl@ss</a>	È la versione di <a href="#">Ecl@ss</a> in cui questi valori dei campi caratteristici sono validi. Le chiavi identificative non saranno mai cambiate nelle successive versioni di <a href="#">ecl@ss</a> . Tutti gli altri campi possono cambiare. Le chiavi identificanti possono scomparire nelle nuove versioni se la parola chiave corrispondente è rimossa

**Tabella 7:Le parole chiave**

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.

### 2.6.6 Le possibilità di ricerca con [ecl@ss](#)

La ricerca di codici o definizioni all'interno di una codifica è un'attività molto frequente e, quindi, è indispensabile predisporre appositi sistemi che la rendano agevole e aderente alle necessità dell'utente. In [Ecl@ss](#) la ricerca può essere effettuata in vari modi :

- **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 si 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.
- **La ricerca gerarchica:** quando si conduce una ricerca gerarchica, inizialmente si visualizza il primo livello. Dopo il ventiduesimo livello 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.

## 2.6.7 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 compagnie 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 compagnie coinvolte nel processo di sviluppo del sistema e sono generalmente in grado di prendere decisioni in breve tempo. L'estensione o il nuovo numero viene intergrato 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, comunque 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. Sarebbe ripetitivo e i risultati sarebbero ripetuti durante il processo di ricerca.

La Figura 9 mostra come l'utente può influenzare [ecl@ss](#). In primo luogo lo può usare senza costi per svolgere le sue attività. Se ritiene può impostare query via Internet, può apportare modifiche o cambiamenti.

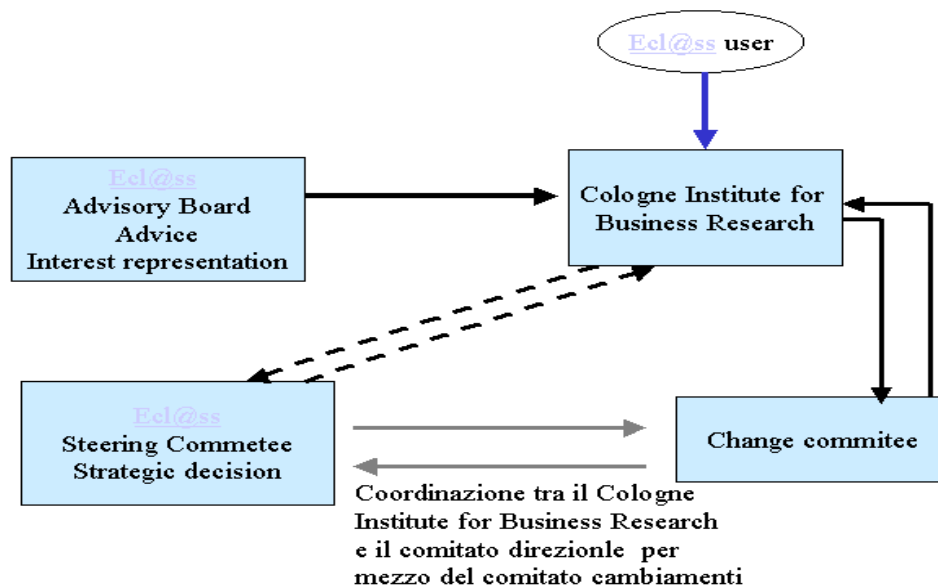


Figura 9: Procedura di aggiornamento

## 2.6.8 Prospettive future di [ecl@ss](#)

In cooperazione con varie compagnie, il Cologne Institute for Business Research continuerà a supportare [ecl@ss](#) e cercherà di introdurre una struttura di notazione per i codici di [ecl@ss](#) su Internet.

È stato pianificato che le compagnie che usano il sistema di qualità di [ecl@ss](#) e seguono le raccomandazioni del Cologne Institute for Business Research ricevano un sigillo di qualità che permette loro di distinguersi dai competitor. Si assume che la classificazione venga usata nelle compagnie e che la competenza in e-commerce sia implementata nella compagnia, in particolare in aziende di piccole o medie dimensioni. A questo fine, il Cologne Institute for Business Research preparerà la lista delle raccomandazioni per le aziende di piccole dimensioni su come devono essere utilizzati i codici di [ecl@ss](#).

## 2.6.9 Il mantenimento e la gestione della qualità

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 il loro 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 adottano [ecl@ss](#) lo porranno tra i loro dati principali. Il mantenimento dei dati principali nelle organizzazioni complesse richiede un'appropriata gestione del sistema qualità. In industrie regolate, come ad esempio l'industria farmaceutica, motivi legali richiedono la gestione della qualità. 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](#) .

## **2.7 UNSPSC (Universal Standard Products and Services Classification)**

### **2.7.1 Obiettivo**

UNSPSC è uno standard open source 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.

### **2.7.2 Ambito**

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 caratteristiche fisiche sono normalmente esclusi. UNSPSC sarà collegato a ECCMA Global Attribute Schema (EGAS).

Il framework di standardizzazione classifica gli attributi, in modo che possano essere facilmente utilizzati per descrivere meglio un servizio o un prodotto specifico offerti nel marketplace. UNSPSC schema comprende ECCMA Global Commodity Identifier (EGCI) e ECCMA Global Commodity Classification (EGCC).

### 2.7.3 Classificazione gerarchica

Il primo obiettivo della classificazione gerarchica UNSPSC (EGCC) 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. EGCC è una classificazione gerarchica a 4 livelli; la rappresentazione raccomandata del EGCC prevede valori numerici a 2 digit separati in periodi: NN.NN.NN.NN.

<b>Livello</b>	<b>Definizione</b>
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 Accomodation furniture
Class	Gruppo di commodities che corrispondono a 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: Struttura di UNSPSC**

## 2.7.4 Commodity Identifier (EGCI)

Il primo scopo del ECCMA Global Commodity Identifier (EGCI) è quello di fornire alla tabella UNSPSC una sequenza di identificazione che permetta un controllo di versione e il collegamento tra tabelle incrociate. EGCI è collegato a titolo e definizione della classificazione; mentre EGCC può cambiare se una commodity viene riclassificata, 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:

<b>Titolo</b>	<b>Definizione</b>	<b>EGCC</b>	<b>EGCI</b>
Segment	Furniture and Furnishings	56.00.00.00	010745
Family	Accommodati on furniture	56.10.00.00	009227
Class	Office furniture	56.10.17.00	010200
Commodity	Desks	56.10.17.03	008288

**Tabella 9: Esempio della struttura di UNSPSC**

## 2.7.5 Classificazione VS identificazione

È molto importante distinguere la classificazione dall'identificazione dei prodotti, in quanto la diversità dello scopo alla base dello schema si riflette non solo in un diverso utilizzo ma in profonde differenze strutturali:

### **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.

## **2.7.6 Distribuzione dei files**

ECCMA mantiene e distribuisce tre file UNSPSC. Il primo è progettato come file pubblico ed è disponibile dal sito di accesso pubblico; 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 attraverso ECCMA Member Lunge. 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 il WRITE UNSPSC e l'Audit file per il READ UNSPSC. Questa raccomandazione assicura compatibilità a ritroso.

## 2.7.7 ECCMA Membership

UNSPSC è uno standard open source; è di pubblico dominio e non soggetto a copyright. UNSPSC può essere 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.

## 2.7.8 Gestione dei processi di 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:

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 distretto, magazzino.

Regola 2.Una commodity sarà classificata con altre prodotte 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

<b>CODE DEVELOPMENT PROJECT REQUEST</b>
---

***\*Code Group or Table to be developed:*** \_\_\_\_\_

<b>1) Submitted by:</b>		
<b>2) Date of Request:</b>		
3) Brief Business Case for Proposed Revision (why the code in its current form would not be sufficient for what you industry needs):		
<b>4) Participants (list anyone that has shown interest in this project):</b>		
Name	Organization	E-mail Address

<b>5) Are you aware of any additional Companies or Organizations that will be interested in or affected by this project work?</b>			
Company Name	URL	Contact Name	Contact E-mail Address

PP# _____  (Assigned by ECCMA Secretariat)
--

### 2.7.9 Servizi di implementazione

ECCMA non fornisce servizi di UNSPSC tagging; comunque ECCMA mantiene una lista sul sito 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'organizzazione confronti prezzi, servizi, qualità di codifica tra i venditori.

## **2.7.10 Electronic Commerce Code Management Association (ECCMA)**

ECCMA è un'organizzazione non-profit registrata in Delaware 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 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). I membri di ECCMA possono associarsi in uno o più code group.

### **2.7.10.1 Principi**

I codici gestiti da ECCMA devono essere disponibili gratuitamente al pubblico senza restrizioni di copyright.

I codici gestiti da ECCMA sono mantenuti e distribuiti esclusivamente su Internet.

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.

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.

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.

I principali doveri e responsabilità dei membri del comitato sono:

- o Approvare i membri del TAG.
- o Approvare i membri del LTG.
- o Assicurare il rispetto delle Code Management Procedures da parte dei membri.
- o 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.

Il TAG (Technical Assesment Group) 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.

LTG (Language Translation Group) di 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.

## **Amministrazione**

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.

## **Il processo ECCMA e il metodo Delphi**

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<sup>17</sup>.

---

<sup>17</sup> <http://www.iit.edu/~it/delphi.html>

<http://eies.njit.edu/~turoff/Papers/delphi3.html>

[http://pespmc1.vub.ac.be/asc/Delphi\\_metho.html](http://pespmc1.vub.ac.be/asc/Delphi_metho.html)

## **Precauzioni Antitrust**

ECCMA è una società. Molti tra i suoi membri sono tra loro concorrenti, la competizione è sia orizzontale sia verticale. Questo significa che ogni attività deve essere valutata rispetto alle leggi Antitrust. ECCMA si 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 si devono essere discussioni su prezzi, boicottaggi, allocazione di consumatori, ogni standard sviluppato deve essere uno standard volontario.

## **2.8 NAICS:dati nuovi per un'economia nuova.**

NAICS è il nuovo sistema del nord America per classificare i business ed elaborare statistiche industriali. Esso rimpiazza il sistema precedente e identifica molte nuove industrie e sarà aggiornato regolarmente. Cambierà le statistiche industriali degli Stati Uniti e il modo in cui sono usate.

### **2.8.1 Background**

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) NAICS. Le prime statistiche industriali basate su NAICS sono i dati schedulati dall'Economic Census nel 1997 per la pubblicazione del 1999.

NAICS è una revisione del sistema usato per classificare i business attraverso le industrie. Diversamente dalle prime revisioni di SIC, i cambiamenti apportati da NAICS sono fondamentali. Riconoscono centinaia di nuovi business nella nostra economia, in



primo luogo nei settori di servizio che sono in forte crescita. Le industrie sono classificate secondo un unico consistente principio economico piuttosto che con vari. Tali cambiamenti sono stati sviluppati congiuntamente dagli Stati Uniti, dal Canada e dal Messico, introducendo per la prima volta comparabilità nel Nord America.

NAICS è un prodotto di estesa, ufficiale e pubblica deliberazione. I piani, la concezione, le proposte di revisione e la proposta completa del sistema sono state dettagliate in numerose notifiche dei registri federali discusse 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 pubblicato a metà del 1998. La classificazione NAICS sarà aggiornata regolarmente per tenersi 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.

## **2.8.2 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.

Il nuovo catalogo industriale di NAICS mostra i vari modi in cui l'industria è cambiata. Alcune parti riconoscono gli sviluppi high tech come il *fiber optic cable manufacturing*, le telecomunicazioni via cellulare, e la riproduzione di software per il computer. Altre riconoscono nuovi business. Si sono fatti anche altri cambiamenti relativi al modo in cui il business è svolto, come servizi di bed and breakfast, warehouse clubs, telemarketing bureaus e l'emissione di carte di credito.

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

**Tabella 10: Le nuove aziende di NAICS**

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 precedenti 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.

### 2.8.3 Più settori economici

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 sole 10 divisioni in SIC. La Tabella 11 elenca ogni divisione di SIC e mostra i settori di NAICS a cui corrisponde.

<b>Nome della divisione(SIC)</b>	<b>Nome del settore(NAICS)</b>
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: Differenze tra SIC e 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.

## 2.8.4 Concetti rivisti della classificazione

NAICS cambia concetti chiave e definizioni della classificazione. Questi cambiamenti tecnici riguardano sostanzialmente come i business sono classificati 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 dei business basata sulla produzione:*** 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 pubblicità mediante mass-media. I commercianti all'ingrosso vendono grandi quantità di beni usando metodi *business-oriented* come cataloghi specializzati, contatti con i clienti e warehouse o office location. Questa ridefinizione fornisce classificazioni e statistiche più ricche di significato per ogni settore, ma le cambia anche. Per 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 compagnia. 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.

### **2.8.5 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 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, entrambi NAICS e SIC includono "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. Gli accordi hanno fissato solo i primi cinque caratteri per l'uso da parte dei tre stati. Il sesto carattere 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.

	<b>Settori mantenuti</b>		<b>Settori nuovi</b>	
<b>Livello</b>	<b>Codice</b>	<b>Titolo</b>	<b>Livello</b>	<b>Codice</b>
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: Settori nuovi e settori mantenuti**

## **2.8.6 Le prospettive del Nord America**

NAICS è stato sviluppato, viene implementato e sarà mantenuto dalle agenzie di statistica del Canada, del Messico e degli Stati Uniti. 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 *cross-border trade flows* e *business market*.

## **2.8.7 Adattamento a NAICS**

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 facendo la traduzione tra statistiche basate su NAICS e SIC sono le tabelle di corrispondenza che dettagliano le relazioni tra NAICS Stati Uniti e SIC.

La Tabella 13 mostra tutte 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 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: 1997 NAICS matched to 1987 SIC**



## **Capitolo 3**

# **Le ontologie: punti di forza nella gestione del commercio elettronico**

### **3.1 Generale**

I computer stanno evolvendo, da singole e isolate applicazioni, verso parte di una rete mondiale per lo scambio delle informazioni e delle transazioni chiamata World Wide Web (WWW). Perciò supportare lo scambio di dati, informazioni e conoscenze è un obiettivo centrale della Information Technology. Le ontologie forniscono una comune e condivisa comprensione di un dominio, che può essere comunicata tra persone e applicazioni. Per questo motivo giocano un ruolo chiave nel supportare i processi di scambio delle informazioni in varie aree.

Le ontologie sono argomento di ricerca comune di vari settori come l'ingegnerizzazione della conoscenza, il processamento del linguaggio naturale e l'Integrazione Intelligente delle Informazioni [Fensel, 2001a]. Sono state sviluppate nell'ambito dell'Intelligenza Artificiale per facilitare lo scambio ed il riutilizzo della conoscenza. Le ontologie forniscono un concettualismo specifico che descrive la semantica dei dati. Hanno funzioni simili allo schema di un database. Le differenze sono:

- Il linguaggio per definire le ontologie è sintatticamente e semanticamente più ricco di un comune approccio per database.

- Le informazioni che sono descritte da un'ontologia consistono in linguaggio naturale semi-strutturato e non di informazioni tabulari.
- Un'ontologia deve essere una terminologia condivisa e consensuale dato che è usata per lo scambio e la condivisione delle informazioni.
- Un'ontologia fornisce una teoria relativa ad un dominio e non la struttura di un contenitore di dati.

Correntemente, le tecnologie Internet e il World Wide Web sono le principali infrastrutture tecnologiche per lo scambio di informazioni on-line. Non sorprende vedere che stanno sorgendo molte iniziative in quest'area per fornire notazioni per la struttura e la semantica dei dati . Per esempio:

- Il *Resource Description Framework (RDF)*<sup>18</sup> fornisce uno standard per descrivere la semantica delle informazioni (per mezzo dei metadati) [Miller,1998].
- *L'Extensible Markup Language (XML)* fornisce uno standard per descrivere la struttura delle informazioni [Connolly, 1997].
- *XML schema* fornisce uno standard per descrivere la struttura e la semantica dei dati.
- *Il linguaggio di trasformazione XSL - T*<sup>19</sup> fornisce lo standard per descrivere il mapping tra differenti terminologie[Clark,1999].
- *Vari linguaggi di interrogazione* per l'XML forniscono standard per descrivere il mapping tra differenti terminologie [QL,1998].

Questi standard permetteranno alle tecniche dell'ontologia di entrare velocemente nei marketplace, supportando vari aspetti dello scambio di informazioni. Ci sono tre settori di business che corrispondono a tre diversi tipi di reti:

- *Intranet*: riservato ai membri di una comunità, compagnia o organizzazione.
- *Internet*: ad accesso libero; rete mondiale di utilizzatori, per esempio, lo shopping on-line ed il Business to Consumer (B2C).
- *Extranet* : limitato accesso dall'esterno (Internet) ad un Intranet, per esempio, nel Business to Business (B2B) nel business elettronico.

---

<sup>18</sup> RDF: Resource Description Framework 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.

<sup>19</sup> 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[Fensel, 2001 a].

## 3.2 Extranet : B2B

Iniziative per supportare lo scambio di dati tra diverse compagnie nei processi di business esistevano già negli anni '60. Per svolgere transazioni gli attori devono accordarsi su standard comuni, in altre parole un protocollo per trasmettere i contenuti e un linguaggio per descriverli. L'uso delle infrastrutture di Internet per le transazioni di business migliorerà significativamente la situazione. Un problema è dato dal fatto che l'HTML non fornisce i mezzi per presentare sintassi ricche e la semantica dei dati. XML che è stato progettato per colmare queste lacune nella Internet Technology, cambierà rapidamente la situazione. XML fornisce una sintassi standard per definire la struttura della semantica dei dati. Perciò, le ontologie giocheranno due ruoli chiave nel commercio elettronico basato sull' XML:

- *Ontologie Standard:* devono essere sviluppate coprendo le varie aree di business. In aggiunta agli standard ufficiali, i marketplace on-line (portali Internet) possono generare standard *de facto*. Se sono in grado di attirare una percentuale significativa delle transazioni on-line in un campo di business creeranno effettivamente un'ontologia standard per quest'area.
- *Servizi di traduzione basati su XSL - T* tra differenti strutture di dati in aree dove non esistono le ontologie standard o dove un particolare cliente vuole usare la sua terminologia e necessita di un servizio di traduzione dalla sua terminologia agli standard. Questo servizio di traduzione deve occuparsi delle differenze legate alla struttura alla semantica e al linguaggio.

In altre parole, gli scambi basati sull'ontologia estenderanno significativamente il livello di automazione degli scambi di dati e creeranno modelli di business completamente nuovi [ McGuinness, 1999].

### 3.3 Le ontologie

Le ontologie sono state sviluppate nell'ambito dell'Intelligenza Artificiale per facilitare la condivisione ed il riutilizzo della conoscenza. Fin dall'inizio degli anni novanta le ontologie sono state un tema di ricerca ricorrente. Il motivo per cui le ontologie sono diventate così popolari si deve ricercare nelle loro promesse: una condivisa e comune comprensione di alcuni domini che può essere trasmessa tra le persone e le applicazioni. Le ontologie sono sviluppate per fornire una semantica delle sorgenti di informazioni processabile dalla macchina e che possa essere comunicata tra vari agenti (software e umani). Negli ultimi anni sono state fornite molte definizioni di ontologia, ma quella che ne caratterizza al meglio l'essenza è: *un'ontologia è una formale, esplicita specificazione di un concettualismo condiviso* [Gruber, 1993]. "Concettualismo" si riferisce ad un modello astratto di un fenomeno che identifica i concetti rilevanti di quel fenomeno. "Esplicito" significa che il tipo di concetti utilizzati e i vincoli sul loro utilizzo sono esplicitamente definiti. "Formale" si riferisce al fatto che le ontologie devono essere leggibili dalla macchina. Di conseguenza sono possibili diversi livelli di formalità. Ampie ontologie come **WordNet**<sup>20</sup>, di cui si parlerà più approfonditamente in seguito, forniscono un thesaurus con oltre 100.000 termini appartenenti al linguaggio naturale spiegati col linguaggio naturale. Il termine "condiviso" riflette il fatto che le ontologie raccolgono conoscenze consensuali, che, non sono ristrette ad alcuni individui, ma sono accettate da un gruppo. Il ruolo dell'ontologia nel processo di ingegnerizzazione della conoscenza è facilitare la costruzione di un modello del dominio. Un'ontologia fornisce un vocabolario di termini e relazioni che modellano il dominio. Lo sviluppo delle ontologie è spesso un processo cooperativo che coinvolge differenti persone, possibilmente da diverse posizioni. In relazione con il loro livello di generalità si possono identificare diversi tipi di ontologie:

- Le *ontologie di dominio* riuniscono la conoscenza valida per un particolare tipo di dominio (elettronico, medico, meccanico).
- Le *ontologie dei metadati* forniscono un vocabolario per descrivere il contenuto delle sorgenti di informazioni on-line [Weibel, 1995].

---

<sup>20</sup> [www.cogsci.princeton.edu/~wn](http://www.cogsci.princeton.edu/~wn): sito dal quale è possibile scaricare la versione 1.7.1 di WordNet e che contiene informazioni ufficiali riguardo tale dizionario lessicale.

- Le *ontologie generiche* o di senso comune mirano a riunire la conoscenza generica riguardo al mondo, fornendo notazioni base e concetti per oggetti come tempo, spazio ed eventi [Fridman-Noy & Hafner,1997]. Di conseguenza, sono trasversali a vari domini.
- Le *ontologie rappresentative* non sono legate ad un particolare dominio. Tali ontologie forniscono entità rappresentative senza precisare che cosa si dovrebbe rappresentare [Gruber, 1993]. Definiscono strutture che permettono di esprimere la conoscenza in modo object-oriented.

### **3.4 Il commercio elettronico come area di applicazione delle ontologie**

In questa parte si esamineranno il commercio elettronico ed il B2B. Si prenderà in considerazione la necessità di standard e formati estesi per lo scambio dei dati e delle informazioni, si evidenzierà come le ontologie giochino un ruolo centrale in quest'area[McGuinness, 1999].

L'area di cui parliamo è ancora molto giovane ed in piena evoluzione. Perciò non esistono tassonomie consolidate. L'obiettivo comune degli approcci che si discuteranno è il miglioramento dell'usabilità del WWW per il commercio elettronico per mezzo dell'aumento dell'accessibilità. Si possono distinguere tre tipi di approcci:

- *Agenti per la ricerca automatica delle informazioni* che aiutano i clienti a trovare i prodotti.
- *Provider di informazioni intelligenti* che aiutano i venditori a presentare i loro beni in una maniera appropriata.
- *Broker di informazioni intelligenti* che mediano tra compratori e venditori.

Attualmente le potenzialità del commercio on-line non sono pienamente sfruttate:

- *Ricerca di prodotti individuale.* Per definizione, i negozi on-line rendono disponibili on-line le informazioni relative ai prodotti. Perciò vengono eliminate barriere fisiche e temporali per accedere a tali informazioni. In teoria dovrebbero bastare solo pochi click per accedere alle informazioni, in

realtà trovare il negozio on-line che vende il prodotto cercato con un prezzo ragionevole è un'operazione che richiede ancora molto tempo.

- *Ricerca di prodotti corporativa.* Si possono costruire automaticamente dei profili utente che supportino nella ricerca del prodotto cercato. Le strategie corporative cercano di individuare utenti simili e usano le loro scelte come consigli.
- *Trasparenza del mercato.* Per i mercati tradizionali le informazioni sui prodotti e sul mercato stesso sono illusorie, per esempio i costi per ottenere informazioni complete sono molto maggiori dei benefici che permettono di ottenere. Con lo shopping on-line si può ottenere una completa trasparenza. Tutte le informazioni sono disponibili on-line e vi si può accedere per leggerle e confrontarle. Non è comunque possibile visitare manualmente tutti i negozi on-line per estrarre informazioni e poi paragonarle.
- *Accesso facile.* L'acquisto di un prodotto è libero da barriere fisiche e di tempo e l'intero processo può essere automatizzato.
- *Negoziazione.* La negoziazione e le aste aiutano ad allocare le risorse in una maniera migliore. La negoziazione e le case d'asta automatiche eliminano elevati costi di transazione e ottimizzano l'allocazione delle risorse.

### 3.4.1 Shopbot

“ I softbots (software robots) sono agenti intelligenti che usano strumenti e servizi software nell'interesse delle persone.” [Etzioni,1997]

Shopbot<sup>21</sup> sono agenti per la ricerca, il filtraggio e l'integrazione delle informazioni, che forniscono una risposta migliore dei generici motori di ricerca e generalmente aggiungono alcuni servizi extra. La loro architettura è basata sul concetto di fornire un accesso integrato a sorgenti di informazioni eterogenee e distribuite [Wiederhold, 1992], [Wiederhold, 1997] e [Wiederhold & Grenesereth, 1997]. I *wrappers* astraggono le informazioni dalla sintassi in cui sono fornite dalle loro sorgenti.

---

<sup>21</sup> Gli shopbot agiscono nell'interesse del compratore, i pricebot agiscono nell'interesse del venditore (cf.[Greenwald & Kephart,1999]).

Il mediatore accede a queste sorgenti per mezzo dei loro wrappers per rispondere a delle query. Deve decidere a quale sorgente di informazioni accedere, può decomporre la query in varie sotto-query per le diverse sorgenti e deve integrare le risposte. Nel caso dello shopping on-line il cliente interagisce con il mediatore per mezzo del browser.

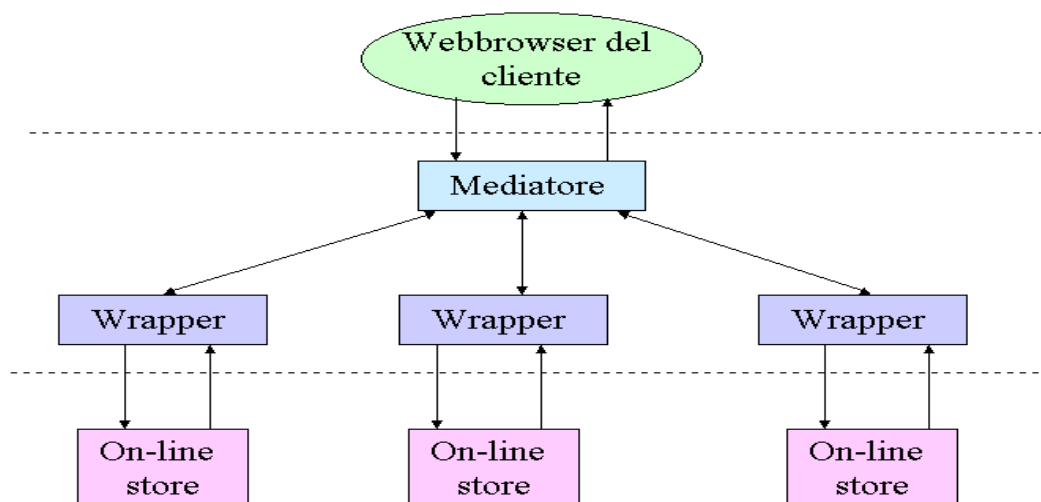


Figura 10: L'architettura degli shopbot.

### 3.4.2 Gli agenti per il commercio elettronico e l'XML

Attualmente il mix di linguaggi naturali, gif e informazioni di layout dell'HTML è la maggior barriera per l'automazione del commercio elettronico, perché la semantica delle informazioni è comprensibile solo per utilizzatori umani. Perciò, non si possono fornire sistemi automatici per processare queste informazioni. Ciò ostacola profondamente la realizzazione e i vantaggi del commercio elettronico. Il servizio di informazioni fornito dagli agenti di shopping è limitato: essi estraggono euristicamente delle informazioni, ma non possono capire pienamente il linguaggio naturale e le attività per mantenere e sviluppare gli agenti di shopping sono molto costose.

Il nuovo standard XML migliorerà molto la situazione. HTML è un linguaggio per presentare documenti di testo, mentre XML è un linguaggio per definire la strutture e la semantica delle informazioni. Perciò, rende possibile la ricerca diretta di informazioni e lo scambio di dati strutturati (per esempio, tra database). Di conseguenza, sarà possibile effettuare il processamento automatico delle informazioni e il commercio elettronico

potrà essere seguito dagli agenti software. Per ora XML fornisce solo una sintassi standardizzata per lo scambio di dati. La definizione di una struttura e di una semantica sono richieste additionally. Questo è precisamente quello che si può ottenere con le ontologie.

### 3.4.3 Sistema di commercio elettronico

Ci sono tre casi di business per il commercio elettronico nell'area del B2B:

- **1:1.** Due compagnie scambiano transazioni di business elettronicamente. Esse necessitano di negoziare uno standard per un sistema comune e lo scambio di dati. Con le nuove tecnologie Internet ciò può essere fatto via TCP/IP e XML. Queste tecnologie Internet forniscono una migliore integrazione con altri processi per lo scambio dei dati ma non cambiano essenzialmente i modelli di business.
- **1:N.** Una compagnia svolge transazioni di business con varie altre compagnie di piccole dimensioni. Abituamente è un grande venditore o un grande compratore che crea tale rete. Esso impone un sistema comune e uno standard per lo scambio dei dati.
- **N:M.** M compagnie svolgono transazioni di business elettronicamente con N compagnie in un mercato frammentato. Un marketplace basato su Internet può aiutare significativamente a mettere insieme entrambe le parti. Esso fornisce una visione globale del mercato e dà la possibilità di effettuare confronti. Tale marketplace cambierà il modello di business di quel frammento di mercato. Fondamentalmente esso rimpiazza o al limite compete con i tradizionali agenti di intermediazione.

Quali sono i colli di bottiglia da superare per potere installare questo tipo di mercati?

Attualmente il commercio elettronico è ostacolato dalla carenza di standard.



- *Carenza di significati delle rappresentazioni e delle traduzioni:* lo standard per Internet HTML non fornisce la sintassi e la semantica per le informazioni, e standard esistenti per l'e-commerce sono isolati e costosi.
- *La carenza di significati per la gestione dei contenuti (ontologie):* non ci sono descrizioni dei prodotti standard (cataloghi) nei vari sottosegmenti.

### **3.4.4 Le ontologie di prodotto e il commercio elettronico**

L'ontologia rappresenta un supporto nell'integrazione di sorgenti di dati eterogenee e distribuite. Perciò ha un ruolo molto importante nella gestione della conoscenza e del commercio elettronico. I mercati attivano la possibilità di nuovi tipi di interazioni tra venditori e clienti. Il problema dei mercati è che devono interagire con descrizioni di prodotti eterogenee e distribuite. Nella maggior parte dei casi non c'è consenso su quali siano i prodotti in un dominio, come descriverli e quale sia la struttura del catalogo adatto. Quando sarà raggiunto un accordo venditori e compratori potranno richiedere differenti viste dei dati sui prodotti. Le industrie hanno percepito queste questioni come i maggiori colli di bottiglia del B2B. Si sono intraprese numerose iniziative per definire un'ontologia destinata a mediare il commercio elettronico.

Un esempio di ontologia con ampia copertura orizzontale è dato da UNSPSC e da [ecl@ss](mailto:ecl@ss). Entrambi si propongono di sviluppare il commercio elettronico in tutte le possibili aree. Le ontologie con ampia copertura orizzontale hanno generalmente una ristretta copertura verticale. Rosettanet è un esempio di ontologia verticale che descrive nel dettaglio prodotti per l'industria software e hardware. Costruire cataloghi di prodotti condivisi e riutilizzabili coincide con il costruire un'ontologia per un certo dominio. L'ontologia è solo un primo passo verso il pieno potere dell'e-commerce. L'ontologia costruisce una semantica di dati comprensibile dalla macchina. Si apre la possibilità di servizi completamente nuovi e automatizzati. Gli agenti software sono in grado di cercare prodotti, possono formare coalizioni di venditori e acquirenti, possono negoziare o possono aiutare a configurare automaticamente i prodotti e i servizi in accordo con la richiesta dell'utente.

La combinazione di una semantica processabile dalla macchina, basata sull'ontologia, e dello sviluppo di un'ampia gamma di servizi porterà il WWW a piene potenzialità, attraverso quello che si chiama il WEB SEMANTICO [Berners-Lee, 1999].

## 3.5 Esempi di ontologia

Nelle pagine successive si descriverà dettagliatamente l'ontologia WordNet. Lo scopo di questa descrizione è da un lato esemplificare un concetto molto astratto come quello di ontologia, dall'altro fornire una base di conoscenze utile alla comprensione della parte successiva dell'elaborato, nel quale WordNet verrà citato frequentemente.

### 3.5.1 WordNet

WordNet è un database lessicale elettronico, la cui struttura è ispirata alle teorie psicolinguistiche legate alla memoria lessicale umana [Ricciardi, 2000]. È stato sviluppato presso l'Università di Princeton dal Cognitive Science Laboratory sotto la direzione del professor A. George Miller<sup>22</sup> ed esiste anche in una versione multilingua europea<sup>23</sup>.

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 (*synsets*), dove ognuno rappresenta un concetto. I *synset* 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. In questo modo il WordNet conserva tutte le caratteristiche dei vecchi dizionari aggiungendo però degli strumenti in più, infatti, inserendo una parola, *machine*, si ottiene una overview (Figura 11) della stessa,

---

<sup>22</sup> <http://www.cogsci.princeton.edu/~wn/>

<sup>23</sup> <http://www.let.uva.nl/~ewn>

completa di tutti i significati, con i sinonimi e la categoria (nome, verbo) di appartenenza, quindi si può scegliere, per ogni categoria e specificando anche l'esatto significato, il tipo di relazione da visualizzare, dopo di che vengono visualizzati tutti i synset legati in modo gerarchico alla relazione scelta (Figura 12).

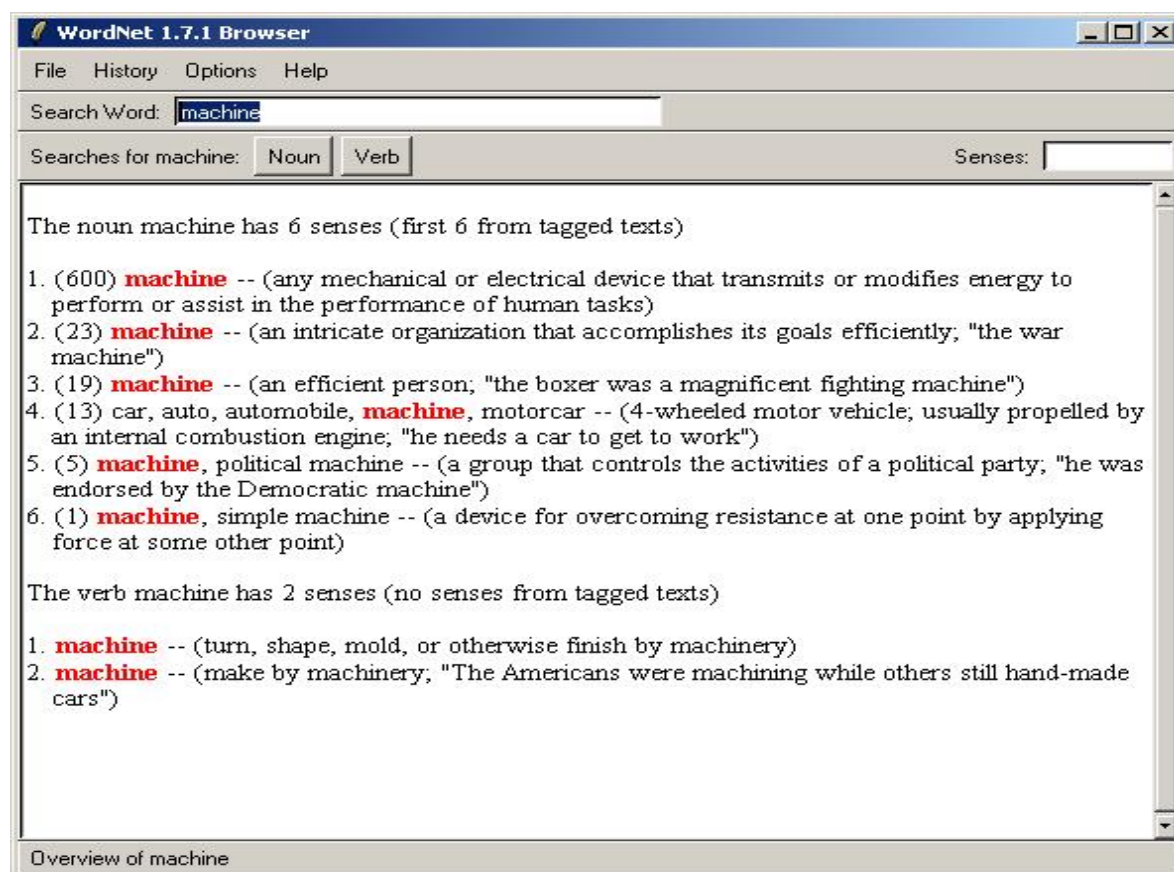


Figura 11: Overview di WordNet

Si può osservare come ogni gerarchia, generata dai diversi significati di *machine*, termina in un nome che viene chiamato beginner. Nel *WordNet* sono presenti 9 diversi beginner (Figura 13), dai quali si generano tutti i nomi presenti nel database.

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 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.

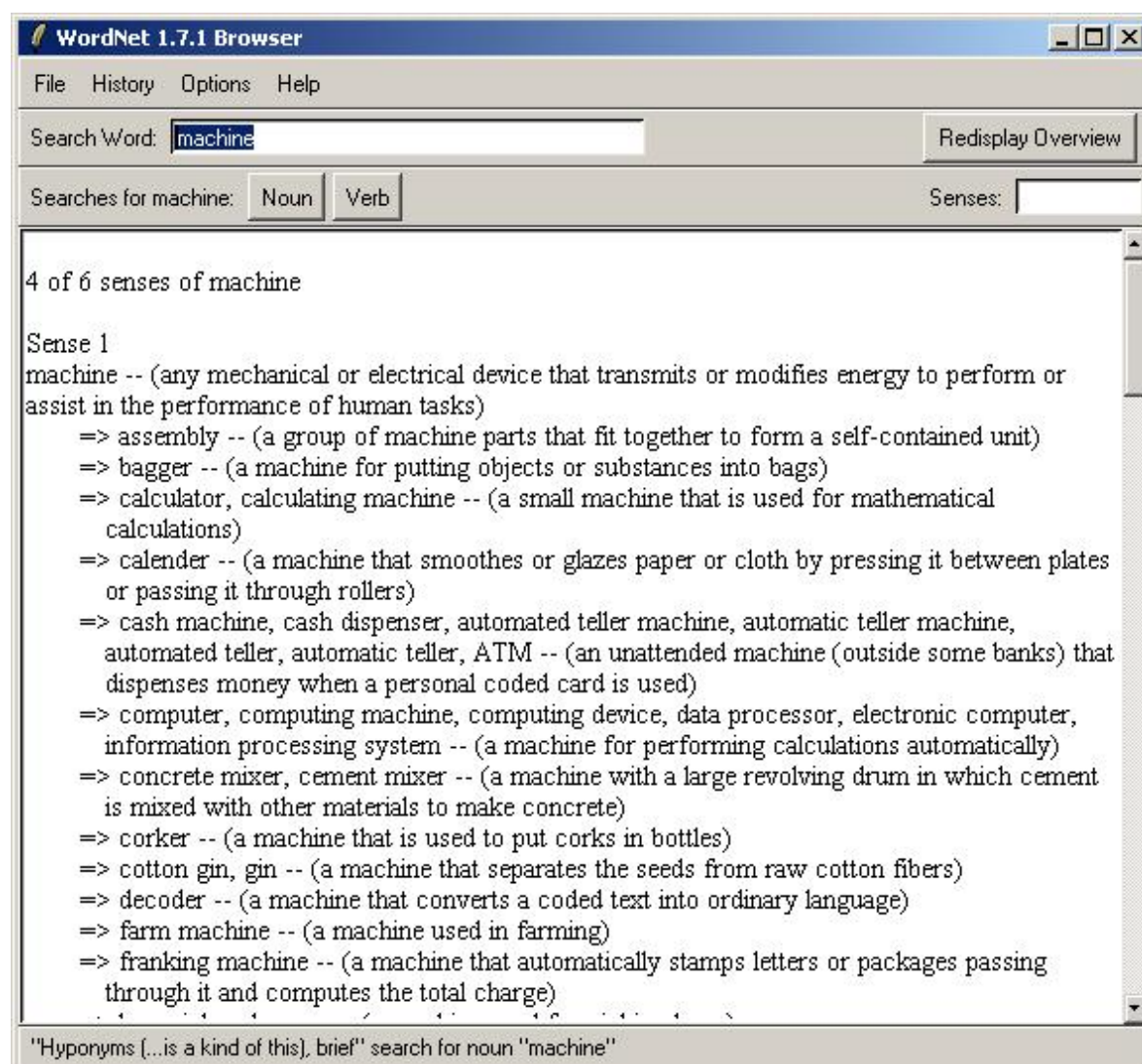


Figura 12: Relazione di iponimia per "machine"

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

Figura 13: Lista dei 9 beginner

### 3.5.1.1 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” o lemma, se ci riferiamo al modo in cui viene letta o scritta.

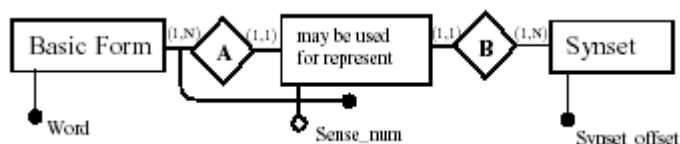


Figura 14: Relazione tra lemmi e significati

Quest'associazione è di molti a molti (vedi Figura 14), 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 15), 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<sub>1,1</sub> implica che il lemma F<sub>1</sub> può essere usato per esprimere il significato M<sub>1</sub>; 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.

Dalla Figura 15 possiamo affermare che F<sub>1</sub> e F<sub>2</sub> sono sinonimi e che F<sub>2</sub> è polisemo. Ogni cella “e” è una definizione (*entry*),  $e=(f, m)$ , dove  $f$  individua la word-form e  $m$  la word-meaning.

Word	Word Forms				
Meaning	F <sub>1</sub>	F <sub>2</sub>	F <sub>3</sub>	-----	F <sub>n</sub>
M <sub>1</sub>	E <sub>1,1</sub>	E <sub>1,2</sub>			
M <sub>2</sub>		E <sub>2,2</sub>			
M <sub>3</sub>			E <sub>3,3</sub>		
-					
-					
M <sub>m</sub>					E <sub>m,n</sub>

Figura 15: Matrice Lessicale

### 3.5.1.2 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 come ognuna di esse è una relazione tra significati; abbiamo anche visto che i significati sono

organizzati in *synset*, quindi è possibile immaginare una relazione semantica come un puntatore tra *synsets*. 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  $\{x, x', \dots\}$ .

### 3.5.1.2.a 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 le 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.

### 3.5.1.2 .b 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*].

### 3.5.1.2.c Iponimia

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 eponimo 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.



### 3.5.1.2.d Meronimia

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

*Un concetto rappresentato dal synset  $\{x,x',\dots\}$  è un meronimo di un concetto rappresentato dal synset  $\{y,y',\dots\}$  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.

### 3.5.1.2.e Relazioni morfologiche

Un'altra importante classe di relazioni lessicali sono le relazioni morfologiche tra word forms. 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.

### 3.5.1.3 Categorie sintattiche

#### 3.5.1.3.a 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).

### 3.5.1.3.b 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 e 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 attualmente contiene approssimativamente 19,500 aggettivi in word forms, organizzati in 10,000 word meanings (*synsets*).

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 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.

### 3.5.1.4 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 *V1* di uno o più verbi *V2* *implica anche V2*. 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 4

## I problemi dell' integrazione

### 4.1 Generale

Il sensazionale aumento di flessibilità offerto dal B2B di Internet pone difficili obiettivi di integrazione tra dati provenienti da diverse fonti. Il marketplace che funziona come livello di comunicazione intermedio riduce il numero di mapping necessari da  $n*m$  a  $n+m$ . Per fornire questo servizio è necessario affrontare il problema dell'eterogeneità dei cataloghi e dei documenti. Sviluppare un approccio scalabile all'integrazione delle informazioni è diventato un prerequisito fondamentale. La carenza di standard, o meglio l'inconsistenza dei nuovi standard generati, rende tutte queste attività complicate [Fensel, 2001b].

### 4.2 La gestione delle informazioni

Un marketplace di successo deve fornire un protocollo comune per lo scambio delle informazioni. Il vero problema è nell'eterogeneità che sorge su tre livelli:

1. contenuti;
2. struttura del catalogo prodotti;
3. strutture dei documenti.

Il contenuto delle informazioni scambiate deve essere modellato. Storicamente, si sono evoluti molti diversi modelli per categorizzare e descrivere i prodotti. Strutturare e standardizzare la descrizione dei prodotti è un obiettivo importante del B2B che assicura che i differenti attori possano comunicare tra loro.

Chi fornisce “Content Management Solutions” può fornire un valore aggiunto aiutando i venditori, costruendo un’ontologia per determinati domini di prodotti.

Il catalogo comprende la descrizione dei prodotti, informazioni sul venditore e sul produttore, il lead time, le informazioni sul controllo della qualità, la versione del catalogo, la data e il codice identificativo. Quando sono coinvolti due cataloghi la loro struttura deve essere allineata nel migliore modo possibile.

Analizziamo l’uso 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 venditore e compratore processare i dati comuni richiede l’uso di un linguaggio comune. Ci possono essere 3 tipi di difficoltà :

1. difficoltà relative ai contenuti → la semantica delle informazioni scambiate;
2. difficoltà relative ai cataloghi → la struttura delle informazioni scambiate;
3. difficoltà relative ai documenti → la struttura delle informazioni scambiate.

### **4.3 L’integrazione delle informazioni**

Gli attori del B2B utilizzano diversi standard nei documenti e differenti “*content standard*” per specificare i prodotti [Omelayenko, 2001a]. Il mercato deve essere in grado di integrare molti documenti e molti standard di contenuto. L’ampio numero di standard e la loro complessità rendono complicati i problemi di integrazione.

Il mercato elettronico deve dare la possibilità ai partecipanti di confrontare facilmente le offerte e svolgere la negoziazione rapidamente grazie allo scambio di informazioni tra i partecipanti. Si devono però risolvere problemi di eterogeneità di prodotti, cataloghi e documenti:

- differenti standard di descrizione dei documenti;
- differenti standard di descrizione dei prodotti;
- differenti cataloghi di prodotto.

Le integrazioni da effettuare hanno come scopo l'ottenimento di:

- mapping dei documenti di business;
- allineamento dei content standard.

## 4.4 I “Content Standard”

Ci sono *content standard* che, come anticipato nel capitolo precedente, possono essere considerati orizzontali e altri che possono essere considerati verticali. Quelli orizzontali generano una classificazione di alto livello di tutti i prodotti appartenenti a molti domini. Forniscono una estesa ma superficiale gerarchia di categorie di prodotti, ma generalmente non contengono gli attributi dei prodotti. Lo standard UNSPSC, per esempio, contiene 15000 categorie, mentre NAICS ne contiene oltre 3000; entrambi gli standard hanno 5 livelli e i loro alberi di classificazione non forniscono attributi.

Gli standard verticali forniscono una classificazione più approfondita di un dominio più ristretto, gli standard verticali forniscono anche gli attributi delle categorie. Per esempio RosettaNet contiene un catalogo di prodotti IT che espande 136 elementi di UNSPSC in 445 categorie con 2660 attributi. Un altro standard è [ecl@ss](#) che usa oltre 13000 categorie per descrivere il dominio industriale delle macchine e dei materiali. Dal punto di vista dell'ingegnerizzazione della conoscenza le gerarchie dei prodotti con i loro attributi possono essere viste come ontologie di prodotto. Lo scambio di documenti richiede il mapping di queste ontologie.

Le ontologie hanno due livelli di composizione gerarchica: all'ontologia generale (che corrisponde agli standard orizzontali) si legano ontologie specifiche dei domini (gli standard verticali). Sono necessari tre tipi di mapping per integrare le ontologie: il mapping tra standard orizzontali, tra orizzontale e verticale e tra verticali [Omelayenko, 2001a].

Il *mapping degli standard orizzontali* ha le seguenti proprietà:

- gli standard differiscono molto nella classificazione a causa dell'assenza di consenso nello schema di classificazione;
- gli standard differiscono molto nel livello di granularità della classificazione;
- non ci sono equivalenze evidenti tra le descrizioni delle categorie.

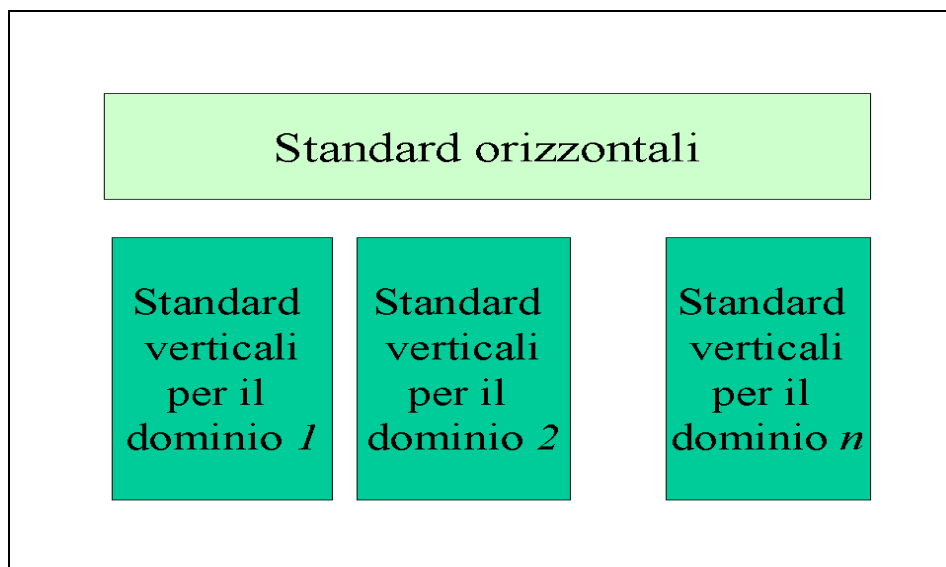


Figura 16: Standard orizzontali e verticali

Il *mapping degli standard verticali* ha le seguenti proprietà (in aggiunta a quelle già citate per il mapping orizzontale):

- richiede il mapping tra i nomi degli attributi e tra i loro tipi;
- richiede la trasformazione tra diversi format di valori e scale di unità;
- i domini degli attributi devono essere mappati;
- ogni catalogo introduce il suo set di attributi da mappare.

Il *legame tra standard orizzontali e verticali* impone i seguenti problemi:

- mapping tra un numero relativamente piccolo di concetti verticali di alto livello e concetti più generali dello standard orizzontale;
- mapping di concetti che sono fuori dal focus dello standard verticale per farli corrispondere a concetti dello standard orizzontale.

Questi problemi sono molto importanti nel B2B, ma non sono nuovi nei database, nell'ingegnerizzazione della conoscenza e in area di modellazione delle informazioni.

In questa tesi verrà proposto un approccio all'integrazione di standard di classificazione dei prodotti che fa uso di un sistema a mediatore. Quindi, le principali caratteristiche degli standard e i problemi incontrati nell'integrazione verranno descritti più ampiamente in seguito.



#### 4.4.1 Problemi semantici relativi ai *content standard*

Pur ipotizzando che anche sorgenti diverse condividano una visione simile del dominio da modellare, e, quindi, un insieme di concetti comuni, niente ci garantisce che i diversi sistemi usino esattamente gli stessi vocaboli per rappresentare questi concetti, né tantomeno le stesse strutture. Poiché le diverse sorgenti sono state progettate e modellate da persone differenti, è molto improbabile che queste persone condividano la stessa concettualizzazione del mondo esterno, ovvero non esiste nella realtà una semantica univoca a cui chiunque possa riferirsi.

Se due consorzi progettano autonomamente gerarchie di classificazione dei prodotti, le due gerarchie presenteranno sicuramente delle differenze semantiche: per esempio in una (UNSPSC) i libri sono classificati all'interno della classe "Printed Publication", nell'altra ([ecl@ss](mailto:ecl@ss)) i libri sono classificati nel ramo di gerarchia relativo a "Communication Technology".

La causa principale delle differenze semantiche si può identificare nelle diverse concettualizzazioni del mondo esterno che persone distinte possono avere, ma non è l'unica [Saltor, 1997]. Le differenze nei sistemi di DBMS (DataBase Management Systems) possono portare all'uso di differenti modelli per la rappresentazione della porzione di prodotti in questione. L'obiettivo dell'integrazione, che è fornire un accesso integrato ad un insieme di sorgenti, si traduce allora nel non facile compito di identificare i concetti comuni all'interno di queste sorgenti e risolvere le differenze semantiche che possono essere presenti tra di loro. Possiamo classificare queste contraddizioni semantiche in due gruppi principali:

1. *eterogeneità tra le classi di oggetti*: benché due classi in due differenti sorgenti rappresentino lo stesso concetto nello stesso contesto, possono usare nomi diversi;
2. *eterogeneità tra le strutture delle classi*: comprendono le differenze nei criteri di specializzazione, nelle strutture per realizzare una aggregazione, ed anche le discrepanze schematiche.

## 4.5 L'infrastruttura dei documenti per il B2B

I documenti utilizzati sono collegati nel modo seguente:

- I cataloghi di prodotto descrivono i prodotti, i loro prezzi, la quantità disponibile. Sono legati ai “*content standard*” che specificano la posizione dei prodotti nella gerarchia dei prodotti, dove ogni categoria fornisce un set di attributi usati per specificare il prodotto. Oltre questi set, il catalogo può fornire attributi addizionali di prodotto. Il catalogo, inoltre, introduce un codice interno di identificazione della compagnia per i prodotti utilizzati da altri documenti di negoziazione.
- Altri documenti, che non contengono la descrizione degli elementi commerciali, sono legati solo ai codici di prodotto interni e al numero dei documenti o non sono linkati.
- I “*content standard*” sono legati l'un l'altro. I *content standard* orizzontali sono mappati a altri *content standard* orizzontali e a quelli verticali e sono anche mappati tra loro.

A causa della separazione tra i documenti e i “*content standard*” ci sono due tipi diversi di ontologie: ontologie dei documenti e le ontologie dei prodotti.

## 4.6 Standard per i documenti

L'area del B2B opera con un elevato numero di documenti di business. Per esempio, Commerce One offre la “XML Common Business Library” (xCBL) con oltre 600 differenti tipologie di documenti, anche Ariba ha una simile offerta. Entrambi gli standard forniscono le DTD dei documenti e stabiliscono che i documenti vengano forniti in XML. Quindi l'integrazione di documenti si trasforma in integrazione di documenti XML.

Ci sono alcuni standard di documenti non in XML già accettati e utilizzati dalle aziende, come ad esempio EDIFACT<sup>24</sup>, sono però disponibili dei wrapper in grado di tradurre uno di questi documenti nel formato XML. C'è anche uno standard ISO che non è

---

<sup>24</sup> EDIFACT: Electronic Data Interchange for Administration, Commerce, and Transport

espresso in XML. Per tutti gli standard che non sono in XML le rappresentazioni in XML sono già state sviluppate e sono in fase di approvazione da parte della comunità. Gli standard di documenti forniscono solo la struttura dei documenti e non l'ontologia dei documenti stessi. La struttura è rappresentata in documenti DTD che contengono solo una parte delle relazioni. Le varie strutture possono fornire rappresentazioni molto diverse anche quando i contenuti sono molto simili.

Un esempio dei problemi tipici che si possono incontrare sono:

- le ontologie hanno nomi diversi per attributi semanticamente equivalenti;
- alcuni attributi sono mancanti nelle ontologie;
- le ontologie hanno diversi livelli di granularità.

## 4.6.1 Esempio di integrazione di documenti

Prenderemo in considerazione il problema dell'integrazione degli indirizzi. Un indirizzo è un semplice concetto di business che ricorre frequentemente in 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 comprensibili da chiunque e rendono l'esempio chiaro. L'integrazione della descrizione degli indirizzi coinvolge vari e interessanti tipi di problemi che ricorrono nell'integrazione dei prodotti. Si prendono in considerazione in particolare quattro standard per la descrizione degli indirizzi [Omelayenko].

Il primo standard analizzato è xCBL 3.0 sviluppato da CommerceOne<sup>25</sup>. 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. Il Document Type Definition (DTD) per un indirizzo nello standard xCBL è presentato nelle righe successive:

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

---

<sup>25</sup> [www.commerceone.com](http://www.commerceone.com)

Il secondo standard è **Internet Open Trading Protocol (IOTP)**, sviluppato in collaborazione con il consorzio Internet Engineering Task Force (IETF<sup>26</sup>), fornisce un formato standard per le operazioni di pagamento per le transazioni svolte via Internet. È indipendente da qualsiasi specifico sistema di pagamento. IOTP fornisce le strutture ed i protocolli di comunicazione per le transazioni di pagamento: acquisto, autenticazione, deposit, 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. Il 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 specificazioni per i contenuti di business nel settore delle applicazioni, con enfasi su **Open Applications Group Integration Specification(OAGIS)**<sup>27</sup>, specifiche per messaggi di business e scenari di integrazione. Fornisce strutture di dati, format e protocolli per le integrazioni di business. OAGIS definisce un vocabolario per i termini di business e si possono 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?, TAXJRSRCTN?,TELEPHONE*, URL?, USERAREA?)>
```

La Versione 1.0 del Real Estate Data Interchange Standard (RETS, 2000) definisce le informazioni per lo scambio di beni. Definisce un'interfaccia standard attraverso la quale un programma Client può comunicare con un property o altri data server. Le specifiche definiscono un protocollo per implementare le transazioni, incorporano una specifica XML per transazioni generiche. Inoltre, fornisce un formato per lo scambio di

---

<sup>26</sup> [www.ietf.org](http://www.ietf.org)

<sup>27</sup> [www.openapplications.org](http://www.openapplications.org)

dati compressi e specifiche per permettere lo scambio di machine-interpretable property information. Le strutture dati per lo scambio sono definite nel **Real Estate Transaction Markup Language (RETML)**<sup>28</sup>, gli indirizzi sono:

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

La rappresentazione dello stesso concetto, l'indirizzo, è differente nei diversi cataloghi. La descrizione dei prodotti può essere espressa in XML utilizzando i tag in modi diversi. Concettualmente uguali 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, il focus su un sistema di spedizione richiede la divisione dell'indirizzo di una compagnia 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>
```

---

<sup>28</sup> [www.rets-wg.org](http://www.rets-wg.org)

```

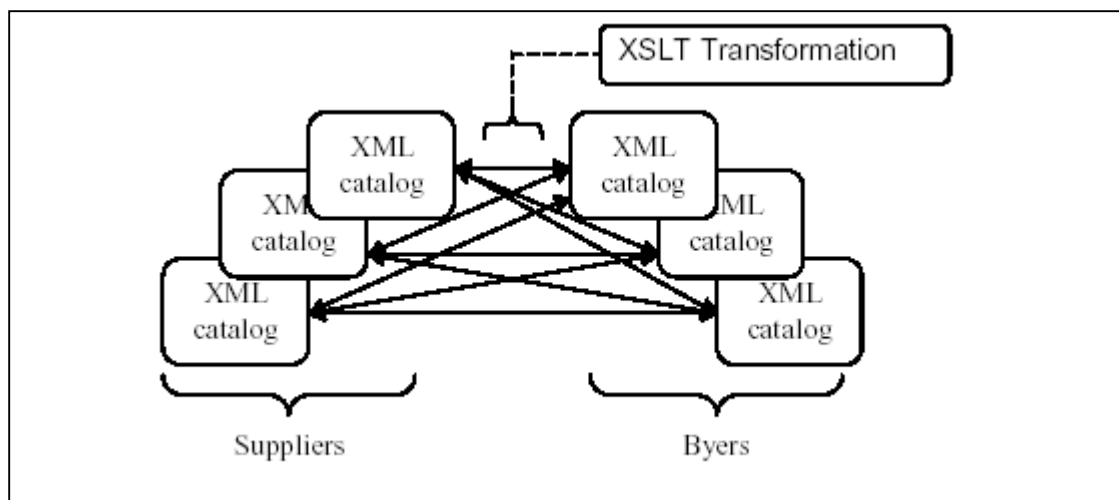
<DESCRIPTN/>
<FAX>3120444765</FAX>
<POSTALCODE>1081 hv</POSTALCODE>
<REGION/>
<STATEPROVN/>
<TAXJRSDCTN/>
<TELEPHONE>3120444765</TELEPHONE>
<URL>www.cs.vu.nl</URL>
<USERAREA/>
</ADDRESS>

```

Se il marketplace media tra  $n$  venditori e  $m$  compratori, deve essere in grado di mappare ciascuno degli  $n$  cataloghi dei fornitori negli  $m$  format dei compratori, svolgendo  $n*m$  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. Ci sono due opposte strategie per selezionare gli elementi da includere nell'UC:

- a) Il catalogo unificato contiene il numero minimo di attributi per ogni prodotto.
- b) 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 strategia (a), 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 strategia (b), l'aggiunta di un catalogo meno dettagliato non influenza l'UC, mentre quella di uno più dettagliato ne richiede l'aggiornamento.



**Figura 17: Integrazione con  $n*m$  mapping**

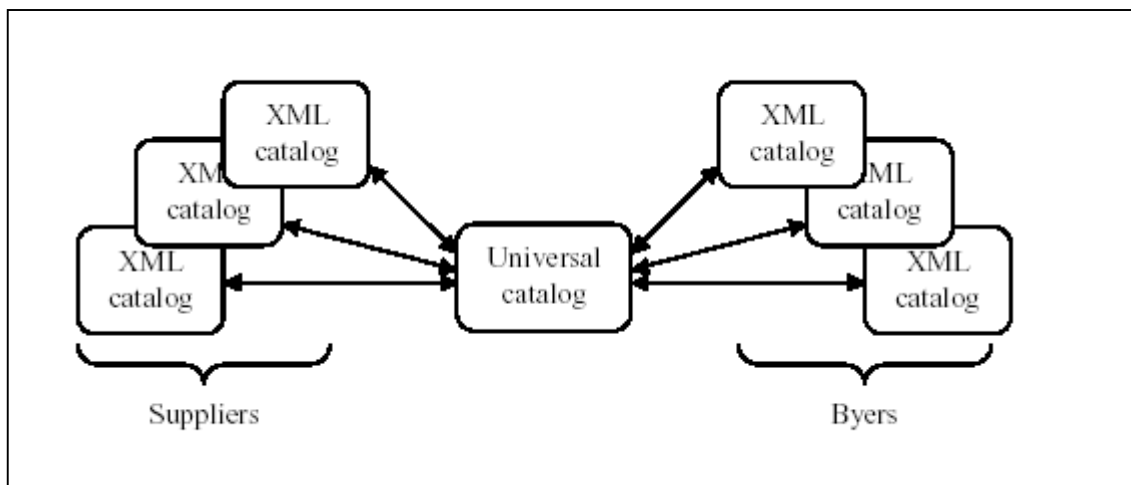


Figura 18: Integrazione con n+m mapping

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
-	-	TAXJRSRCTN?	-	TaxDistrict
-	-	USERAREA?	-	Area
TradingPartner	-	-	-	TimeZone
Timezone?				

Tabella 14: Mapping degli elementi

## 4.7 Modello per l'integrazione del catalogo

Il catalogo può essere integrato con una serie di attività XSL-T che traducono direttamente i vari formati [Omelayenko]. Il problema di questo approccio è che mixa tasks indipendenti, come:

- allineamento di diverse terminologie;
- allineamento della granularità della rappresentazione;
- trasformazione del valore degli attributi;
- ricostruzione di un necessario formato in accordo con il target dei documenti standard.

Svolgere tutto ciò *one-shot* è molto complesso e causa dei seri problemi di implementazione. Si è, dunque, creato un approccio stratificato che permette di separare vari aspetti del processo di trasformazione.

***Livello sintattico:*** corrisponde ai documenti rappresentati con la serializzazione XML, che specifica elementi e attributi utilizzati con il loro ordine.

***Livello del modello dei dati:*** è il ponte tra il livello sintattico e l'ontologia. Le rappresentazioni sono astratte dalla differenza imposta dalla sintassi e gli oggetti sono rappresentati come oggetto – proprietà - valore. La normalizzazione viene svolta in accordo con la corrispondente ontologia. La terminologia usata a questo livello è definita dall'ontologia e generalmente coincide con la terminologia usata a livello sintattico.

***Livello ontologico:*** definisce una terminologia usata per rappresentare le informazioni fornite dai vari standard di prodotti e documenti. Specifica la terminologia ad un livello sufficiente per definire un mapping tra i cataloghi. L'ontologia contiene sia gli elementi specificati che quelli opzionali, assenti nella serializzazione XML.



## 4.8 Mapping dei modelli di dati

La traduzione di un documento richiede tre passi:

1. traduzione di un documento da un modello di dati normalizzato (ABSTRACTION STEP) 1:N;
2. mapping del modello di dati (TRASFORMATION STEP) N:1;
3. serializzazione del modello di dati in XML (REFINEMENT STEP) 1:1.

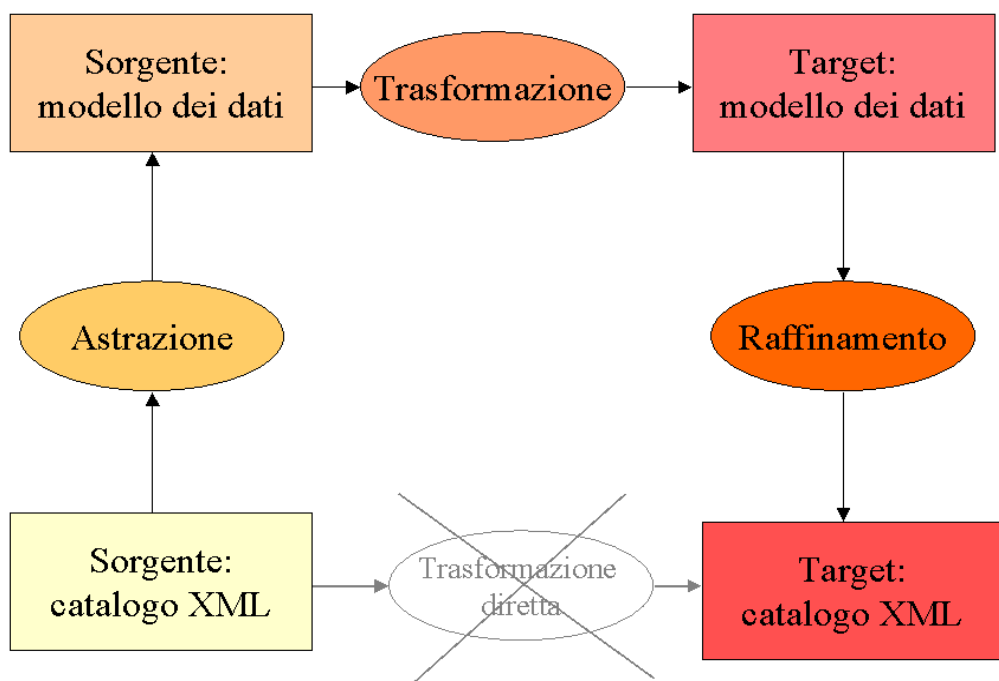


Figura 19: Mapping dei modelli di dati

### *Abstraction Step*

Nel passo di astrazione i cataloghi XML sono tradotti nel loro modello di dati codificato con triple RDF, ciò richiede le seguenti trasformazioni:

- 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.

### ***Translation Step***

Tutto il mapping all'interno del catalogo è fatto a livello del modello di dati. Si assume che tutti gli split tra elementi necessari vengano fatti nel passo di astrazione (sono necessari solo mapping uno ad uno).

### ***Refinement Step***

Sono ristabilite tutte le restrizioni sintattiche del format:

- 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.

## **Capitolo 5**

# **L'Integrazione Intelligente con MOMIS**

### **5.1 L'Integrazione delle Informazioni**

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, dei 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 una eterogeneità semantica (o logica) non risolvibile da questi standard.

Problematiche aggiuntive sono dovute al cosiddetto sovraccarico di informazioni, 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 *datawarehouse* 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 ( $I^2$ ).

### 5.1.1 Integrazione Intelligente delle 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<sup>3</sup>*).

### 5.1.2 Il programma $I^3$

Dal 1992 è operativo il Programma  $I^3$  [Hull, 1995]: un progetto di ricerca fondato e sponsorizzato dall'ARPA (Advanced Research Projects Agency agenzia che fa capo al Dipartimento di Difesa degli Stati Uniti), 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:

- l'orizzontale, divisa in tre livelli: livello utente, moduli intermedi che fanno uso di tecniche di IA, risorse di dati;
- la 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'utilizzatore, 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.

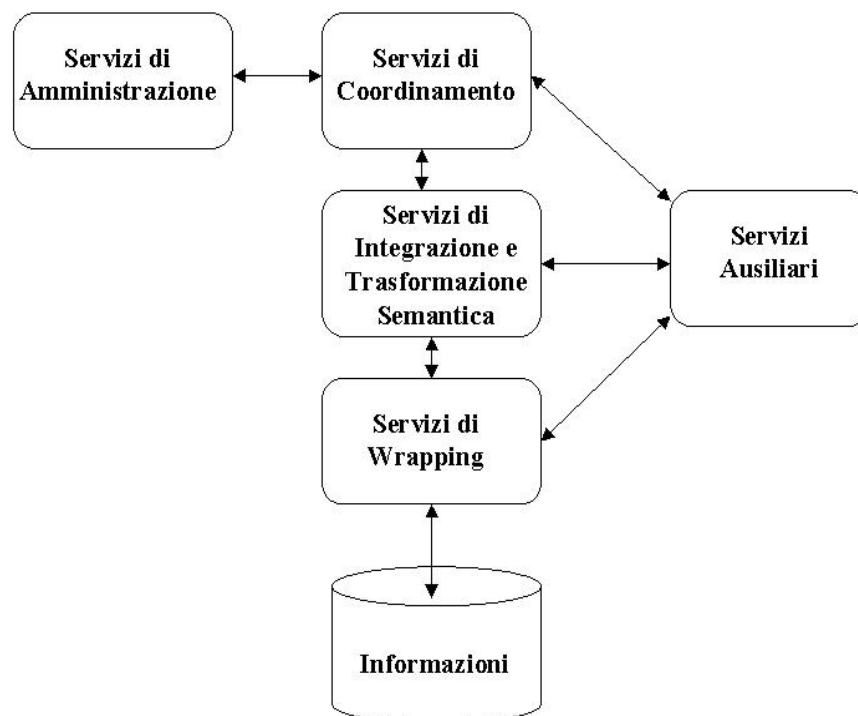


Figura 20: Diagramma dei servizi  $I^3$

### 5.1.3 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 20 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 20 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.

### **5.1.4 Il mediatore**

Secondo la definizione proposta da Wiederhold [Duschka, 1996]:

“un mediatore è un modulo software che sfrutta la conoscenza su un certo insieme di dati per creare informazioni per una 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, del quale si parlerà nelle pagine successive, 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 è 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  $P^3$  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

Il Mediator EnvirOnment for Multiple Information Sources (MOMIS) [Bergamaschi, 1999], è un sistema per l'integrazione intelligente delle informazioni, sviluppato in una collaborazione tra l'Università di Modena e Reggio Emilia e l'Università di Milano. Come si può osservare in Figura 21 la struttura di MOMIS ricalca le specifiche  $I^3$ .

Al livello più basso abbiamo i wrapper, che svolgono la funzione di interfaccia tra il mediatore e le sorgenti, che per il sistema MOMIS possono essere sia strutturate, database, che 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 ODL $I^3$  utilizzato dal sistema;
2. nella fase di query processing: traducono la query, inviata dal mediatore nel linguaggio OQL $I^3$ , 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. 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 ODL $I^3$  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,1998] che, con l'uso di tecniche di logica descrittiva, scompone e ottimizza la query che fornirà in linguaggio OQL $I^3$  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.

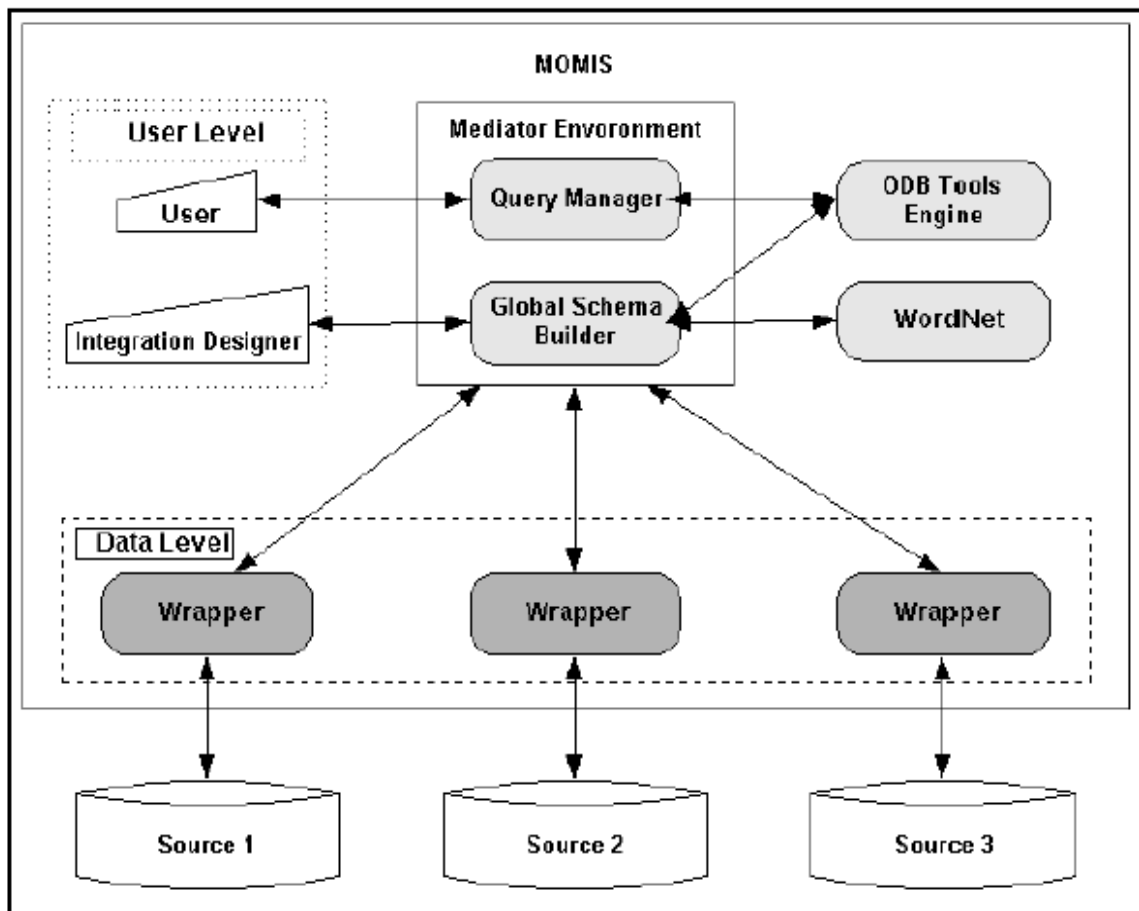


Figura 21: Architettura del sistema MOMIS

### 5.2.1 Il linguaggio ODL<sup>3</sup>

Il linguaggio ODL<sup>3</sup>, realizzato in accordo con le raccomandazioni ODMG [Cattel,1994,1997], 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<sup>3</sup>;
- 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 olonimia **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, 1997a] 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;

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

**Figura 22: Struttura di ODB Tools**

2. l'espansione semantica dei tipi, sempre attraverso l'algoritmo di sussunzione.

Come si osserva in Figura 22 ODB tools è costituito da due parti:

- ODB Designer;
- ODB Qoptimizer[Beneventano, 1997b] .

ODB Designer acquisisce e valida schemi OODB 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.

### **5.3 Il processo d'integrazione**

La prima parte del processo di integrazione del sistema MOMIS[Bergamaschi,1999], come si vede in Figura 23, 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. 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.

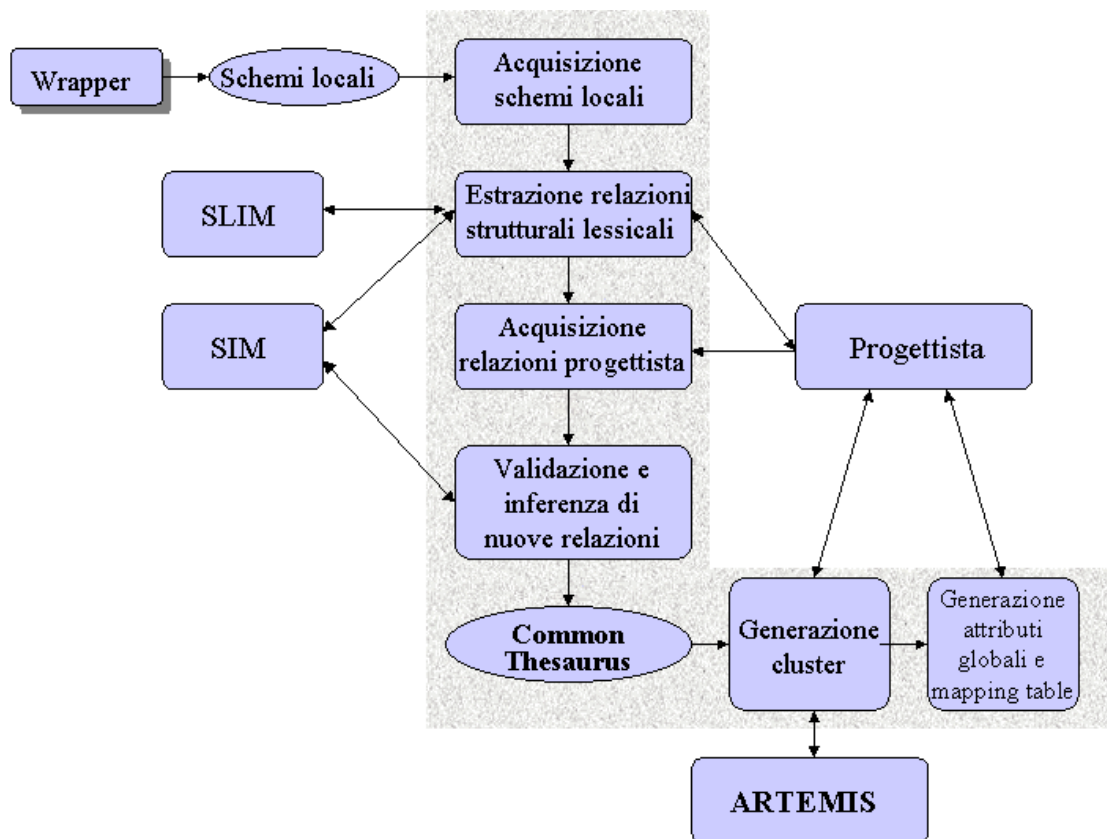


Figura 23: Fasi del processo di integrazione

## 5.4 Generazione del Thesaurus comune

Questa parte dell'integrazione viene operata, come si vede in Figura 23, 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.

### **5.4.1 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.

### **5.4.2 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<sup>3</sup> 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'ipernimia una BT;
- l'iponimia una NT;
- l'olonimia e la correlazione vengono invece tradotte in relazioni RT.

### **5.4.3 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.

### **5.4.4 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, li 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.

### **5.4.5 Inferenza di nuove relazioni**

Anche questa parte del processo di generazione del Thesaurus Comune, come la precedente, ricade su ODB tools.

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 una 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.

## 5.5 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. lo *Structural Affinity Coefficient* che viene determinato tra due classi, in base alle relazione tra i loro attributi;
2. il *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  $\sigma$  a ogni relazione, peso che sarà tanto maggiore quanto più stringente sarà il legame imposto dalla relazione ai due termini considerati:

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

Tipicamente vengono considerati i seguenti valori:

$$\sigma_{sin}=1$$

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

$$\sigma_{rt}=0.5$$

## 5.6 La generazione dei cluster

Calcolato il grado di affinità tra le classi, tutte quelle che hanno una 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.

## 5.7 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.

## **5.8 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.

### **5.8.1 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.

## 5.8.2 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.

## 5.9 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.

Data una classe globale G con m attributi globali ( $A_1, A_2, \dots, A_m$ ) e rappresentativa di n classi locali ( $L_1, L_2, \dots, L_n$ ) la mapping table associata a G è una tabella con n righe e m colonne, nella quale l'elemento ( $i, j$ ) rappresenta in che modo l'attributo globale  $A_i$  è mappato nella classe locale  $L_j$ .

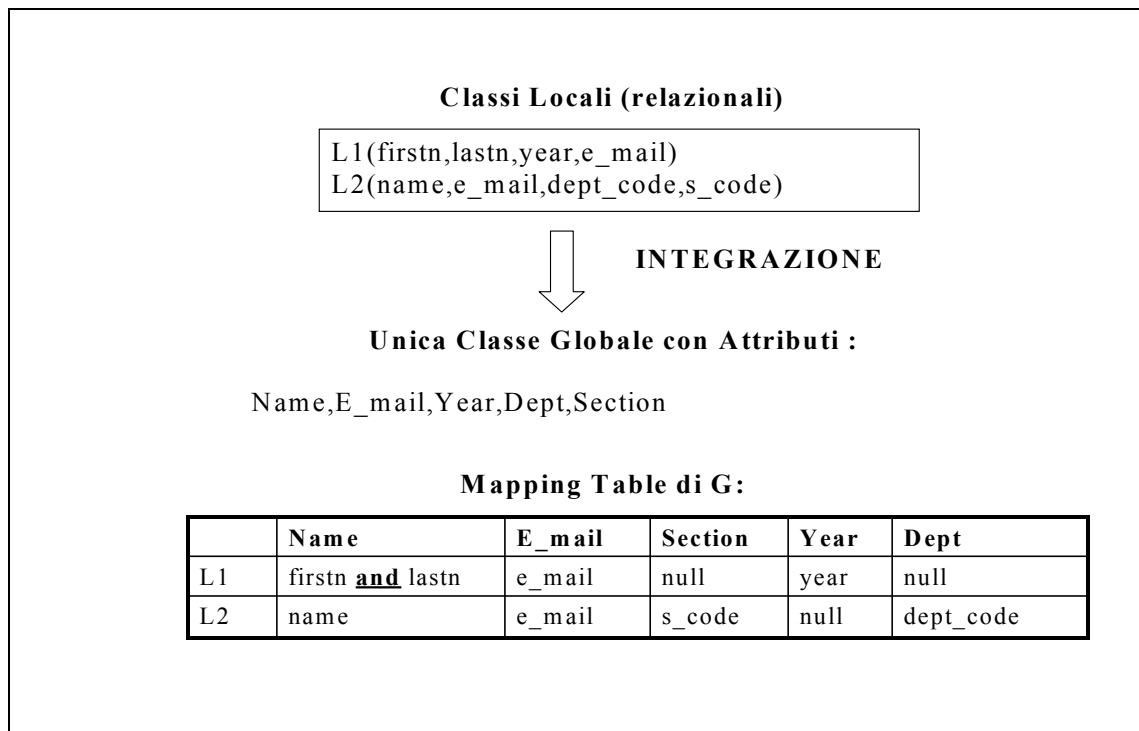


Figura 24: Esempio di Mapping table

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.

## 5.10 SI-Designer

SI-Designer [Beneventano, 2000] 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 25 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 (TUNing of the Mapping table)

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

### 5.10.1 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. Valutate 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.

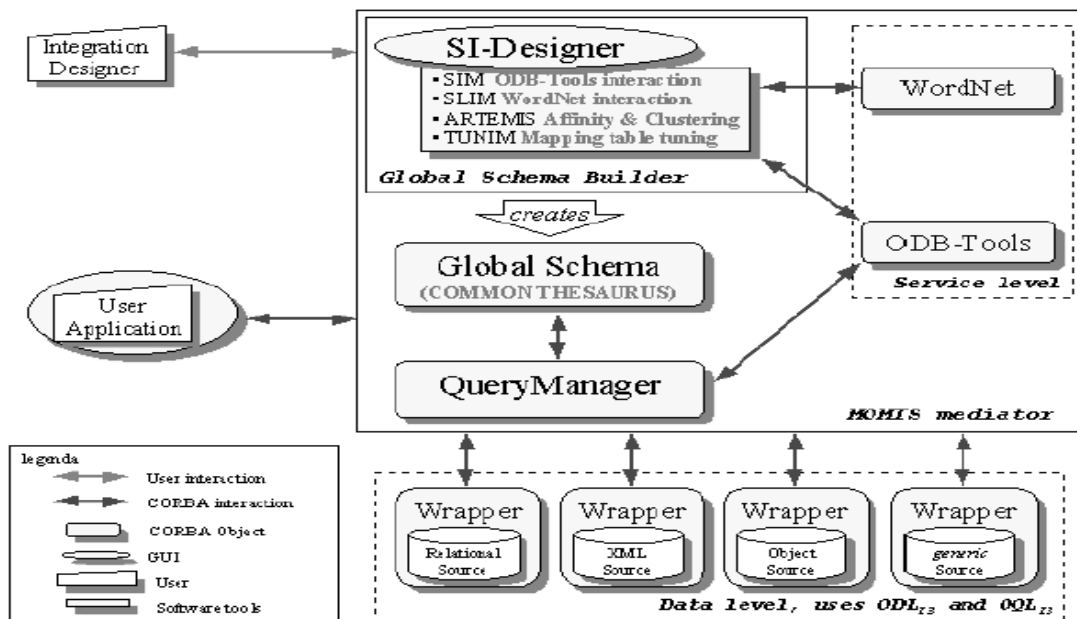


Figura 25: L'architettura del sistema MOMIS

## 5.10.2 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 26. La parte in basso di questo pannello, denominata *textitNew wrapper data input*, è quella necessaria al caricamento dei wrapper; qui 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 *textitWrappers* e la visualizzazione della struttura della sorgente, con la specificazione del tipo (relazionale, a oggetti, semistrutturata), nella finestra di destra denominata *textitSource acquired*.

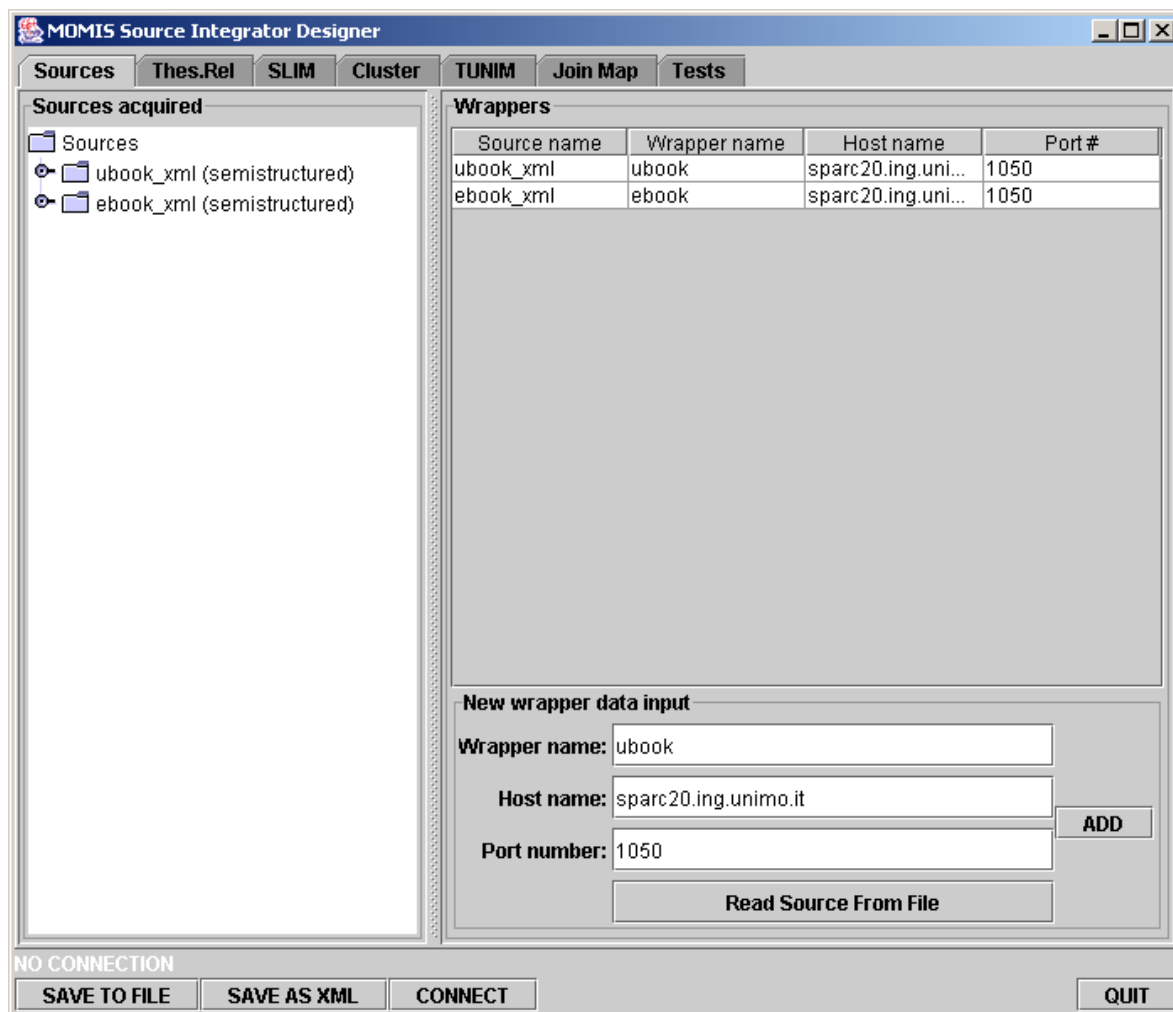


Figura 26: Pannello Source

### 5.10.3 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; per fare questo si accede al pannello *Thes.Rel*, Figura 27, e si seleziona il pulsante run SIMA. 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.



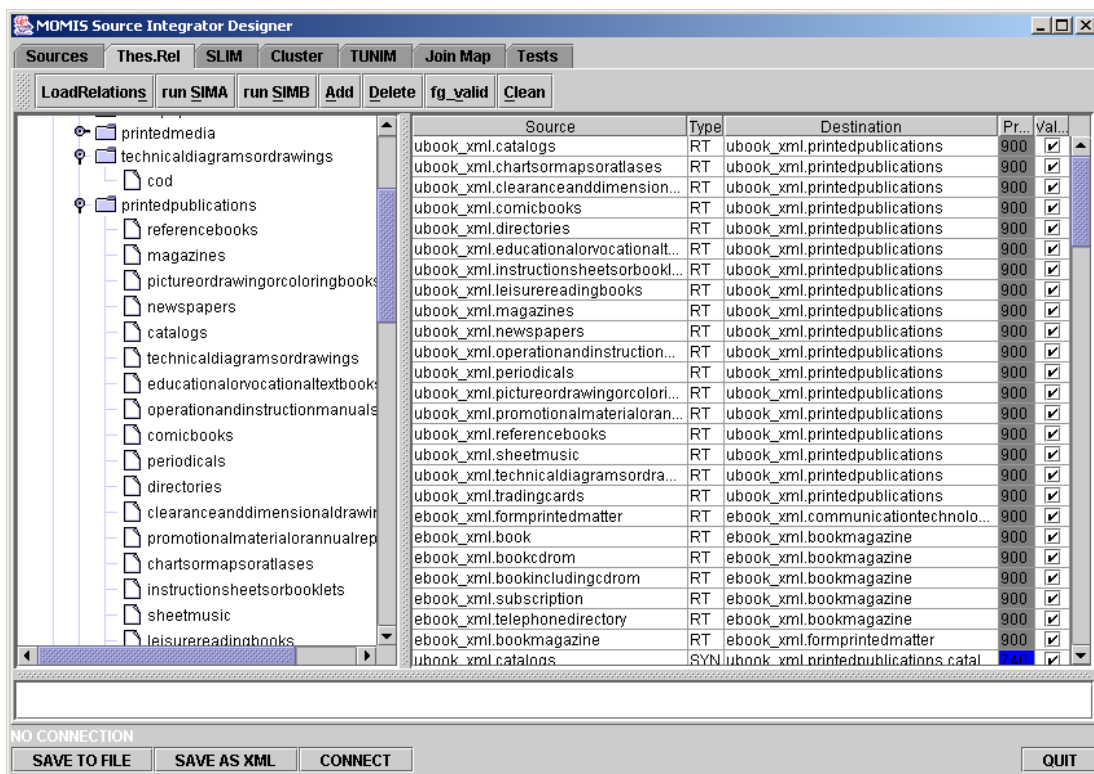


Figura 27: Pannello Thes.Rel: run SIM

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.

### 5.10.4 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*. 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 28, con i seguenti elementi che conducono alla selezione del significato del termine:

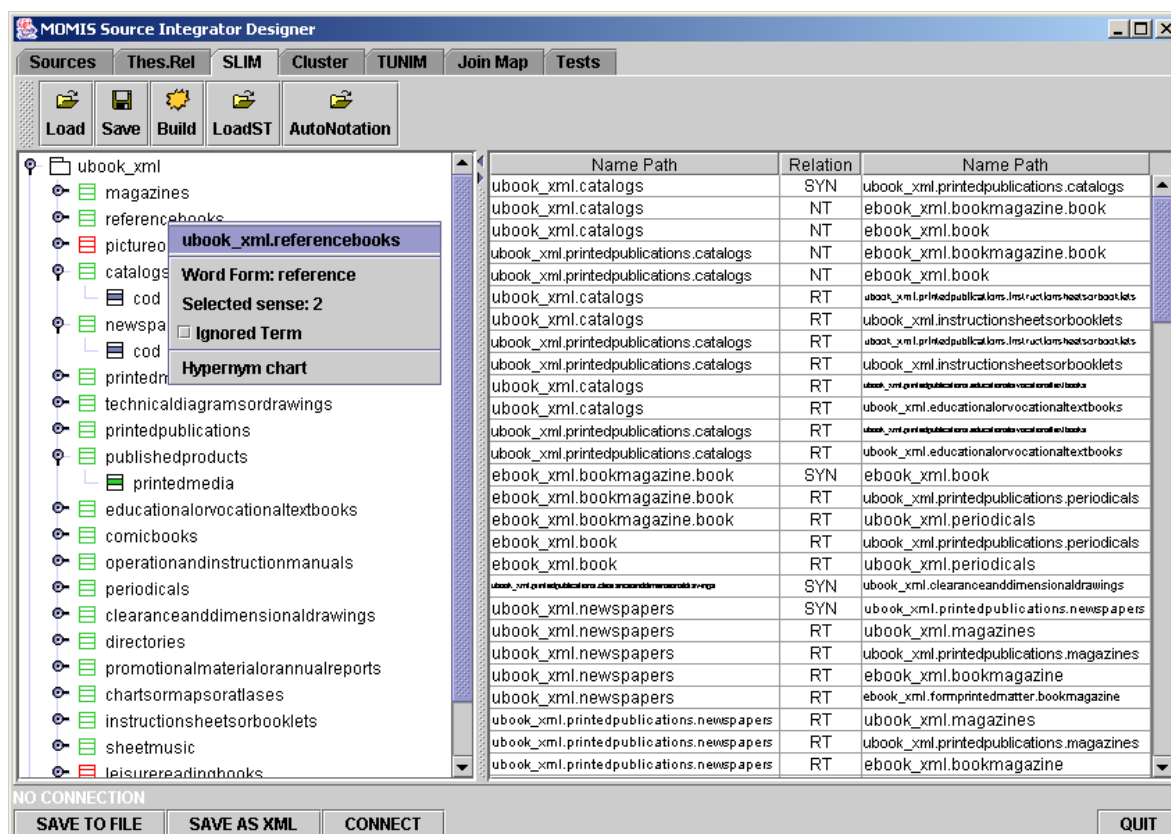


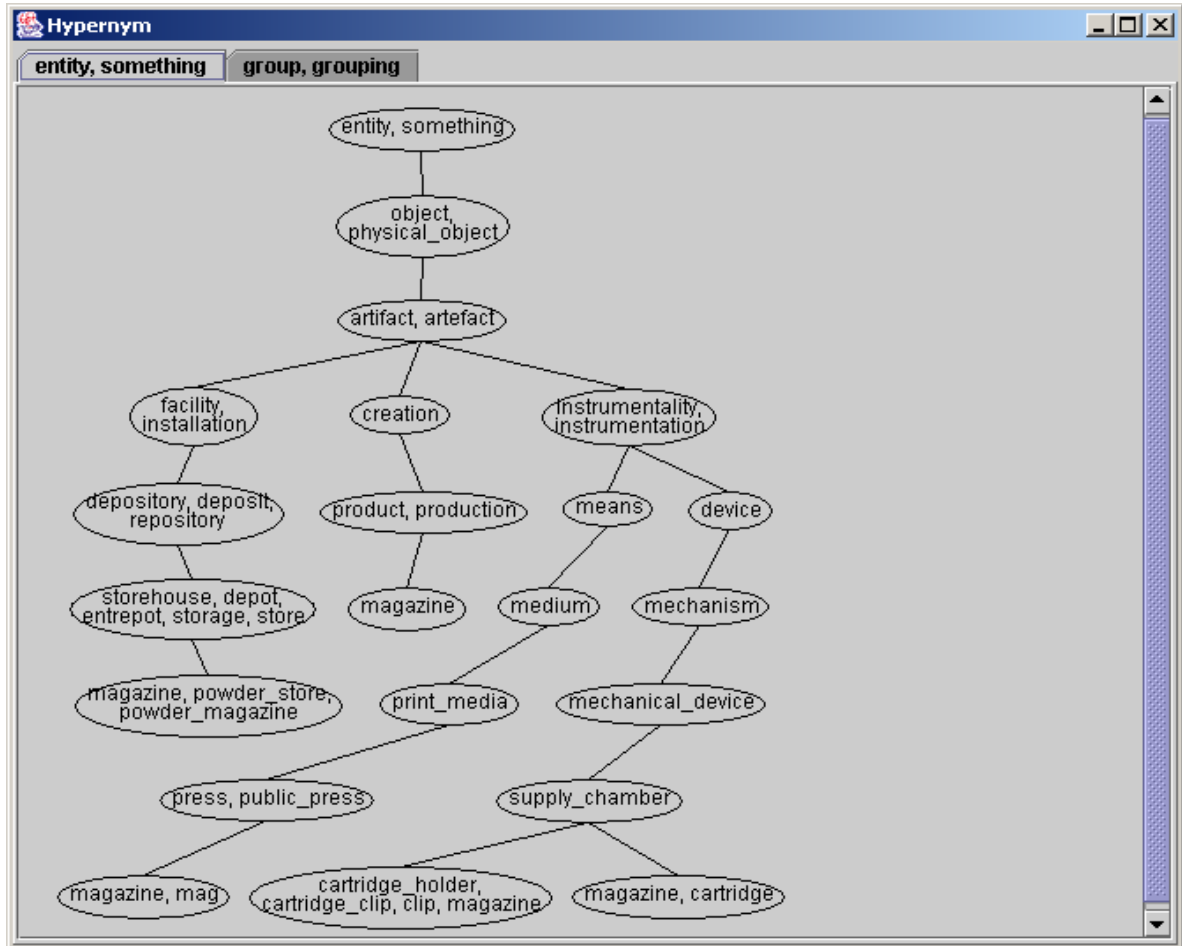
Figura 28: Pannello di SLIM

- ◆ 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;

- Hypernim: 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 (Figura 29).



**Figura 29: Pannello Hypernym**

- 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 previsti due tasti, LOAD e SAVE, per caricare e salvare il significati già assegnati ai termini. In alternativa al pannello SLIM si può utilizzare il pannello ARM, che svolge le medesime funzioni, solamente con una interfaccia a pannelli invece che a menu.

### 5.10.5 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.

### 5.10.6 EXTM Rel

Questo pannello, Figura 30, permette al progettista di sfruttare la sua conoscenza sui legami estensionali tra le classi in modo da costruire nuove relazioni per il Thesaurus.

Nella parte destra del pannello, denominata *New rule data input*, si possono selezionare tramite i menu a tendina le due classi di cui si vuole esprimere il legame, quindi si può scegliere il tipo di relazione estensionale che sussiste tra i due tra i due termini proposti dal menu a tendina, centrale: EQU, ISA.

Una volta stabilito se le due classi sono una la specializzazione dell' altra o se sono coincidenti, si può premere il tasto ADD che provoca l' inserimento della relazione estensionale e la sua conversione in relazione intensionale rispettivamente di tipo SYN o NT, relazione che viene poi inserita nel Thesaurus.

Per la descrizione teorica \ sistemistica si veda[Venuta, 2000].

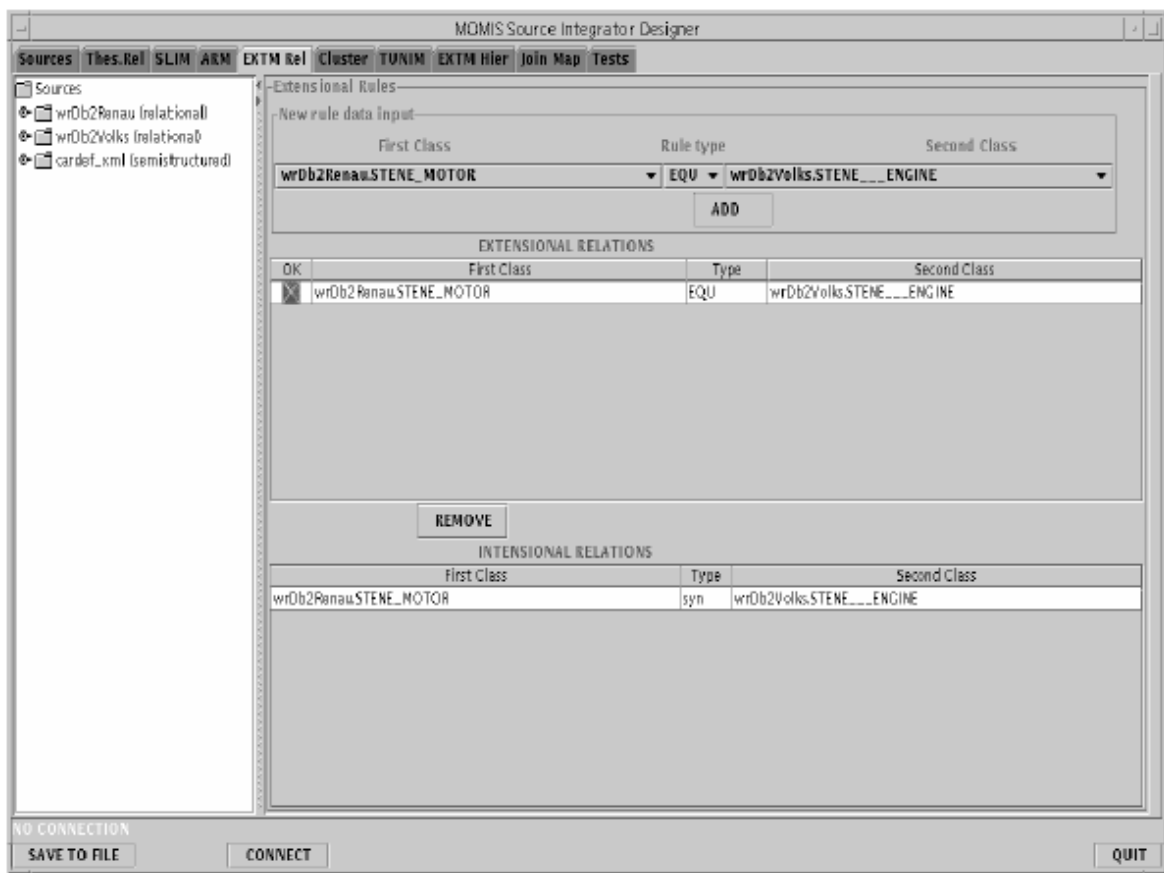


Figura 30: Pannello EXT M Rel

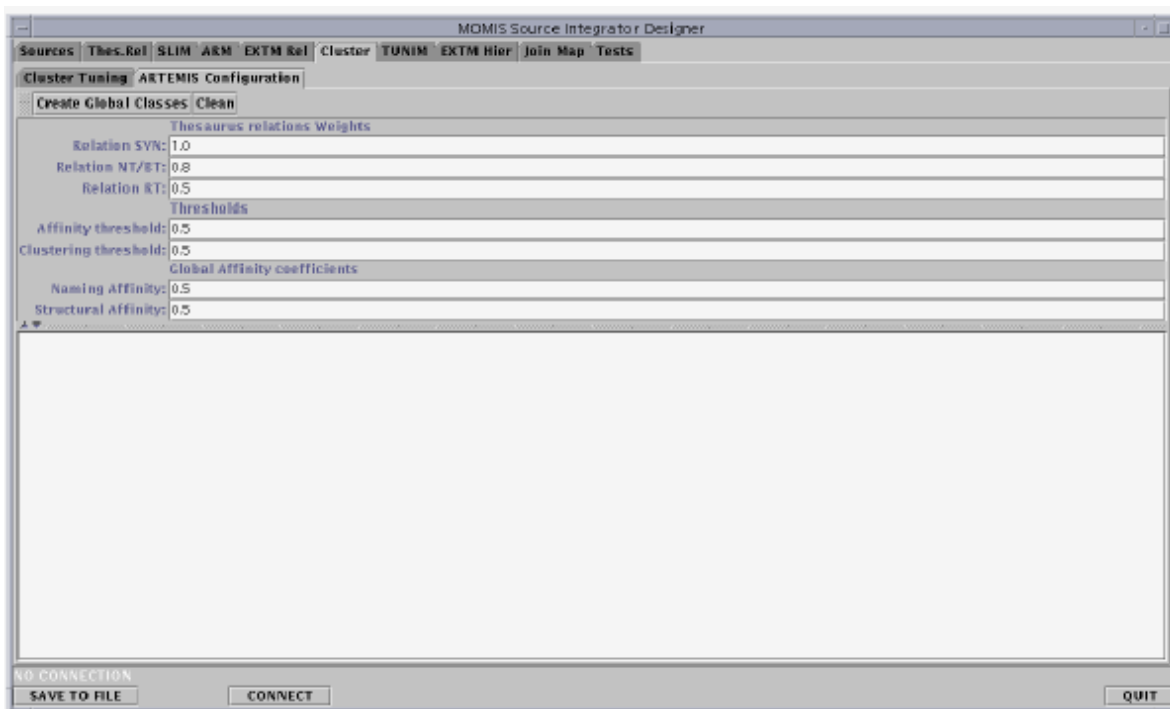
### 5.10.7 Cluster

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*, visibile in Figura 31, si configura il modulo ARTEMIS settando:

- i pesi delle relazioni del Thesuarus;
- i valori di soglia per le affinità e il clustering;
- i coefficienti per l'affinità globale.



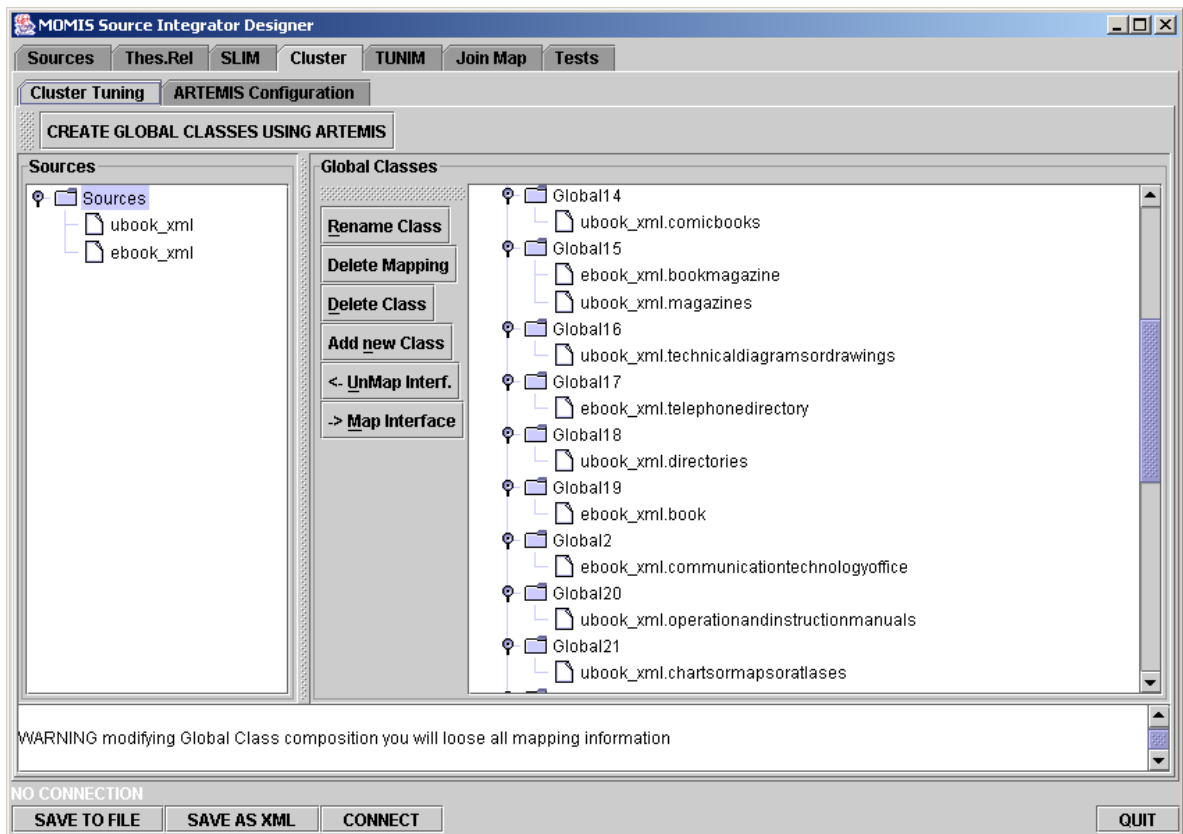
**Figura 31: Pannello Cluster: ARTEMIS Configuration**

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 32) invocando il modulo per mezzo del pulsante CREATE GLOBAL CLASSES USING ARTEMIS.

Il risultato viene presentato nella finestra di sinistra. 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 una 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.



**Figura 32: Pannello Cluster: Cluster Tuning**

## 5.10.8 TUNIM

Il pannello TUNIM, che possiamo osservare in Figura 33, è quello che permette la costruzione degli schemi globali dopo che le classi locali sono state raggruppate in cluster dal modulo ARTEMIS.

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.

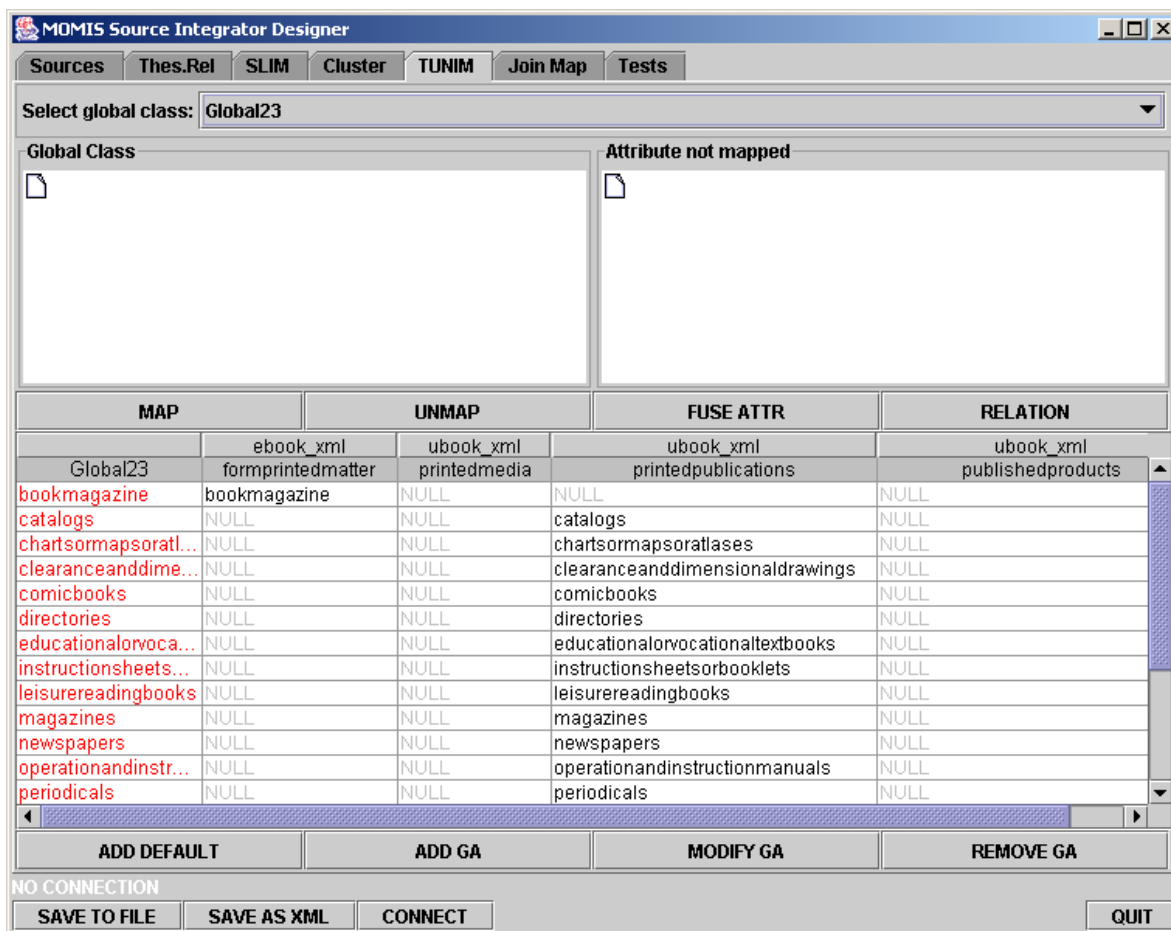


Figura 33: Pannello TUNIM

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 degli standard di classificazione dei prodotti con MOMIS**

### **6.1 Generale**

L'obiettivo di questa tesi è analizzare la validità del sistema MOMIS nell'integrazione degli standard di classificazione dei prodotti. Questo studio si colloca in un contesto di ricerca molto ampio, infatti, l'integrazione di diverse codifiche è diventata un problema cruciale per lo sviluppo dell'e-commerce. Ci sono due possibili strategie per l'ottenimento dell'interoperabilità tra i sistemi:

- l'adozione di un unico standard di classificazione dei prodotti, che assumerebbe il ruolo di una Lingua Franca per agevolare gli scambi;
- l'individuazione di un mapping tra i diversi sistemi di classificazione dei prodotti che permetterebbe ad ogni utente di continuare ad utilizzare la codifica che ha scelto ma allo stesso tempo di comunicare con altri utenti.

MOMIS permette di adottare entrambe le strategie, in quanto colloca i codici corrispondenti in un'unica classe globale. Per esempio nell'integrazione tra i codici relativi all'argomento PEN (in sezione 7.4.2) è stata individuata una classe globale:

### **Global Class 10:**

- Eclass Communication technology – office;
- UNSPSC Office equipment and accessories;

Questa classe globale può essere interpretata come mapping tra i codici che la compongono, ma può anche essere considerata come la base per la creazione di una Lingua Franca. Si potrebbe, infatti, creare un “nuovo standard” partendo dalle classi globali individuate e assegnando loro un nome significativo. In queste tesi non si prenderà in considerazione tale opportunità, in quanto, a causa del tempo che richiede e della continua evoluzione dei prodotti, non viene ritenuta vantaggiosa [Schulten, 2002]. Nei paragrafi successivi si analizzeranno più dettagliatamente i problemi imposti dall'utilizzo di diverse codifiche, le varie soluzioni proposte fino ad ora e le possibilità offerte dal sistema MOMIS in questo ambito.

## **6.2 Lo stato dell'arte**

Come ampiamente evidenziato nei capitoli precedenti, la codifica dei prodotti e dei servizi in accordo con sistemi di classificazione standardizzati è utile per velocizzare il commercio tra le compagnie. Inoltre, lo sviluppo di soluzioni per l'e-commerce, e in particolare per gli e-marketplace, ha fatto crescere notevolmente la richiesta di nomi di prodotti leggibili alla macchina e che agevolino le funzioni di marketing e di vendita nel trovare clienti e nel fornire servizi migliori.

Un marketplace è un luogo dove la domanda e l'offerta di compratori e venditori partecipanti ad un processo di business si incontrano. I marketplace sono una soluzione interessante, in quanto mostrano prodotti commercializzati da vari venditori e danno la possibilità di effettuare confronti tra prodotti affini.

Inserendo i codici in vari documenti elettronici per il commercio, come i cataloghi di prodotto, i siti Web, gli ordini di acquisto e le fatture, si possono processare le transazioni automaticamente. Sorgono, però, delle difficoltà dovute al fatto che i compratori devono utilizzare un solo standard per classificare e descrivere prodotti forniti da vari venditori, in modo da assicurarsi una facile integrazione con il sistema ERP (Enterprise Resource Planning). Per questo motivo il marketplace deve fornire uno strumento per *mediare* tra i differenti standard utilizzati dai partecipanti alla

transazione, in modo che i diversi attori del processo di business possano scambiare informazioni utilizzando i loro formati di dati (standard).

Negli ultimi anni molti gruppi di ricerca si sono concentrati su queste tematiche. Come anticipato, questa tesi ha preso spunto principalmente da un articolo pubblicato sul sito [www.ontoweb.org](http://www.ontoweb.org) e da un successivo studio svolto nell'ambito del nostro Ateneo [Bergamaschi, 2002]. Tale studio suggerisce l'uso di una metodologia semiautomatica sviluppata all'interno di MOMIS, per definire il mapping tra i diversi standard di classificazione dei prodotti.

### **6.2.1 La ricerca di una Lingua Franca o della Stele di Rosetta**

La dura realtà dell'e-business e la lenta adozione degli acquisti via Internet sono da attribuire alla complessità della sostituzione delle interazioni umane con il computer. Certamente questo non è difficile da comprendere. Il dialogo umano è strutturato in base ad una semantica, ad una sintassi e ad una grammatica che si collocano in un contesto di convenzioni sociali e culturali. L'e-commerce è privo di tale background. Ciò vale specialmente per il business tra le imprese, dove si devono prendere in considerazione l'interoperabilità tra diversi sistemi, l'integrazione con il sistema di back-end e i complessi processi di business. Nonostante la carenza di appropriati standard di linguaggio, molte imprese hanno iniziato a popolare il Web in alcune Community, formando un Intranet, un Extranet o un marketplace. Il multilinguismo non è un problema in sé, spesso apporta creatività e una costruttiva evoluzione. Ma quando le traduzioni sono assenti la situazione diventa critica. Questo è esattamente il caso del B2B e-commerce. Si potrebbe paragonare questa situazione con la costruzione della torre di Babele dove ognuno apporta il proprio contributo. I risultati sono leggendari.

Le industrie e le comunità di ricerca, attualmente, impiegano risorse per iniziative no-profit di standardizzazione e di armonizzazione. L'analogia con la torre di Babele ricorre spesso nelle metafore usate nell'ambito di queste iniziative. Esempi sono la Lingua Franca (riferita all'XML) o la stele di Rosetta (da cui prende il nome il consorzio Rosettanet). Lingua Franca è un linguaggio commerciale utilizzato all'inizio del diciottesimo secolo da numerose comunità del Mediterraneo per rendere possibili gli

scambi. La stele di Rosetta è un bassorilievo scritto in tre lingue che ha dato agli studiosi la possibilità di intuire la traduzione tra i linguaggi. Le due strade che si possono percorrere per arrivare all'interoperabilità richiamano questi esempi: da un lato si può definire un nuovo standard dall'altro si può costruire una piattaforma che fornisca un mapping tra i differenti standard, o che, in modo più realistico, individui un percorso che li connette.

La necessità di consenso in una comunità di business sorge a differenti livelli. Questo si riflette in diverse aree di focalizzazione delle iniziative di armonizzazione. Un massiccio impegno è dedicato alla ricerca di un consenso sui processi e sui documenti di business. Ma in paragone ci sono poche iniziative che si focalizzano sull'armonizzazione, apparentemente semplice, delle fondamenta di ogni transazione di business: la descrizione dei prodotti. Si possono associare due funzioni agli standard di descrizione dei prodotti:

- limitare la ricerca del cliente alla gamma di prodotti utilizzati;
- fornire al cliente una descrizione di prodotto che si adatti alle sue esigenze.

Nella situazione ideale, tutto il commercio elettronico tra le aziende si svolgerebbe utilizzando un unico standard di classificazione dei prodotti. Ma ci sono due ragioni per cui questa non sembra un'opzione possibile. La prima è che i prodotti cambiano in continuazione, quindi, la classificazione sarebbe in continua evoluzione. La seconda è che le aziende non possono aspettare un decennio perché questo standard sia pronto. L'attuale stato dell'arte si traduce in un dizionario dei prodotti particolarmente user-unfriendly e difficile da mantenere che usa diversi indici non legati che contengono sezioni sovrapposte senza appropriati riferimenti incrociati. Ciò può rallentare fortemente lo sviluppo e l'adozione di soluzioni di e-commerce.

L'unica possibilità per mettere in comunicazione utenti che utilizzano standard differenti è quella di creare un mapping.

L'obiettivo del mapping è fare sì che un potenziale compratore che accede ad un marketplace e inserisce il codice del prodotto cercato, ottenga come risultato della ricerca tutti i prodotti del tipo desiderato, anche se alcuni venditori li hanno codificati in accordo con differenti standard di classificazione. Praticamente, se un compratore ha adottato la codifica [ecl@ss](#) e cerca all'interno di un marketplace un libro inserirà il codice: 24.15.03.01, la ricerca deve fornire, come risultato, anche tutti quei libri in

vendita che sono classificati in accordo con un'altra codifica. Grazie al mapping esistente tra le codifiche si dovrebbero individuare anche i libri commercializzati da fornitori che utilizzano la codifica UNSPSC e che di conseguenza hanno codice 55.10.15.XX.

## **6.2.2 L'obiettivo: arrivare ad un ontologia comune tra i diversi sistemi di classificazione**

Un gruppo di ricercatori che fa riferimento a OntoWeb<sup>29</sup> ha proposto ai gruppi di ricerca di partecipare ad una sfida relativa alla classificazione dei prodotti per il B2B e-commerce<sup>30</sup>. L'obiettivo dei partecipanti a questo challenge è creare un modello (ontologia, metodo di risoluzione dei problemi, una soluzione di lavoro implementata) per arrivare ad un metodo computazionalmente effettivo, utile e basato su principi teorici per descrivere il dominio dei prodotti e per stabilire un mapping semi-automatico tra i diversi sistemi di classificazione dei prodotti.

I progressi della ricerca forniranno importanti input per la futura modellazione di standard di classificazione dei prodotti.

Le soluzioni al problema dovrebbero mostrare solo come risolvere casi di classificazioni multiple in e-commerce in modo generico.

Le attività richieste da questo challenge sono:

**Progettare un modello generico (ontologia, metodo, implementazione, risultati computazionali sperimentali) per arrivare ad un mapping automatico o semiautomatico tra Ecl@ss e UNSPSC, visti come esempi di standard di classificazione dei prodotti.**

Le proposte inviate saranno analizzate da un comitato formato dai membri di OntoWeb, in accordo con i seguenti criteri:

---

<sup>29</sup> OntoWeb: The UE Thematic Network on Ontology-Based Information Exchange for Knowledge Management and Electronic Commerce.

<sup>30</sup> "unique B2B E-Commerce Product Classification Challenge".

1. La proposta deve portare ad una soluzione operativa che sia concettualmente e computazionalmente adeguata al caso (schemi di [ecl@ss](#) e di UNSPSC; frammenti come quello utilizzato nell'esempio citato nelle prossime pagine). Sono importanti la adeguatezza epistemologica, la complessità, la precisione raggiunta nel mapping e l'approccio alla classificazione.
2. Le soluzioni proposte devono avere un valore generico, devono, infatti, mostrare come risolvere una situazione di questo tipo legata all'e-commerce.
3. Se si ritiene opportuno si possono inserire note e suggerimenti per il futuro sviluppo di standard di classificazione dei prodotti.
4. Si possono fornire anche raccomandazioni o richieste per i linguaggi del Semantic Web e per gli standard.
5. Tutte le assunzioni fatte prima di svolgere il lavoro e la lezione appresa devono essere esplicitate.

### **6.2.3 Esempio: Mapping tra un frammento di [Ecl@ss](#) e UNSPSC**

Lo staff di OntoWeb propone anche un esempio di integrazione svolto su due frammenti di [ecl@ss](#) e UNSPSC relativi ai "Paper product".

UNSPSC

Paper: 14.11.15.11

14 Paper Material and Products

11 Paper Products

15 Printing and Writing Paper

11 Writing Paper



[Ecl@ss](mailto:Ecl@ss)

Paper: 24.11.05.34

24 Communication Technology

11 Office Supplies

05 Office Supplies (other)

34 Writing Paper

I connettori in Figura 34 mostrano i mapping creati, alcuni sottendono una corrispondenza 1:1, altri una corrispondenza di tipo 1:N. Si nota, per esempio, che [ecl@ss](mailto:Ecl@ss) printer paper ed [ecl@ss](mailto:Ecl@ss) writing paper vengono mappati con UNSPSC writing paper.

Si vede immediatamente che i domini considerati non coincidono con rami delle codifiche, ma sono stati selezionati facendo attenzione ad individuare dei match. In questo esempio è stata tralasciata la codifica NAICS, che, come si illustrerà approfonditamente in seguito, presenta molte differenze concettuali con i due standard in esame.

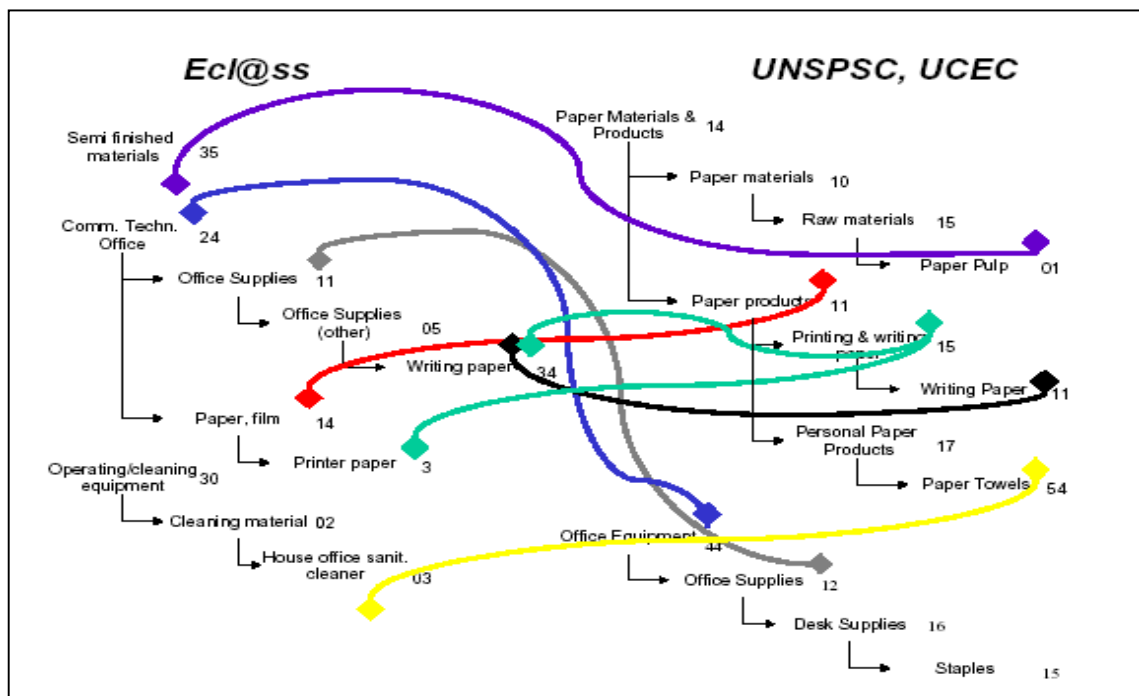


Figura 34: Mapping tra le codifiche

## 6.2.4 La riconciliazione degli standard mediante il sistema MOMIS

In [Bergamaschi,2002] sono state presentate alcune idee preliminari per un'integrazione semiautomatica degli standard di classificazione dei prodotti tramite il sistema MOMIS. In particolare, è stato dimostrato che MOMIS, al livello dei metadati<sup>31</sup> (per esempio gli schemi coinvolti nel processo di integrazione che descrivono i due frammenti di codice), è in grado di svolgere il processo di mapping tra i due standard in modo semiautomatico. Nelle Figure 35 e 36 sono riportati i frammenti di codice integrati, mentre nella Figura 37 è mostrato il mapping creato da MOMIS tra loro.

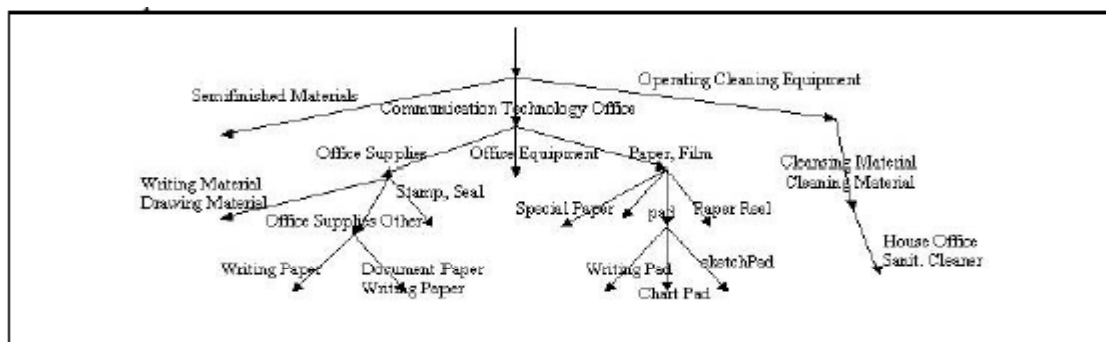


Figura 35: [Ecl@ss](#)- Paper Product

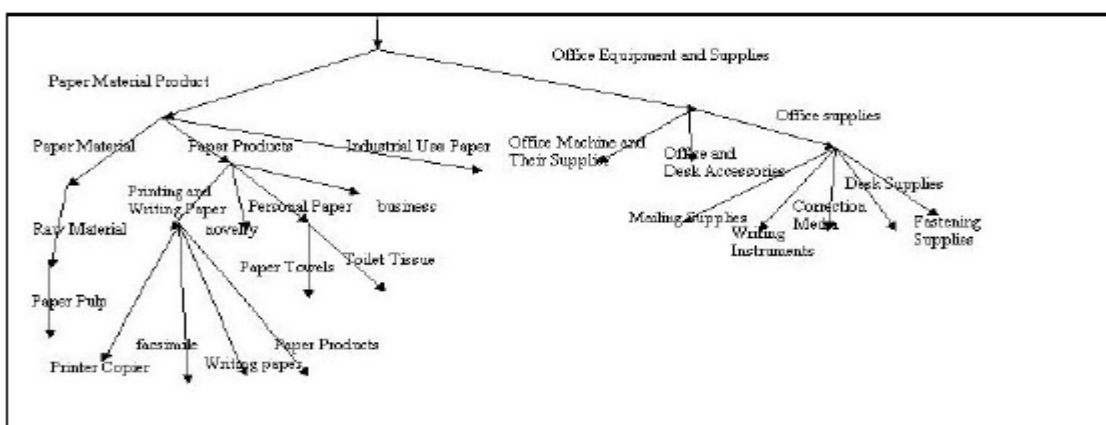


Figura 36: UNSPSC- Paper Product

<sup>31</sup> Metadati: informazione descrittiva relativa ai dati di una risorsa, compresi il dominio, la proprietà, le restrizioni, il modello di dati.

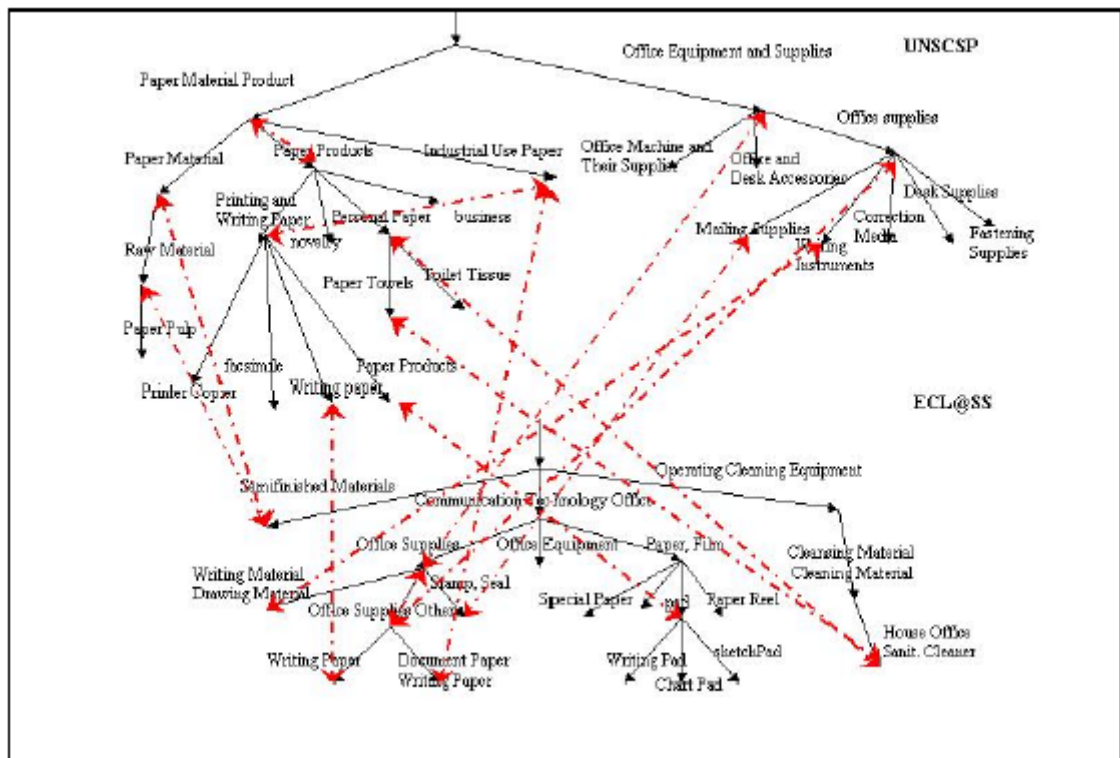


Figura 37: Integrazione Paper Product con MOMIS

## 6.3 Analisi dell'approccio MOMIS

Partendo dalle idee e dai risultati preliminari ottenuti in [Bergamaschi, 2002], la presente tesi ha lo scopo di estendere e approfondire questi concetti, valutando l'efficacia di un sistema semiautomatico a mediatore, quale MOMIS, nelle attività di integrazione degli standard di classificazione dei prodotti.

### 6.3.1 Acquisizione dei codici

Il primo passo verso l'individuazione del mapping tra gli standard di classificazione è l'acquisizione delle codifiche.

[Ecl@ss](http://www.ecl@ss.de) è disponibile on-line e si può scaricare dal sito [www.ecl@ss.de](http://www.ecl@ss.de), si compone di una gamma di fogli di calcolo che opportunamente collegati tra loro ricostruiscono la complessa struttura del codice. A tal fine si può costruire un database che permette una

gestione più agevole e integrata dei codici. Sul sito in questione si possono trovare anche pubblicazioni relative alla classificazioni e informazioni sulla sua storia.

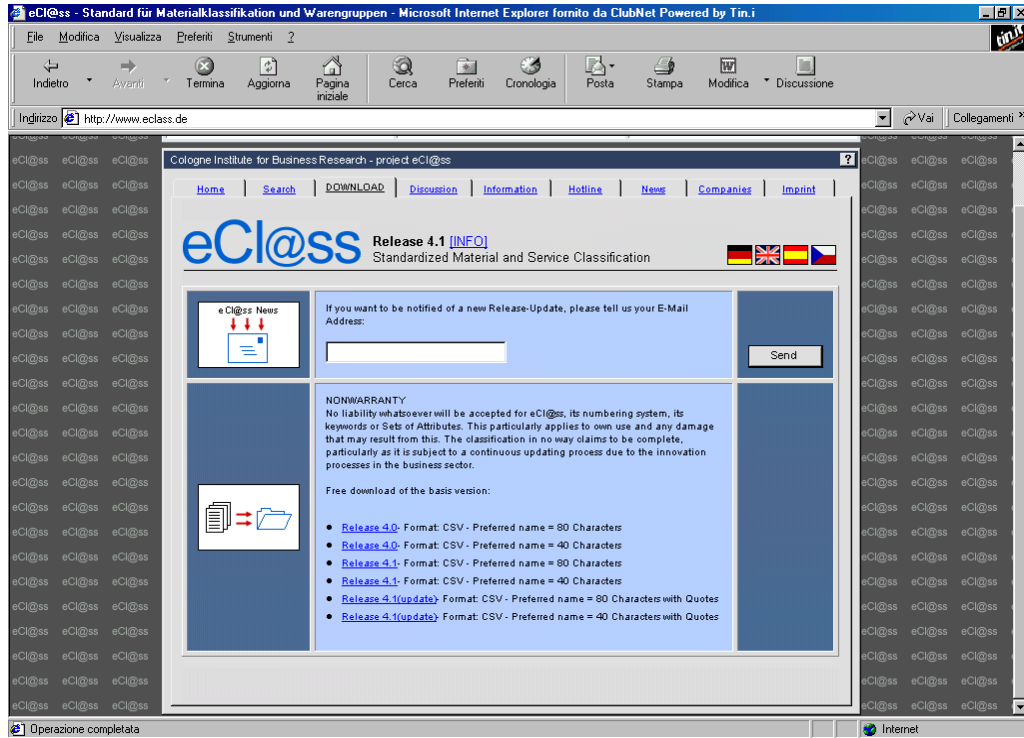


Figura 38: Download di [ecl@ss](http://www.eclss.de) dal sito Internet

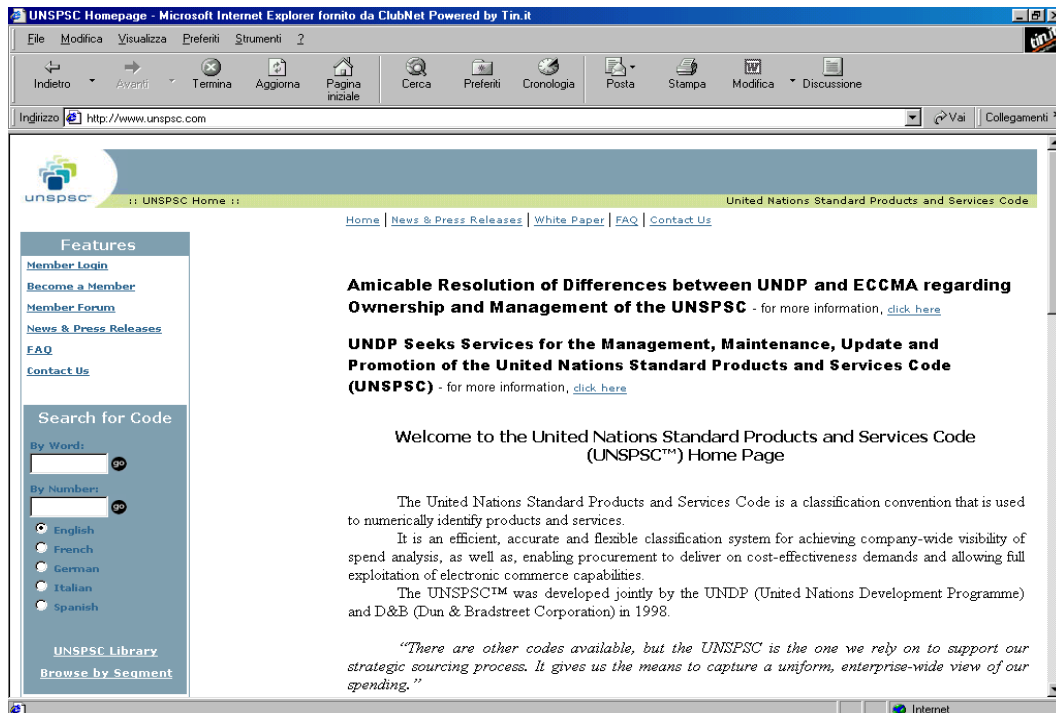


Figura 39: Sito di UNSPSC

UNSPSC è disponibile on-line all'indirizzo [www.unspsc.com](http://www.unspsc.com)<sup>32</sup> in questo sito viene fornita l'ultima versione del codice con il relativo file di audit che mostra le correzioni apportate alla versione precedente. Nel sito citato la codifica è presente in Inglese, Francese, Tedesco, Italiano e Spagnolo. I file sono forniti anche in questo caso in formato Excel, l'unica differenza rispetto ai file di [ecl@ss](mailto:ecl@ss) è che la codifica è raccolta in un unico foglio di calcolo.

Le informazioni relative a NAICS si possono trovare nel sito ufficiale [www.ntis.gov/products/naics](http://www.ntis.gov/products/naics), sono disponibili informazioni molto generiche e riferimenti a testi e articoli che si possono acquistare per avere maggiori informazioni. La codifica non può essere scaricata, la versione disponibile per gli utenti è a pagamento. La soluzione adottata è stata dedurre la classificazione NAICS da tabelle che effettuano il confronto codice per codice tra NAICS e SIC, reperibili nel sito [www.census.gov](http://www.census.gov). Con questo metodo, attraverso alcune semplici operazioni, si è ottenuta la codifica NAICS in formato Excel.

**U.S. Census Bureau**  
**1997 NAICS and 1987 SIC Correspondence Tables**

How to navigate these tables  
 How to interpret these tables  
 Download files in database or word processor format

Enter a search keyword:

Table 1: 1997 NAICS Matched to 1987 SIC		Table 2: 1987 SIC matched to 1997 NAICS	
11	<a href="#">Agriculture, Forestry, Fishing, and Hunting</a>		<a href="#">Agriculture, Forestry, and Fisheries</a>
21	<a href="#">Mining</a>		<a href="#">Mineral Industries</a>
22	<a href="#">Utilities</a>		<a href="#">Construction Industries</a>
23	<a href="#">Construction</a>		<a href="#">Manufacturing</a>
31-33	<a href="#">Manufacturing</a>		<a href="#">Transportation, Communications, and Utilities</a>
42	<a href="#">Wholesale Trade</a>		<a href="#">Wholesale Trade</a>
44-45	<a href="#">Retail Trade</a>		<a href="#">Retail Trade</a>
48-49	<a href="#">Transportation and Warehousing</a>		<a href="#">Finance, Insurance, and Real Estate</a>
51	<a href="#">Information</a>		<a href="#">Service Industries</a>
52	<a href="#">Finance and Insurance</a>		<a href="#">Public Administration</a>
53	<a href="#">Real Estate and Rental and Leasing</a>		
54	<a href="#">Professional, Scientific and Technical Services</a>		
55	<a href="#">Management of Companies and Enterprises</a>		
56	<a href="#">Administrative and Support and Waste Management and Remediation</a>		

Click [HERE](#) to view Tables 1 and 2 together in a frame

Figura 40: Tabella delle corrispondenze tra NAICS e SIC

<sup>32</sup> Si deve notare che questo sito non è quello indicato da OntoWeb in relazione alla codifica UNSPSC, ma dall'Ottobre 2002 tale codifica non è più sotto il controllo di ECCMA. È stato emesso un accordo tra UNDP ed ECCMA che dichiara che dal 2 Ottobre 2002 la codifica UNSPSC è gestita e distribuita da UNDP.

### 6.3.1.1 Motori di ricerca

La ricerca del codice di uno determinato prodotto può essere effettuata per mezzo di specifici motori di ricerca. Nel sito ufficiale di [ecl@ss](http://ecl@ss) si trova un sistema di ricerca che permette di:

- individuare il prodotto a cui corrisponde un codice;
- individuare il codice di un prodotto;
- navigare agilmente attraverso la gerarchia.

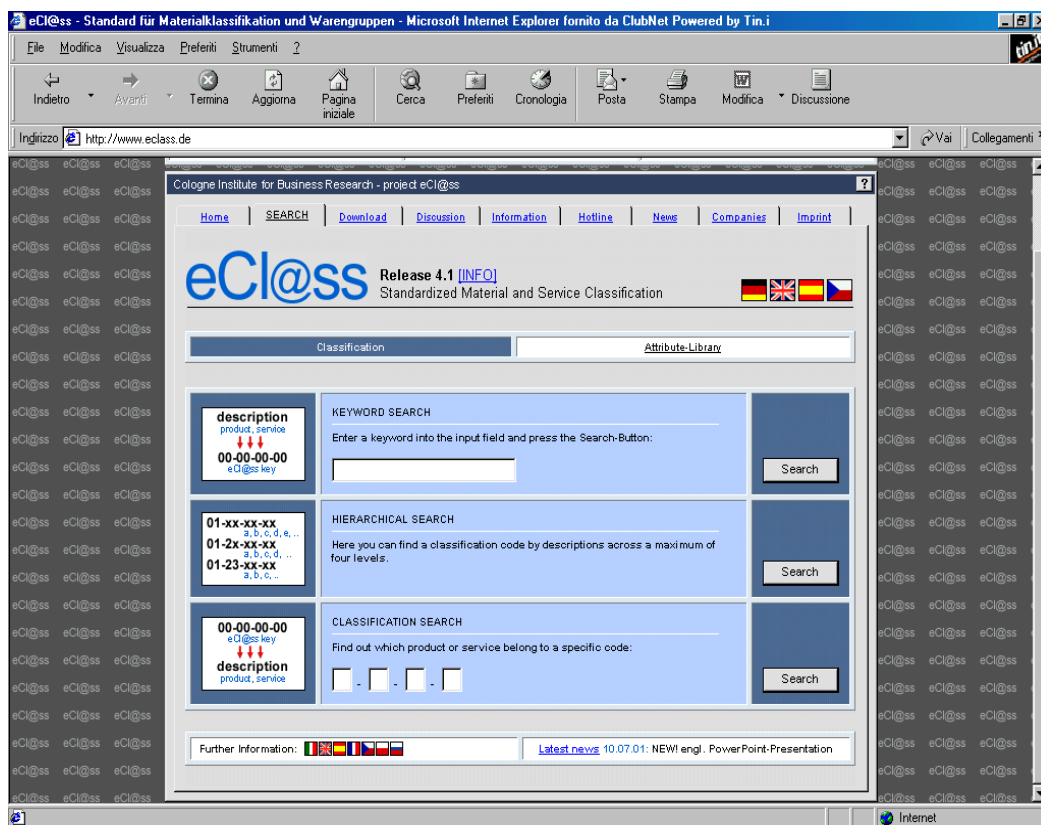


Figura 41: Motore di ricerca su ecl@ss

Per svolgere ricerche in UNSPSC si può utilizzare il motore di ricerca messo a disposizione dal sito [www.unspsc.com](http://www.unspsc.com), per mezzo del quale si possono cercare codici a partendo da una parola chiave oppure si può trovare la descrizione associata ad un codice noto. Anche il sito [www.eccma.com](http://www.eccma.com) “teoricamente” dispone di un motore per le ricerche all’interno della codifica ma non è funzionante. Per ricercare un codice all’interno della codifica NAICS ci si può servire del motore di ricerca messo a disposizione dal sito [www.ntis.gov/products/naics](http://www.ntis.gov/products/naics), che permette di effettuare ricerche mediante l’inserimento di parole chiave.

### **6.3.2 Selezione dei frammenti di codice**

Il primo passo verso l'integrazione è l'individuazione di frammenti, cioè porzioni delle codifiche che possono coincidere con un ramo della gerarchia oppure comprendere solo i primi livelli del ramo stesso. Tali frammenti devono essere compatibili per l'integrazione, cioè devono presentare il maggior numero possibile di match tra i codici che includono. La selezione dei frammenti di codice è un passo inevitabile, in quanto non è possibile inserire tutti i codici contemporaneamente, poiché ciò comporterebbe la gestione di una mole di dati eccessiva. Ne conseguirebbero dei tempi di processamento del sistema molto ingenti e un carico di lavoro troppo elevato per il progettista.

La selezione dei frammenti, che all'apparenza è un'operazione banale, è molto complicata a causa delle profonde differenze che sussistono tra i sistemi di classificazione. Ogni sistema di classificazione ha un metodo diverso per suddividere i prodotti tra i vari rami della gerarchia e assegna ad un dominio una quantità di codici proporzionale all'importanza che associa al dominio stesso o alle richieste dei principali fruitori della codifica. Capita spesso di individuare frammenti di dimensioni molto diverse e con diversi livelli di approfondimento associati ad uno stesso dominio. L'operazione di selezione ha molta influenza sui risultati del mapping, infatti, se i frammenti scelti non sono effettivamente compatibili non è possibile ottenere dei buoni risultati in fase di integrazione.

Ci possono essere fondamentalmente due punti di partenza: si può scegliere un particolare prodotto e cercarlo in entrambi i sistemi di classificazione, oppure, si possono scorrere le codifiche cercando argomenti in cui risultino simili, per poi procedere all'integrazione dei frammenti corrispondenti. Il primo approccio si può definire di tipo Bottom-up, il secondo di tipo Top-down.

### 6.3.2.1 Approccio Top-Down

Questo approccio porta alla costruzione di grandi frammenti generici che prendono in considerazione un dominio piuttosto che uno specifico prodotto. Si basa sull'individuazione di corrispondenze tra le codifiche ai primi livelli della gerarchia. L'esplorazione dei domini risulta completa in quanto si considerano tutte le diramazioni derivanti dal nodo scelto, anche se non si scende in profondità.

Si esamina 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 cercano similitudini tra i diversi standard, nel caso vengano individuate, i frammenti si costruiscono scendendo di uno o due livelli a partire dal segmento \ settore scelto.

UNSPSC è costituito da 54 segmenti, [Ecl@ss](#) da 22 segmenti e NAICS da 20 settori, da questo dato s'intuisce una certa disuniformità tra i codici che rende difficile l'individuazione di corrispondenze 1:1 tra due segmenti o tra un segmento e un settore. La soluzione a questo problema sarebbe considerare anche le unioni, per esempio si potrebbe individuare un segmento di [Ecl@ss](#) che corrisponde a 2 o 3 segmenti di UNSPSC. Questo tentativo ha avuto poco successo poiché amplia enormemente lo spettro di prodotti considerati senza portare ad una vera corrispondenza.

S'intuisce che questo approccio deve essere usato solo quando c'è una corrispondenza evidente tra due segmenti, quindi non è uno strumento per ottenere un mapping completo delle codifiche, ma solo per svolgere prove di integrazione.

Esempi in cui si è utilizzato un criterio di scelta Top-down sono i frammenti di codice relativi a FARM e MINING, entrambi coinvolgono solo due delle tre codifiche esaminate. Negli esempi citati si è scesi di due livelli nella gerarchia, considerando un ampio spettro di prodotti. Non si scende ulteriormente per evitare di gestire una grande quantità di dati eterogenei e per evitare di avere termini troppo specifici, che sarebbero di difficile comprensione per un dizionario generico come WordNet.

Il principale difetto 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 difetto è che 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. Nel caso di FARM, ad esempio, si può verificare che in entrambi i frammenti si trovino termini come: crop, seed, tree e floricoltura.

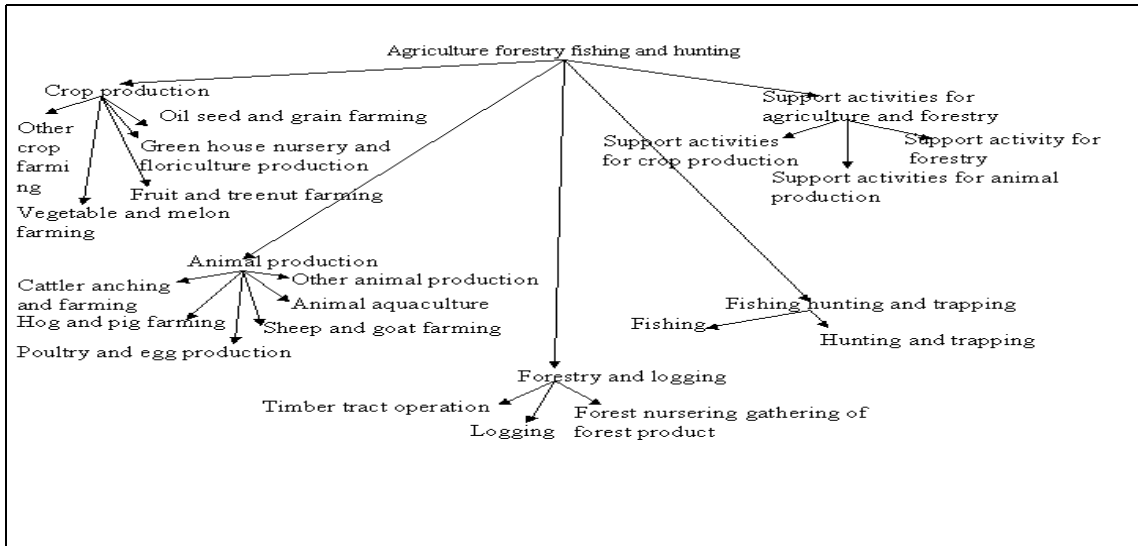


Figura 42: NAICS- Agriculture, forestry, fishing and hunting

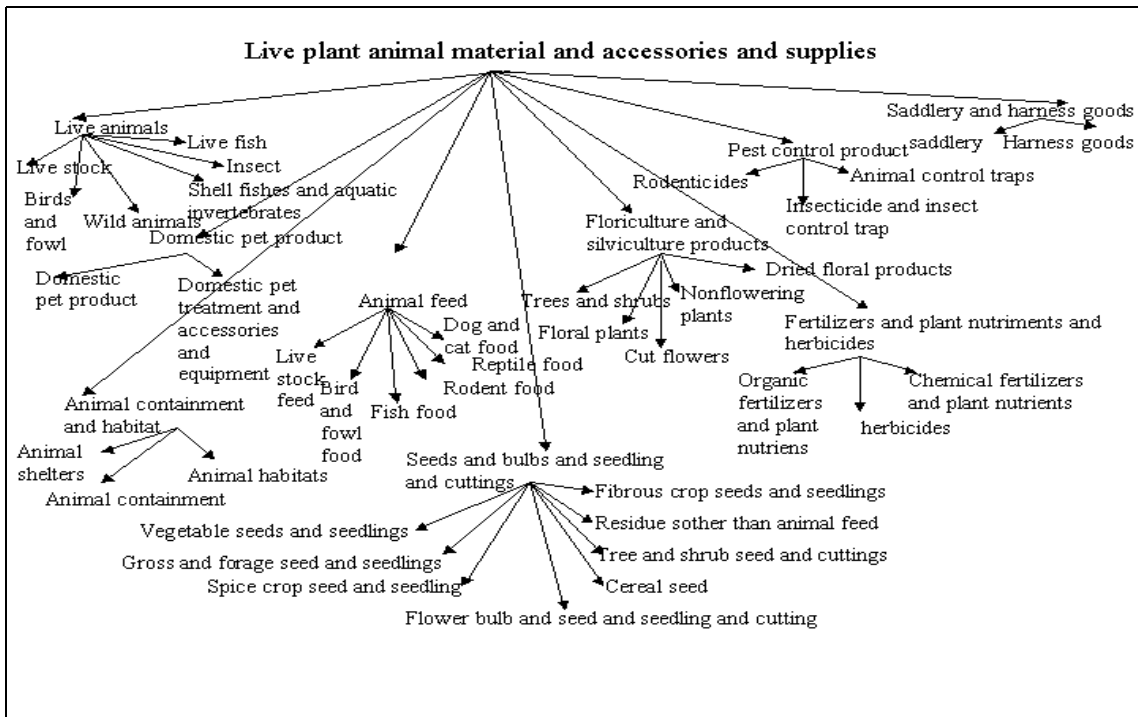


Figura 43: UNSPSC -Live plant, animal material and accessories and supplies

### 6.3.2.2 Approccio Bottom-up

Questo approccio porta alla costruzione di frammenti relativi ad un prodotto piuttosto che ad un dominio, l'idea di base è considerare una descrizione dettagliata di una ristretta gamma di prodotti.

Il punto di partenza è la scelta del prodotto da cercare all'interno delle codifiche, per esempio PEN (penna), in seguito si cercano i codici abbinati al prodotto. Nel caso delle penne si è immesso il termine PEN all'interno dei motori di ricerca specifici dei sistemi di classificazione, tali motori di ricerca hanno fornito, come risultato, i codici che all'interno della loro descrizione contengono il termine "pen". Tra i codici individuati si sceglie il più pertinente o quello che, essendo ad un livello superiore, comprende anche gli altri. Una volta individuato il codice 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.

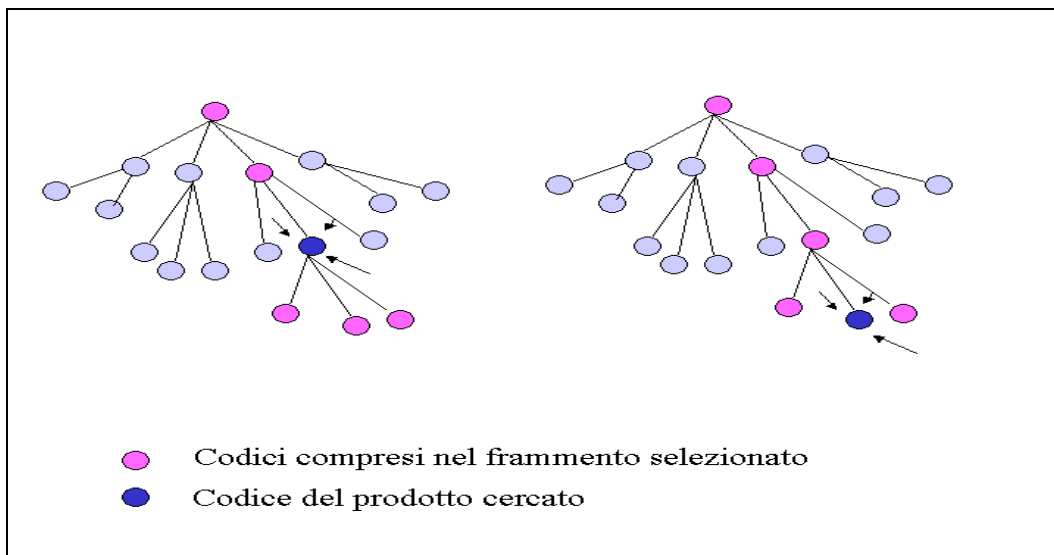
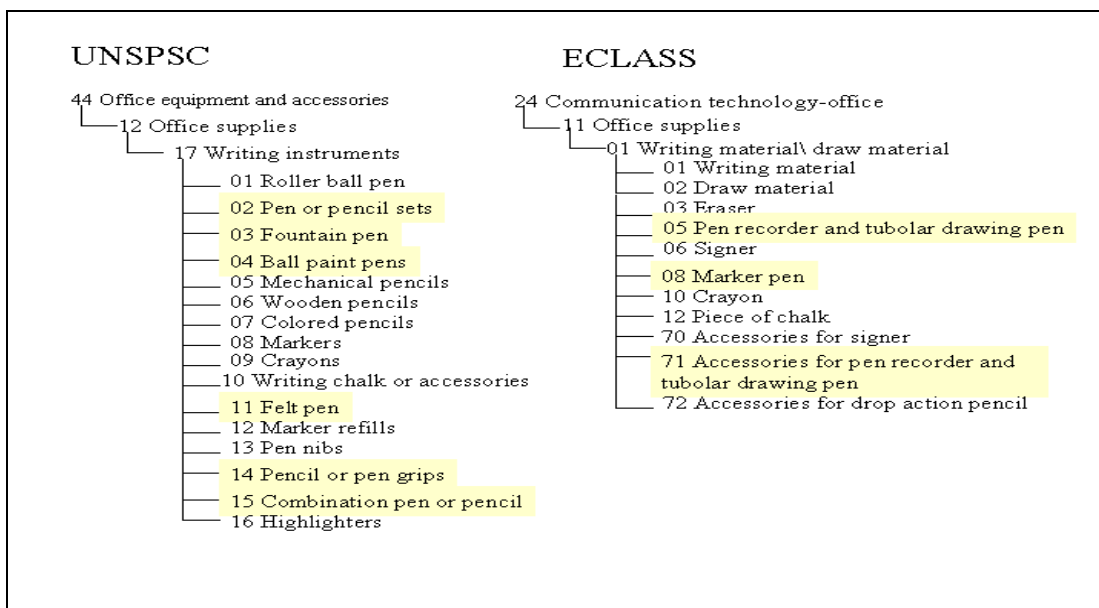
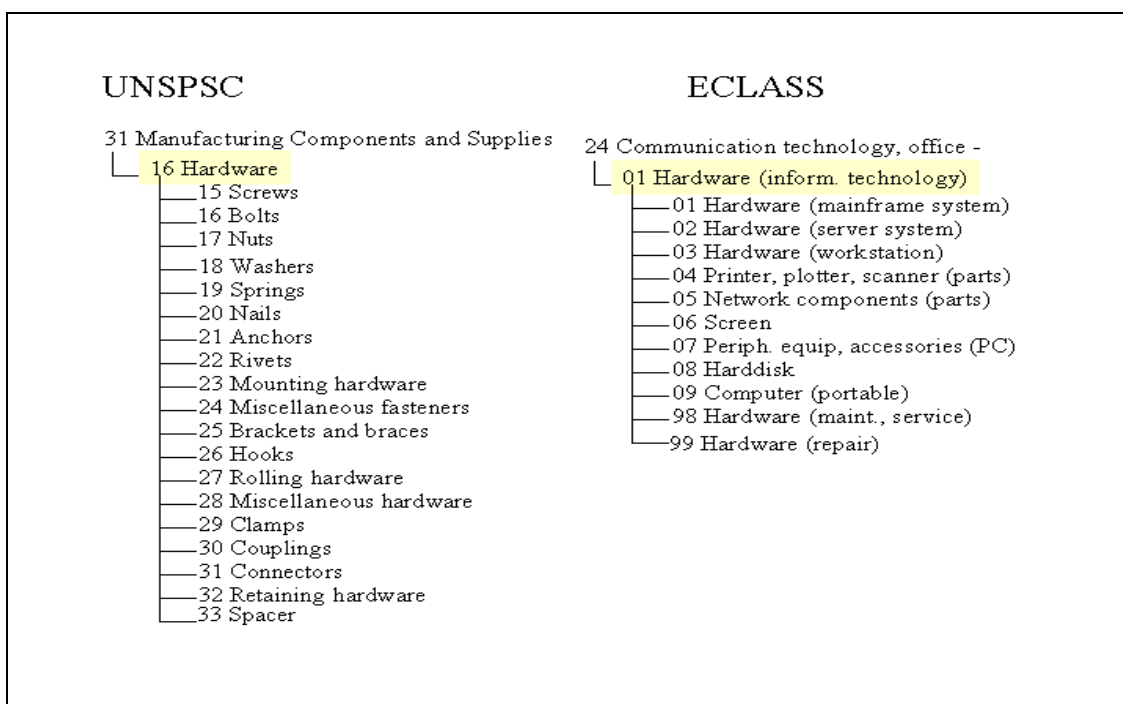


Figura 44: Individuazione dei frammenti di codice



**Figura 45: UNSPSC -ECL@SS : PEN**

È un approccio molto semplice ed immediato, infatti, da parte del progettista è sufficiente scegliere il prodotto e controllare che i risultati della ricerca siano pertinenti. Si deve, però, considerare che è impossibile mappare le intere codifiche procedendo con questo metodo. Un altro problema legato a questo approccio è che i sistemi di classificazione in alcuni casi assegnano significati diversi ad uno stesso vocabolo. Per esempio: se si ricerca il prodotto “HARDWARE”, [ecl@ss](#) classifica con tale definizione la componentistica del computer, mentre UNSPSC classifica la componentistica meccanica.



**Figura 46: UNSPSC- ECL@SS: HARDWARE**

Se ne deduce che seguendo il metodo citato sopra, senza controllare la descrizione dei codici selezionati, si corre il rischio di trovarsi con frammenti di codice assolutamente incompatibili.

### 6.3.2.3 Difficoltà incontrate nella selezione dei frammenti

La selezione dei frammenti si è rivelata un'operazione molto più difficoltosa del previsto, non tanto per l'individuazione di un ramo di codici appartenenti ad una codifica ma per l'individuazione di prodotti trattati da tutte le codifiche con un paragonabile livello di approfondimento. Le codifiche differiscono molto tra loro, quindi, capita, in molti casi, che il numero di codici che dedicano ad uno stesso prodotto sia molto diverso. Spesso uno stesso prodotto è posizionato a livelli diversi della gerarchia. Vediamo per esempio il caso dei frammenti selezionati con l'approccio Bottom-up cercando il prodotto TOY (giocattolo).

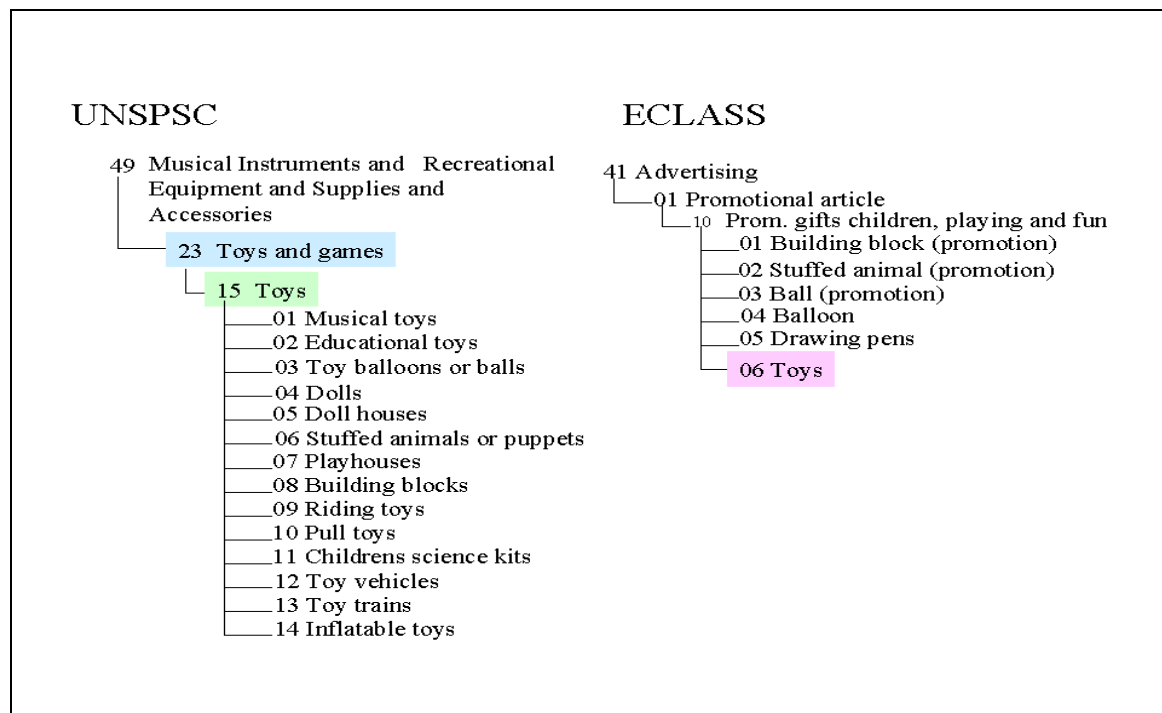


Figura 47: UNSPSC- ECL@SS: TOY

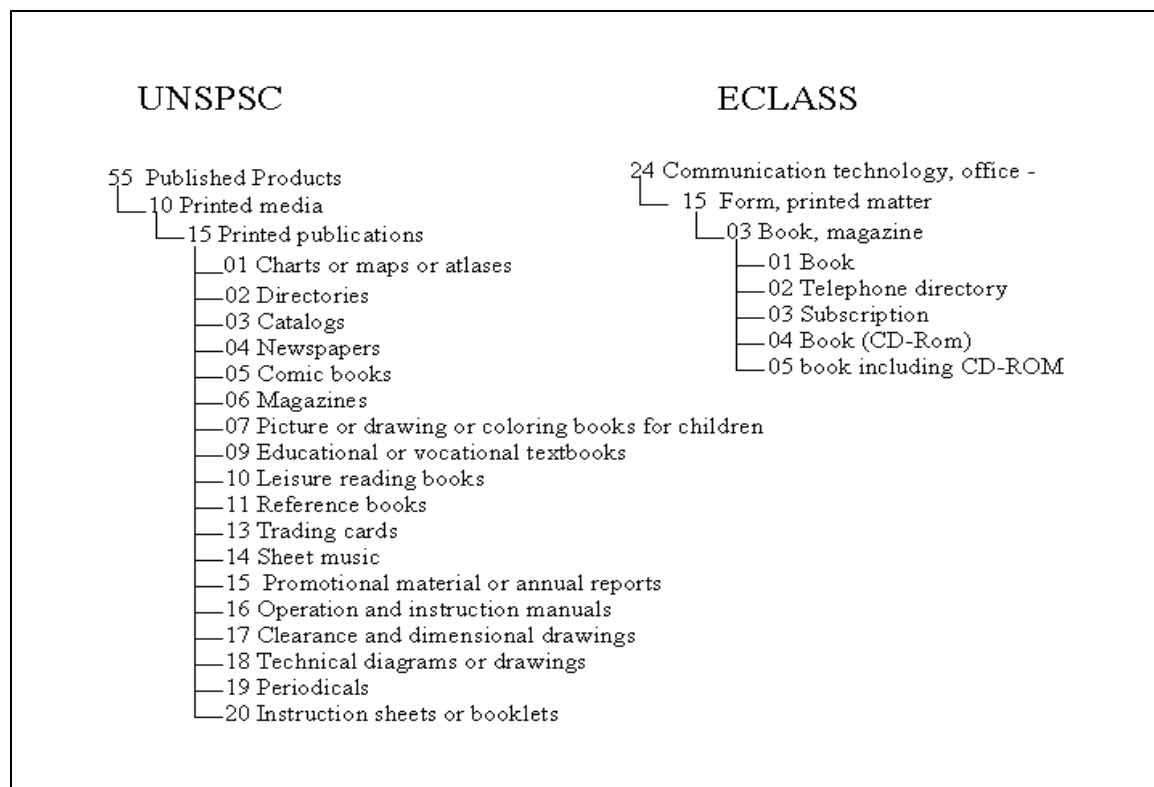
Le due codifiche hanno una struttura diversa: [ecl@ss](#) posiziona “Toy” all’ultimo livello, viene considerato un figlio di “Promotional gift children, playing and fun”, mentre UNSPSC classifica “Toy” al terzo livello e utilizza i codici del quarto per dettagliare ulteriormente il dominio. Se si considerano complessivamente i frammenti si nota che, nonostante le differenze strutturali, le due codifiche hanno molti codici corrispondenti.

Ci sono casi in cui i codici individuati ricercando lo stesso prodotto differiscono sostanzialmente in quanto posizionano lo stesso prodotto in domini molto diversi. Un esempio di questa situazione è dato dal dominio “shoes”, infatti, [ecl@ss](#) classifica le scarpe tra i prodotti per la sicurezza sul lavoro e la prevenzione degli incidenti, mentre UNSPSC le considera come parte dell’abbigliamento. Risulta chiara in questa circostanza la diversa focalizzazione delle codifiche: UNSPSC abbraccia l’intera gamma di prodotti e dei servizi, mentre [ecl@ss](#) considera solo i prodotti e servizi che hanno qualche attinenza con le attività manifatturiere. In particolare ci sono domini che vengono approfonditi molto di più da una codifica rispetto alle altre, [ecl@ss](#) approfondisce molto il dominio dei prodotti chimici, infatti, in origine era un sistema di classificazione riservato all’industria chimica. UNSPSC, come mostrato dall’esempio precedente, considera tutti i prodotti in commercio, dedicando ad ogni dominio una quantità di codici proporzionata all’influenza del dominio in questione sul mercato.

Gli standard di classificazione differiscono molto anche nella capacità di sfruttare le possibilità offerte dalla gerarchia. Se si ricerca all’interno delle codifiche la parola “Book” si nota che UNSPSC dedica 20 codici a questo argomento, dettagliandolo in modo pertinente ed ordinato, [Ecl@ss](#) utilizza soltanto 5 codici e nessuno di questi è in diretta corrispondenza con uno di quelli di UNSPSC. I metodi utilizzati dai due sistemi di classificazione per suddividere questo dominio di prodotti sono completamente diversi. UNSPSC distingue le pubblicazioni in giornali, cartine, periodici e varie tipologie di libri, mentre [Ecl@ss](#) si limita a distinguere i libri dall’elenco telefonico, dai CD-ROM e dai libri con allegato CD-ROM.

La scelta di frammenti di questo tipo porta inevitabilmente ad un mapping infruttuoso indipendentemente dalla validità dell’approccio di integrazione scelto.

Quando si sceglie di utilizzare la selezione Top-down dei frammenti si incontrano ostacoli diversi. Al primo livello le codifiche si compongono di 54 segmenti nel caso di UNSPSC, di 22 e 20 rispettivamente nel caso di [ecl@ss](#) e UNSPSC.

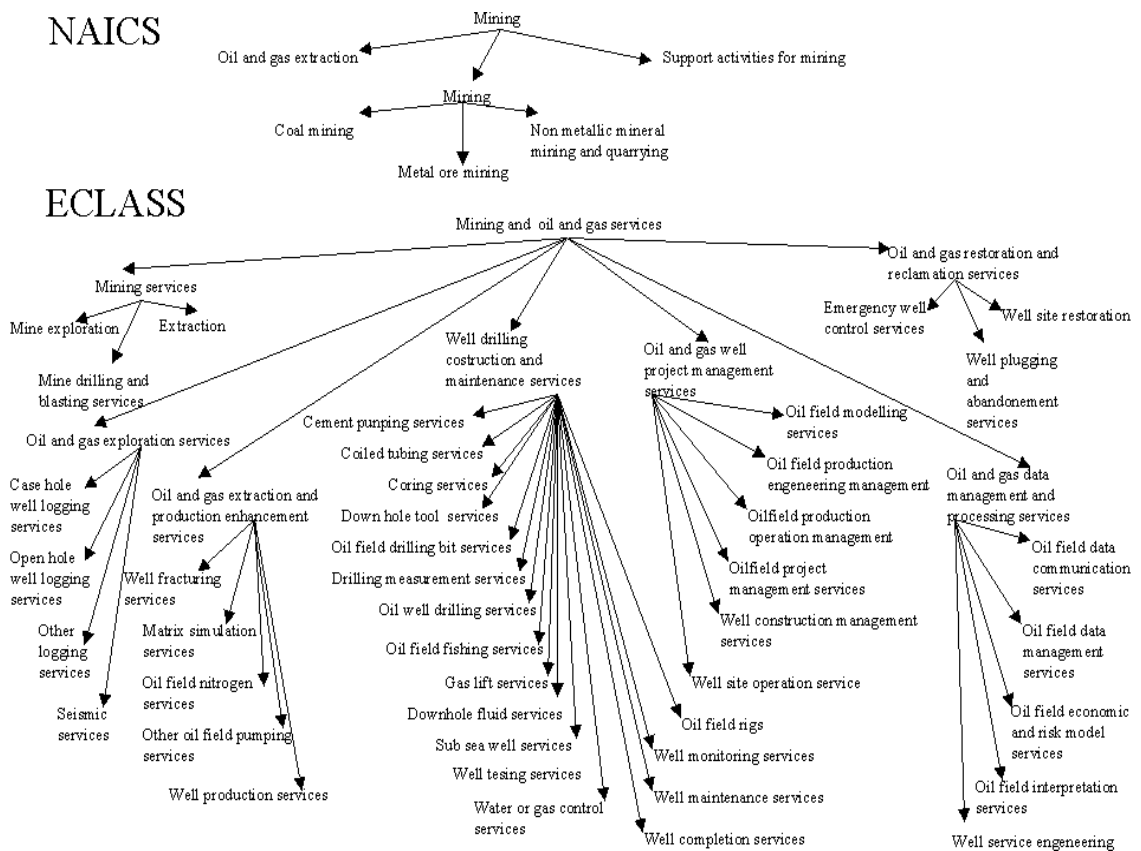


**Figura 48: UNSPSC- ECL@SS: BOOK**

La prima osservazione da fare è che già al livello dei segmenti le codifiche presentano un grado di approfondimento molto diverso. Individuare a prima vista una corrispondenza tra i segmenti è molto difficile e in alcuni casi in cui questa corrispondenza si intuisce immediatamente le codifiche si differenziano talmente tanto al secondo livello da rendere difficoltosa l'integrazione.

Un esempio sono i segmenti relativi all'argomento "Mining" appartenenti a NAICS ed [ecl@ss](#).

Al terzo livello [ecl@ss](#) si compone di oltre 40 codici, NAICS di 5, la soluzione per ottenere frammenti facilmente integrabili sarebbe scendere di un ulteriore livello nella gerarchia di NAICS, ma si è constatato che a tale livello si differenzia molto da [ecl@ss](#), approfondendo in modo molto diverso il dominio.



**Figura 49: NAICS- ECL@SS: MINING**

Una considerazione molto importante ai fini dell'integrazione è che [ecl@ss](#) e UNSPSC classificano i prodotti, mentre NAICS classifica le attività produttive o di vendita. Se si cerca all'interno di NAICS la parola libro si trovano i codici delle aziende che producono libri o li vendono, le altre codifiche applicano, invece, il codice al prodotto. Questa divergenza assume peso in fase di selezione dei frammenti ma anche in fase di integrazione. Per ottenere dei buoni risultati l'unica soluzione è fingere che anche NAICS classifichi i beni e, di conseguenza, considerare in fase di integrazione il prodotto che corrisponde all'attività produttiva codificata.

### 6.3.3 Traduzione dei frammenti

I codici, come si è detto in precedenza, sono forniti come fogli di calcolo Excel, per renderli leggibili dal sistema MOMIS ci sono due alternative:

- Rappresentarli direttamente come classi del modello ODL<sub>13</sub> usato da MOMIS.
- Rappresentarli in un formato, per esempio schema relazionale o XML, e usare in seguito il wrapper.

E' stata scelta questa seconda strada, utilizzando il formato XML; in questo modo si ha anche il vantaggio di riferirsi ad uno standard per gli scambi di dati nel Web.

#### 6.3.3.1 XML

L'XML (Extensible Markup Language) è uno standard approvato dal W3C<sup>33</sup>, è nato come linguaggio di descrizione e per le sue caratteristiche si pensa possa diventare uno standard universale di scambio dati nel Web. È necessario porre in evidenza che questo linguaggio nato per operare all'interno del WWW, viene sviluppato con lo scopo di incentivare lo scambio dei dati in maniera strutturata attraverso le varie sorgenti di informazioni inserite nella rete Internet. Lo scopo è quello di creare un linguaggio che sia orientato allo scambio elettronico di dati "machine readable". Questa visione orientata ai dati si traduce in tre sue principali peculiarità:

- L'utente può definire i tag, che rappresentano lo strumento per descrivere la realtà da modellare e che, contestualmente, racchiudono il valore della qualificazione introdotta.
- La struttura del documento può essere dotata di elementi innestati a qualunque livello di profondità.
- Qualsiasi documento può contenere una rappresentazione opzionale della propria grammatica, chiamata DTD, che può essere utilizzata attraverso specifiche applicazioni.

---

<sup>33</sup> W3C: World Wide Web Consortium



L'XML da una parte è orientato allo scambio di dati tra gli elaboratori ( come l'HTML è orientato allo scambio di documenti tra le persone), mentre dall'altra l'XML è in grado di esprimere molte caratteristiche dei dati semistrutturati: può rappresentare, infatti, dati con strutture irregolari e non conosciute in anticipo, e può modellare dati che cambiano frequentemente e senza preavviso. In altre parole è semplice convertire i dati sia strutturati sia semistrutturati appartenenti ad una determinata sorgente nel linguaggio XML, e poi utilizzare questo linguaggio come veicolo di condivisione dei propri dati con altri presenti nel Web. Il linguaggio XML potrebbe perciò rappresentare allo stesso tempo sia sorgenti di dati strutturati sia sorgenti di dati semistrutturati.

### 6.3.3.2 Document Type Declaration (DTD)

Un documento XML che rispetti tutti i vincoli di well-formedness è corretto sintatticamente e verrà accettato da tutti i parser XML, ma chi assicura la correttezza degli elementi? Un documento XML può facoltativamente dichiarare per esplicito la sua struttura. Può indicare quali elementi sono ammessi e dove devono essere collocati. Per ogni elemento è possibile indicare quali attributi possono comparire. Esiste una apposita sezione nel documento XML adibita alle dichiarazioni che ne descrivono la struttura: la DTD.

L'uso della DTD equivale a definire la grammatica del documento XML e comporta l'introduzione del concetto di *validità*. Un documento XML si dice valido se rispetta tutti i vincoli di validità imposti dalla DTD. I vincoli imposti restringono gli elementi e gli attributi che possono comparire in un documento XML. La DTD individua quindi una classe di documenti XML aventi caratteristiche simili. Ogni documento XML valido è spesso indicato come documento istanza di una determinata DTD.

Nel corso della trattazione si mostreranno, a scopo esemplificativo, i passi seguiti per svolgere l'integrazione dei frammenti di codice relativi al dominio "Toy". Dai file XML e DTD relativi alla codifica [ecl@ss](#), riportati nelle righe successive, si nota che il DTD descrive dettagliatamente la struttura del ramo della gerarchia considerato (Figura 47). Il file XML nella seconda riga richiama il nome del file DTD a cui fa riferimento e nelle righe successive sviluppa con l'ausilio dei tag un ramo della gerarchia in questione.

ECLASS-“Toy”:

DTD:

```
<!ELEMENT Advertising (PromotionalArticle)>
<!ELEMENT PromotionalArticle
(Promgiftschildrenplayingandfun)>
<!ELEMENT Promgiftschildrenplayingandfun (buildingblock |
stuffedanimal | ball | balloon | drawingpens | toy)>
<!ELEMENT buildingblock (Cod)>
<!ELEMENT stuffedanimal (Cod)>

<!ELEMENT ball (Cod)>

<!ELEMENT balloon (Cod)>
<!ELEMENT drawingpens (Cod)>
<!ELEMENT toy (Cod)>
<!ELEMENT Cod (#PCDATA)>
```

XML:

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE Advertising SYSTEM "toy.dtd">
<Advertising>
  <PromotionalArticle>
    <Promgiftschildrenplayingandfun>
      <buildingblock>
        <Cod />
      </buildingblock>
    </Promgiftschildrenplayingandfun>
  </PromotionalArticle>
</Advertising>
```

### 6.3.4 Wrapper XML

Il wrapper XML sovrintende ai documenti di tipo XML e, come ogni altro wrapper utilizzato nell’ambito del progetto MOMIS, deve soddisfare due funzionalità. La prima, consiste nel fornire una descrizione in ODLI3 della sorgente alla quale è connesso, la seconda, nel consentire l’esecuzione di query, generate dal query manager sulla specifica sorgente. La prima funzionalità indicata viene svolta completamente dal parser

DTD e dal traduttore DTD-ODLI3, utilizzando le tabelle e gli algoritmi di conversione. In particolare è necessario osservare che, oggetto del parser, non è il file dati XML, ma il file (o la sezione del file XML) contenente la rappresentazione della struttura del sorgente. Per questo motivo il wrapper sovrintende unicamente a file XML valid. I file well-formed, non avendo DTD associato, non vengono trattati.

Il funzionamento del wrapper XML, per quello che concerne la parte di analisi del file sorgente, può essere suddiviso in tre blocchi funzionali che vengono eseguiti in successione temporale.

- **Parsing della DTD:** in questa fase viene effettuata la lettura della DTD e vengono create delle strutture dati corrispondenti agli elementi individuati.
- **Determinazione delle chiavi.**
- **Traduzione del documento:** in questa fase avviene la traduzione del documento XML nel documento ODLI3, sulla base delle informazioni che si sono ottenute negli step precedenti.

### 6.3.5 SI-Designer

SI-Designer è l'interfaccia grafica del sistema MOMIS; il progettista accede a questa per svolgere l'Integrazione Intelligente delle Informazioni. SI-Designer è accessibile online mediante il collegamento al server<sup>34</sup> *sparc20.ing.unimo.it*.

E' disponibile anche una versione software di Si-Designer che permette di lavorare ad una velocità più elevata e, quindi, di gestire con minori difficoltà ingenti moli di dati. La versione software non consente l'utilizzo di tutte le funzionalità offerte dal sistema, ma permette di effettuare tutte le operazioni necessarie all'integrazione degli standard di classificazione di prodotto. L'interfaccia è user-friendly, è possibile svolgere tutte le operazioni anche per un utente che non sia a conoscenza dei principi su cui si basa il sistema. L'unica conoscenza pregressa richiesta riguarda la funzione e la modalità d'uso dei vari tool che compongono il sistema. L'integrazione può essere svolta in diverse fasi, in quanto il sistema, sia in modalità online che software, dà la possibilità di salvare il lavoro svolto e riprenderlo in qualsiasi momento.

---

<sup>34</sup> Un server è semplicemente un computer che fornisce risorse, file o informazioni.

Nei paragrafi che seguono si percorreranno le fasi dell'integrazione attraverso i vari tool utilizzando sempre come esempio i frammenti di [ecl@ss](mailto:ecl@ss) e UNSPSC relativi a "toy".

### 6.3.5.1 Source

La prima operazione da svolgere è l'inserimento delle sorgenti di dati da integrare. Le sorgenti si ottengono effettuando il wrapping dei file XML mediante l'apposito wrapper XML. Il wrapping deve essere svolto on-line, quindi, per questo motivo l'inserimento delle sorgenti è un'attività che deve essere obbligatoriamente svolta in connessione col server. Per avere la possibilità di operare off-line sui dati si deve creare uno stato XML del sistema dopo l'inserimento delle sorgenti.

In fase di immissione delle sorgenti in sistema richiede:

- **Nome del wrapper:** questo nome viene assegnato dal progettista in fase di wrapping;
- **Nome dell'host:** la macchina su cui il wrapper è installato;
- **Numero della porta** attraverso cui vi si può accedere.

Dopo aver inserito i dati si clicca sul bottone ADD e i nomi delle sorgenti vengono aggiunti nella finestra di label textitWrappers e la visualizzazione della struttura della sorgente, con la specificazione del tipo (relazionale, a oggetti, semistrutturata), nella finestra di destra denominata textitSource acquired.

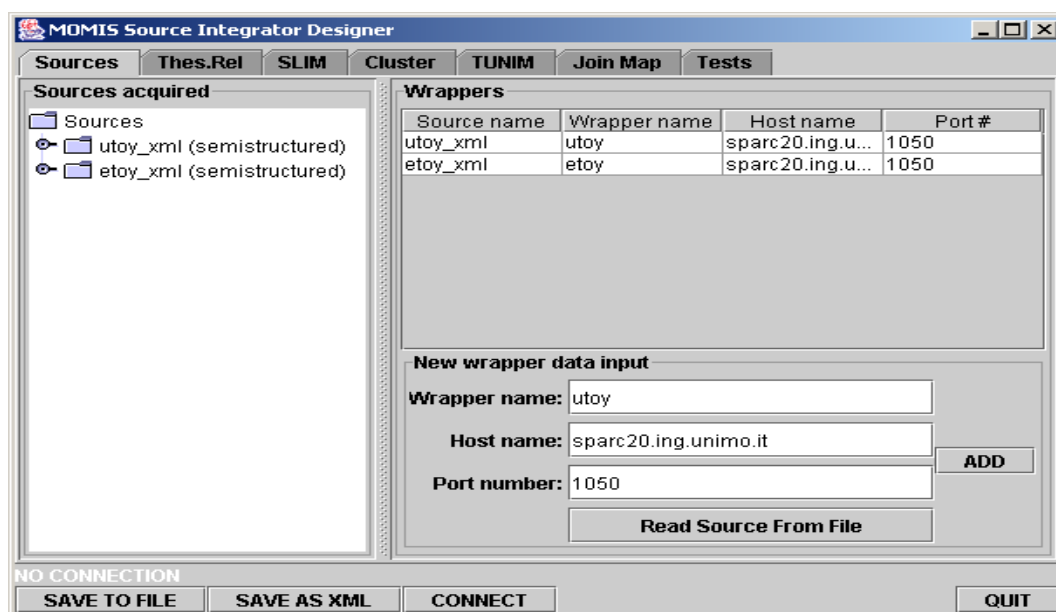


Figura 50: MOMIS- Sources

### 6.3.5.2 SIMA

Una volta caricate le sorgenti si prosegue nel processo di integrazione con la formazione del Thesaurus Comune, il primo passo consiste nell'estrazione delle relazioni intra-schema; per fare questo si accede al pannello Thes.Rel e si seleziona il pulsante run SIMA. Il risultato di questa operazione sarà la comparsa nella finestra di destra delle relazioni tra le classi e gli attributi, di una stessa sorgente. Nella finestra di destra si visualizzano i termini che partecipano alla relazione: Source, Destination separati dal tipo di relazione che li lega: SYN, BT, NT e RT. A destra compare una colonna in cui viene identificato, per mezzo di un colore e di un codice il costruttore della relazione. Nella colonna successiva si indica se la relazione è stata validata o no.

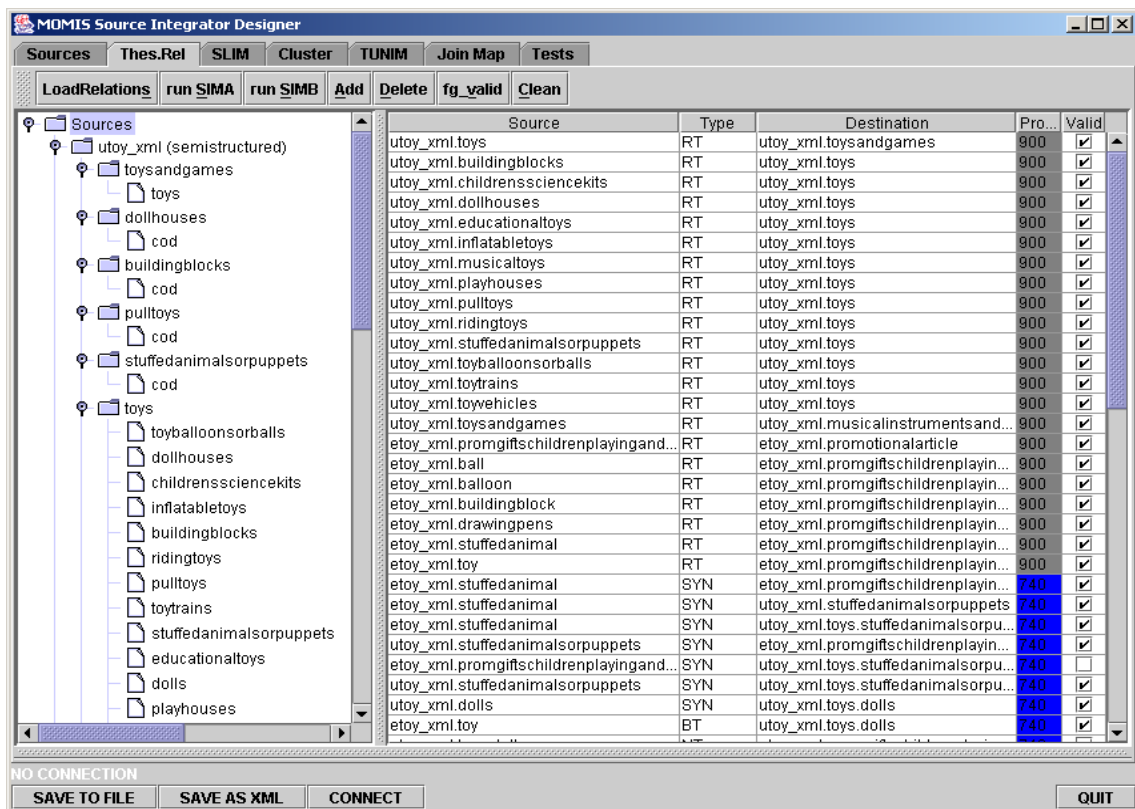


Figura 51: MOMIS-SIMA

### 6.3.5.3 SLIM

Dopo aver costruito le relazioni intrascema è necessario costruire le relazioni tra i dati appartenenti a schemi diversi, questa operazione si compie con l'ausilio del pannello SLIM. Il primo passo è l'assegnazione dei significati ai termini, vediamo come si svolge questa operazione:

- si seleziona con il tasto destro del mouse il termine che si vuole analizzare;
  - compare una tendina che elenca le seguenti voci, che permettono di assegnare un significato al termine:
- **Word form:** selezionando questa voce si apre una finestra in cui si può inserire la forma base per il termine in analisi, questa operazione è stata svolta molto frequentemente, visto che nell'integrazione in esame i termini sono costituiti dalle descrizioni abbinate alle voci di codice. Per rispettare la sintassi dell'XML le descrizioni sono state scritte senza spazi bianchi, unendo tutte le parole, tali termini sono, nella maggior parte dei casi, incomprensibili alla macchina a meno che non si inserisca una forma base. Per esempio, il codice 49. 23.15.06 "Stuffed Animal or Puppets", appartenente ad UNSPSC, viene tradotto nel DTD con: `<!ELEMENT StuffedAnimalandPuppets (Cod)>` e viene acquisito sotto forma di sorgente come "stuffedanimalandpuppets". Una stringa di questo tipo non può essere decifrata da WordNet, a meno che non si inserisca una appropriata forma base come, per esempio, "Puppet". Si deve ricordare anche che WordNet non accetta termini al plurale e non permette l'immissione di più di un vocabolo, fanno eccezione termini composti che vengono inseriti con l'interposizione dell'underscore;

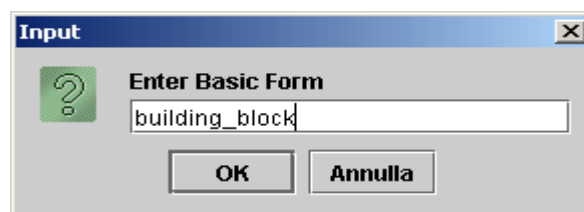


Figura 52: MOMIS-Inserimento forma base

- **Select sense:** questo comando causa l'apertura di una finestra contenente i significati che WordNet associa alla forma base scelta per il termine. In questa finestra è possibile selezionare il significato o i significati più adeguati;

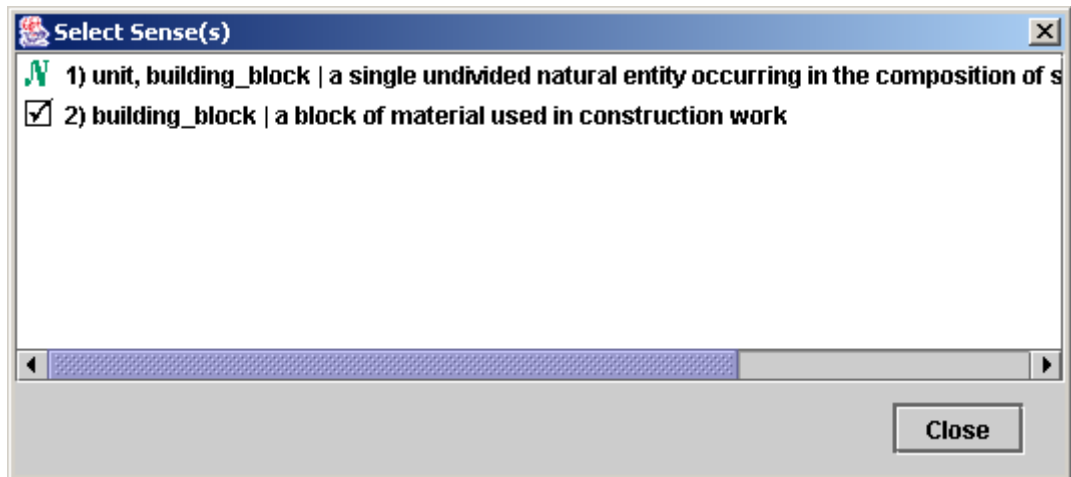


Figura 53: MOMIS-Assegnazione del significato

- **Ignored term:** questa voce permette di indurre SLIM a non considerare il termine. Questa operazione si rende necessaria quando la forma base che si assegna al termine ha un significato eccessivamente generico che potrebbe provocare 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.
- Dopo aver svolto queste operazioni si clicca sul tasto **Build** e SLIM costruisce tutte le relazioni che possono derivare dalle forme base e dai significati assegnati ai termini.

Questa è una delle fasi che richiede il maggior sforzo da parte del progettista in quanto richiede l'immissione di molte forme base e l'assegnazione di altrettanti significati. Il carico di lavoro è ulteriormente appesantito dal fatto che il sistema riconosce soltanto due livelli della gerarchia, quindi, per alcuni termini forma base e significato devono essere assegnati due volte. Ad esempio al termine "Toys" si deve considerare una volta come padre di "Dolls", "Stuffed animals or puppets", ecc. e una volta come figlio di

“Toys and games”; è molto importante assegnare la stessa forma base e lo stesso significato in entrambe le circostanze.

Il lavoro può essere svolto in più fasi grazie alle funzioni SAVE e LOAD che permettono di salvare i significati assegnati e di ripristinarli in un momento successivo.

#### 6.3.5.4 SIMB

Costruite le relazioni dal modulo lessicale si ritorna al pannello Thes.Rel, qui il progettista ha la possibilità di inserire nuove relazioni in base alla sua conoscenza del dominio e alla sua esperienza. 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 e modifica 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.

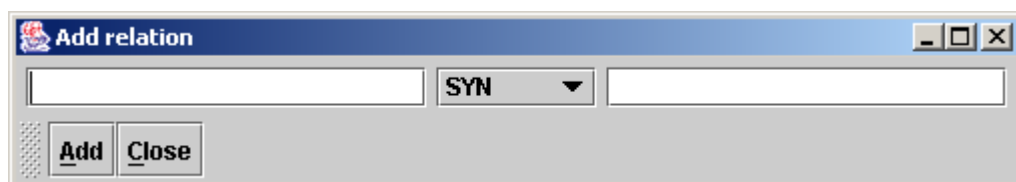


Figura 54: MOMIS-inserimento relazioni

#### 6.3.5.5 Cluster

Il pannello Cluster rappresenta l'interfaccia grafica del modulo Artemis e serve alla creazione di Cluster che raggruppano le classi locali degli schemi da integrare. Il pannello si articola in due parti:

- *Cluster Tuning*;
- *Artemis Configuration*;

Il pannello Artemis Configuration permette di configurare il modulo Artemis assegnando:



- il peso delle relazioni del Thesaurus;
- i valori di soglia per le affinità e il clustering;
- i coefficienti per l'affinità globale.

Dopo aver settato le variabili di ARTEMIS si procede premendo il pulsante CREATE GLOBAL CLASSES che fa eseguire dal sistema il modulo ARTEMIS e fa comparire nella finestra in basso i cluster costruiti secondo le specifiche assegnate.

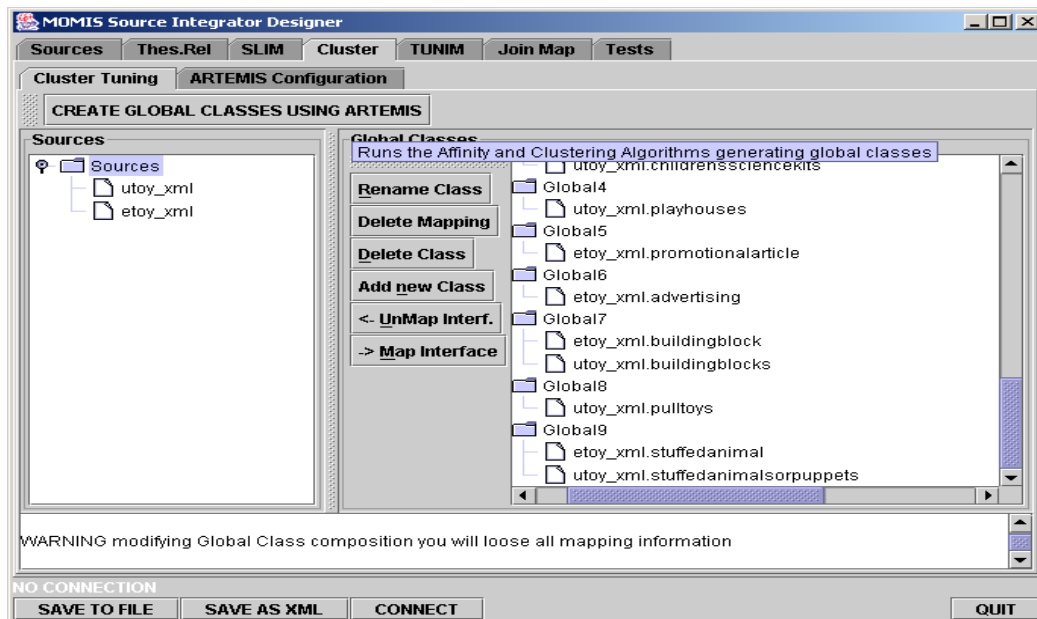


Figura 55: MOMIS-Global Classes

Nel caso non si vogliono modificare i valori standard dei parametri di ARTEMIS si può eseguire la costruzione dei cluster anche dal pannello Cluster Tuning, richiamando il modulo ARTEMIS per mezzo del pulsante CREATE GLOBAL CLASSES USING ARTEMIS. Il risultato viene presentato nella finestra a destra. Su tale risultato il progettista può intervenire manualmente con i pulsanti che si trovano al centro per effettuare delle modifiche. Selezionando una classe e premendo il pulsante Rename Class è possibile cambiare il nome a una classe globale; questa operazione è frequente poiché i cluster creati non hanno nomi significativi ma solo un numero.

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.

### 6.3.5.6 TUNIM

Il pannello TUNIM permette la costruzione degli schemi globali dopo che le classi locali sono state raggruppate in cluster dal modulo ARTEMIS.

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.

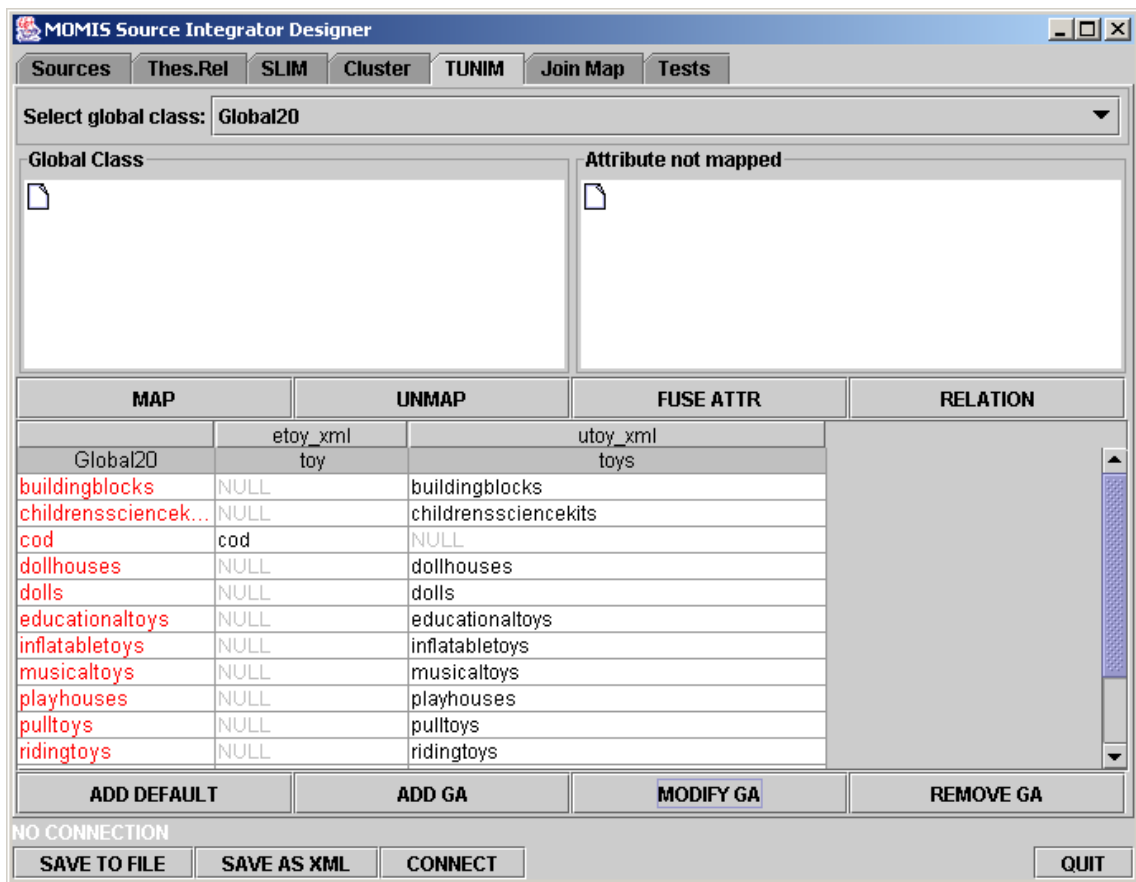


Figura 56: MOMIS-TUNIM

### 6.3.6 Analisi dei risultati

Come illustrato nelle pagine precedenti, MOMIS fornisce classi globali che raggruppano varie classi locali appartenenti alle sorgenti immesse. Questo non è il risultato che ci si era proposti di ottenere, lo scopo di questo lavoro è ottenere un mapping tra le diverse codifiche in relazione ad uno specificato dominio. Per ottenere un risultato di questo tipo è richiesta da parte del progettista un'ulteriore operazione di interpretazione.

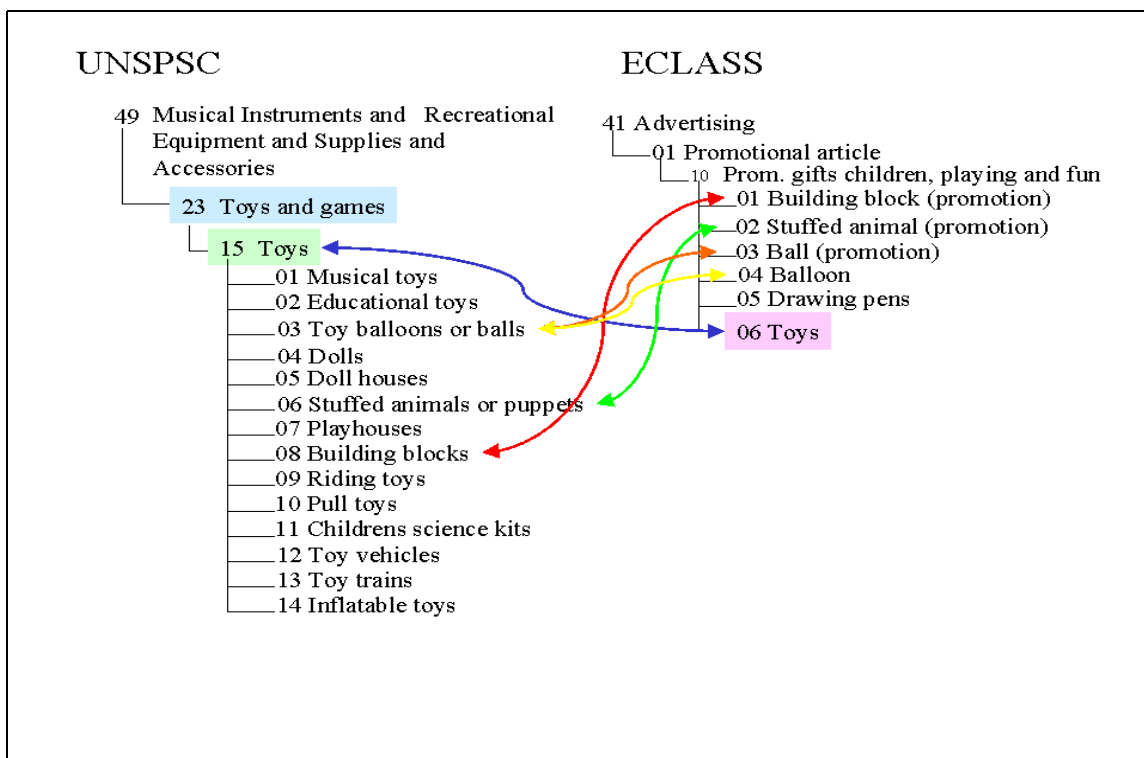
Le classi globali in alcuni casi contengono la descrizione di un solo codice in altri ne raggruppano varie che possono o meno appartenere alla stessa codifica.

A scopo esemplificativo elenchiamo le classi globali ottenute nell'esempio relativo a "Toy" e mostriamo come da queste si è dedotto il mapping tra le codifiche.

Global Class	Description
Global 0	utoy_xml.ridingtoy
Global 1	utoy_xml.inflatabletoys
Global 2	utoy_xml.educationaltoys
Global 3	utoy_xml.childrensciencekits
Global 4	utoy_xml.playhouses
Global 5	etoy_xml.promotionalarticle
Global 6	etoy_xml.advertising
Global 7	etoy_xml.buildingblock utoy_xml.buildingblocks
Global 8	utoy_xmlpulltoys
Global 9	etoy_xml.stuffedanimal utoy_xml.stuffedanimalorpuppets
Global 10	utoy_xml.toytrains
Global 11	etoy_xml.drawingpens
Global 12	etoy_xml.promgiftschildrenplayingandfun
Global 13	utoy_xml.toysandgames
Global 14	utoy_xml.musicalinstrumentsandrecreationalequipment andsuppliesandaccessories
Global 15	utoy_xml.toyvehicles
Global 16	utoy_xml.musicaltoys
Global 17	utoy_xml.dollhouses
Global 18	etoy_xml.ball
Global 19	utoy_xml.dolls
Global 20	etoy_xml.toy utoy_xml.toys

**Tabella 15: MOMIS-Classi globali**

Ogni volta che una classe globale comprende due o più codici se ne può dedurre un mapping tra tali codici, infatti, il sistema li ha considerati equivalenti e per questo motivo li ha raggruppati. Nella figura riportata in seguito si mostra quale è il mapping dedotto dalle classi globali create



**Figura 57: Mapping “Toy”**

Questo è il mapping ottenuto sul dominio in esame, ovviamente non è l’unico possibile, infatti, le scelte che il progettista ha fatto influenzano fortemente i risultati ottenuti. Nell’assegnazione dei significati e delle forme base il progettista ha l’opportunità di pilotare l’integrazione, generalizzando o al contrario specificando in modo particolare alcuni termini. Inoltre, come già specificato, si possono di aggiungere manualmente delle relazioni in base alla conoscenza che si ha del dominio.

I criteri per valutare la buona riuscita di un mapping possono essere diversi, per esempio, si può ritenere buono un mapping che:

- individua tutte e soltanto le corrispondenze esatte tra i codici, che, quindi, è in grado di ricostruire una serie di relazioni 1:1 tra i frammenti esaminati. Un esempio è dato dal mapping tra UNSPSC Toys 49.23.15 e [Ecl@ss](#) Toy 41.01.10.06.
- individua anche corrispondenze tra codici che rappresentano diversi livelli di dettaglio, per i quali sarebbe plausibile una relazione di tipo NT o BT. Una

relazione di questo tipo sussiste tra [Ecl@ss](#) Toy e Riding Toys, Pull Toys, Toy vehicles e Toy trains che appartengono a UNSPSC.

- individua per ogni codice un codice corrispondente appartenente alle altre codifiche in esame. Per ottenere un risultato di questo tipo i frammenti devono corrispondere perfettamente, sia per quanto riguarda il livello di dettaglio che per quanto riguarda la concettualizzazione del dominio in esame. È un caso che si ritiene abbastanza improbabile viste le profonde differenze tra le codifiche esaminate.

La scelta dell'obiettivo del mapping dipende fortemente da quale è il suo utilizzo finale, noto lo scopo dell'integrazione il progettista potrà fare delle scelte coerenti e finalizzate al suo raggiungimento.

# **Capitolo 7**

## **Analisi dei mapping**

### **7.1 Generale**

In questo capitolo si analizzeranno le situazioni che si sono incontrate frequentemente in fase di integrazione, mettendo in evidenza i problemi dell'approccio utilizzato. In una fase successiva, si prenderanno in esame i mapping tra le codifiche ottenuti con il sistema MOMIS. Si considereranno frammenti molto diversi tra loro, alcuni scelti con l'approccio Bottom-up e altri con il Top-down, in modo da valutare il comportamento del sistema in circostanze molto diverse. Inoltre, si svolgeranno vari tentativi di integrazione su uno stesso frammento, basandosi su distinte scelte del progettista, per valutare la ripercussione che hanno sul mapping finale.

### **7.2 Difficoltà legate al sistema**

Nello svolgimento delle prove di integrazione si sono incontrati molti ostacoli, alcuni dei quali derivano dalla struttura e dalle caratteristiche del sistema MOMIS.

Nei paragrafi successivi si descriveranno le difficoltà affrontate e si mostreranno le strategie adottate per superarle.

## 7.2.1 WordNet: un dizionario generico

Nei capitoli precedenti, sono state ampiamente descritti sia la struttura di WordNet sia il suo utilizzo da parte del sistema MOMIS. In fase di integrazione ci si è resi conto dell'importante ruolo di WordNet nell'ottenimento del mapping. Tramite l'uso di questo dizionario si assegnano, infatti, le forme base e i significati ai dati, questa è una fase di fondamentale importanza affinché il sistema sia in grado di individuare il maggior numero possibile di relazioni corrette tra le sorgenti. Il problema sorge quando ci sono dati per cui non è possibile individuare una forma base inclusa in WordNet. Questo problema si presenta di frequente visto che WordNet, in quanto dizionario generico, non contiene termini molto specifici. Il progettista, a questo punto, può optare per diverse soluzioni:

1. *Individuare un sinonimo*: più un termine è specifico più sarà difficile individuarne un sinonimo che non comporti alcuna perdita di significato.
2. *Generalizzare*: se il livello di specificità è eccessivo per trovare un vocabolo all'interno del dizionario in esame, lo si può sostituire, al momento dell'immissione della forma base, con un termine meno specifico ma che comporti il minor numero possibile di perdite di significato.
3. *Il termine può essere ignorato*: se non si individua una forma base adeguata, si può sempre decidere di ignorare il termine. Questa scelta porta il sistema a non individuare relazioni che lo coinvolgono. Si può ovviare a questa mancanza con l'immissione manuale da parte del progettista delle relazioni.

I termini che comportano le maggiori difficoltà sono i termini composti, nei quali un aggettivo o un altro sostantivo specifica il significato di un termine generico. Se, ad

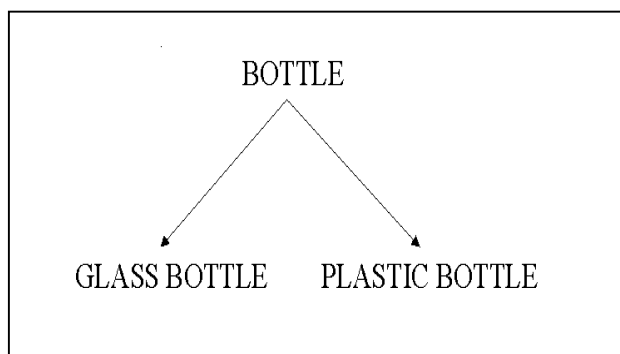


Figura 58:Gerarchia sul dominio Bottle

esempio, si ha un ramo di gerarchia costituito da termini come quelli in Figura 58, nascono immediatamente dei problemi. WordNet riconosce il termine BOTTLE ed è in grado di riconoscere anche i termini relativi ai materiali come GLASS e PLASTIC. Il problema sorge quando



si immettono come forma base “PLASTIC BOTTLE” e “GLASS BOTTLE”, infatti, questi vocaboli non vengono riconosciuti. L’inserimento della forma base “Bottle” in tutti i tre casi rappresenta una generalizzazione eccessiva che comporterebbe la perdita della completezza e della ricchezza delle codifiche. Un’altra possibilità è data dall’inserimento manuale di tutte le relazioni che riguardano questi codici. Nel caso in cui il dominio non sia troppo specifico, si possono adottare soluzioni alternative che non sono “formalmente corrette” ma che consentono di ottenere buoni risultati. In questo caso, si può assegnare come forma base di “Glass Bottle” “Glass” e di “Plastic Bottle” “Plastic”, a patto che si adotti la stessa strategia per tutte gli altri codici con descrizione simile a questa. Questa soluzione ha consentito di ottenere buoni risultati, ma non è consigliabile adottarla per sorgenti di grandi dimensioni, per le quali risulterebbe difficile da parte del progettista effettuare un controllo sulle relazioni create.

## **7.2.2 Problemi di ripetizione**

Lo scopo dell’integrazione è individuare il maggior numero possibile di corrispondenze tra due differenti codifiche, quindi, la presenza di termini comuni tra le sorgenti immesse è una condizione auspicabile. Nella scelta dei frammenti, si presta molta attenzione alle similitudini tra le gerarchie considerate e si dà un peso positivo alla presenza di definizioni uguali appartenenti a codifiche diverse. Questa situazione, vantaggiosa ai fini dell’integrazione, rappresenta un problema per MOMIS nel momento in cui si salva lo stato ottenuto. Lo stato viene, infatti, salvato come file XML e la presenza di un vocabolo ripetuto all’interno di tale file lo fa considerare non valido dal sistema che ne impedisce il riutilizzo. Quindi, al progettista è richiesto di prestare attenzione affinché nei file XML e DTD inseriti come sorgenti non compaiano termini comuni. All’apparenza è un problema banale ma, analizzando le sorgenti, ci si rende presto conto che non è immediato individuare similitudini, specialmente se i frammenti sono di grandi dimensioni. Se non si riesce a individuare la ripetizione nelle sorgenti se ne intuirà la presenza quando dopo aver salvato correttamente uno stato non si riuscirà a ripristinarlo. L’unica soluzione, a questo punto, è ripartire dal principio, dopo aver individuato il termine ripetuto (viene indicato nel segnale di errore quando si cerca di aprire SI-Designer), si sostituisce in una delle sorgenti con un sinonimo o

semplicemente con il suo plurale. Di conseguenza, si deve ricaricare il wrapper sulla sorgente modificata e ripartire dall'inizio nell'immissione delle sorgenti in SI-Designer. In appendice si riporta il file XML relativo allo stato salvato relativo all'integrazione sul dominio "Toy", entrambi i frammenti integrati in questo esempio contengono il termine "toy", quindi, per poter salvare e ripristinare lo stato, in una delle due sorgenti in questione è stato sostituito con "toys". Si riportano alcune sezioni del file XML in cui compare la ripetizione, al fine di illustrare l'ambiguità creata e giustificare l'inserimento del plurale:

```
<Sourcename="etoy_xml" type="semistructured" wrapperName="etoy"
  hostname="sparc20.ing.unimo.it" portNumber="1050">
```

...

```
- <Interfacename="toy" persistent="false">
  <extent name="toy" />
- <Attributename="cod" type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="toy">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
```

...

```
- <Attributename="toy" type="toy">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toy">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
```

...

```
<Sourcename="utoy_xml" type="semistructured" wrapperName="utoy"
  hostname="sparc20.ing.unimo.it" portNumber="1050">
```

...

```
- <Attributename="toy" type="toy">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toy">
```

```

    <SlimNodeSenso numero="1" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
...
- <Interfacename="toy"persistent="false">
  <extent name="toy" />

```

Si nota immediatamente che lo stesso termine è utilizzato in due diverse posizioni per rappresentare due diverse sorgenti, tale ambiguità rende il file XML non valido e, quindi, deve essere rimossa.

### 7.2.3 SLIM: un tool rigido

Si è ribadito, in varie circostanze, che il tool SLIM ha un ruolo centrale nell'ottenimento del mapping, infatti, mediante tale pannello si inseriscono le forme base ed i significati e, infine, si generano le relazioni interschema. Ci si è resi conto che questo tool limita molto l'elasticità del sistema, infatti, se, durante un'integrazione, si desidera inserire una sorgente aggiuntiva, il lavoro svolto, fino a quel momento, con SLIM viene perso nel momento stesso in cui la sorgente viene introdotta. Per esempio, se dopo aver immesso due sorgenti, mentre si inseriscono le forme base, ci si rende conto che sarebbe proficua l'introduzione di una terza sorgente, in quanto contenente codici coerenti con le altre, al momento dell'inserimento di questi nuovi dati le forme base già immesse e le eventuali relazioni vengono cancellate.

Questo problema non pregiudica la buona riuscita dell'integrazione, ma comporta una forte limitazione nell'utilizzo del sistema. Capita spesso di rendersi conto, solo in fase di integrazione, che sarebbe necessario l'inserimento di un'altra sorgente al fine di ampliare il dominio in esame o di trovare nuove corrispondenze. È difficile che i frammenti selezionati corrispondano perfettamente, quindi, si possono cercare corrispondenze in altri rami non ancora esplorati.

Lo stesso problema sorge nel momento in cui si desidera ripristinare uno stato salvato, infatti, anche in questo caso, non è possibile inserire ulteriori sorgenti, salvo il caso in cui si accetti la perdita di tutto il lavoro svolto sui dati precedentemente.

## 7.2.4 Lettura parziale delle gerarchie

Gli standard di classificazione dei prodotti si avvalgono di una struttura gerarchica che permette di gestire le informazioni al meglio. I file XML e DTD, scritti per descrivere i frammenti di codice mantengono perfettamente la struttura gerarchica, il problema sorge in fase di wrapping, infatti, le sorgenti immesse in MOMIS non conservano interamente la gerarchia delle codifiche ma ne mantengono solo due livelli. Vediamo, per esempio, come viene letto il ramo di gerarchia nelle sorgenti.

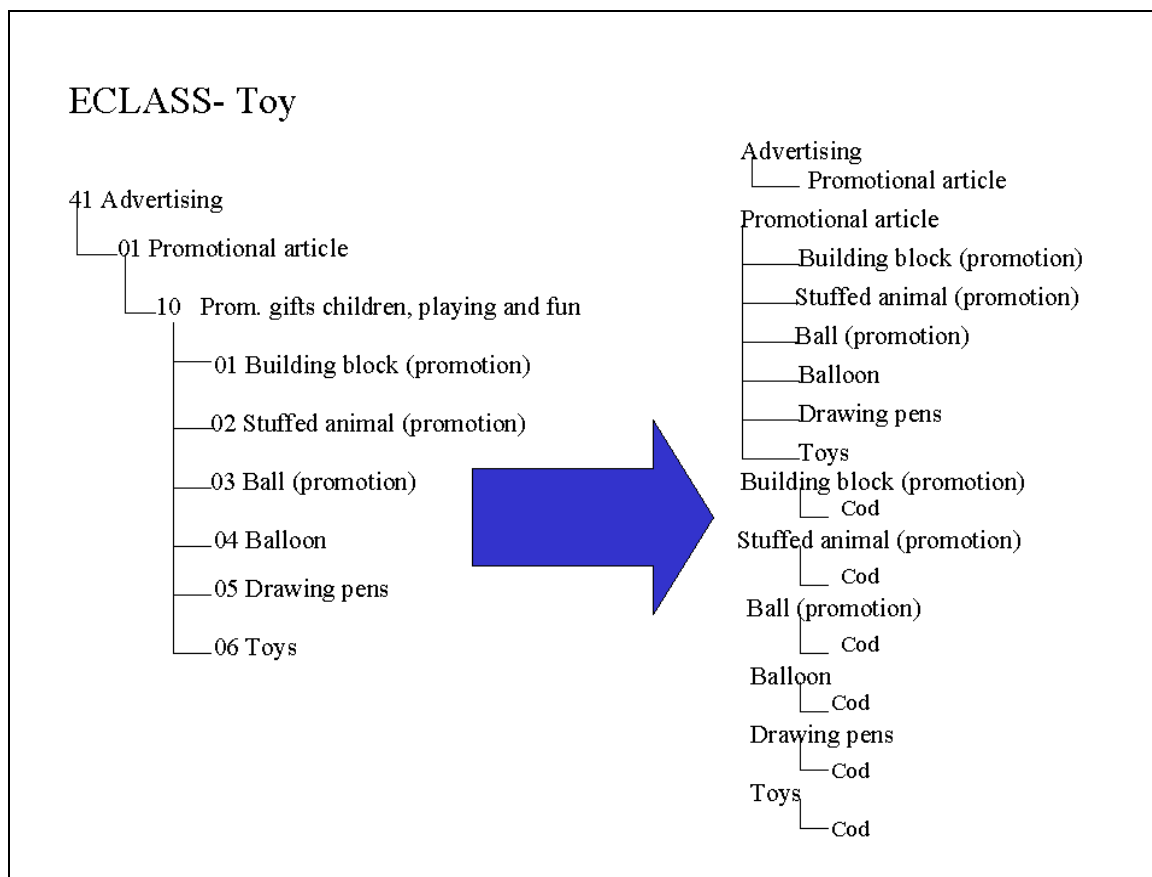


Figura 59: Trasformazione delle gerarchie nelle sorgenti.

Questo modo di leggere la gerarchia comporta molti inconvenienti nell'inserimento delle forme base e nell'assegnazione dei significati. Infatti, per tutti i termini che non appartengono al primo livello del frammento di gerarchia considerato, i significati e le forme base devono essere inseriti due volte: una considerandoli come padri una come

figli. Si può, dunque, prevedere come miglioramento dell'applicazione l'assegnamento automatico dell'annotazione ai due concetti corrispondenti.

Il motivo per cui nelle sorgenti si considera un solo livello di gerarchia risiede nel fatto che il sistema è preposto alla lettura e integrazione di oggetti con struttura:

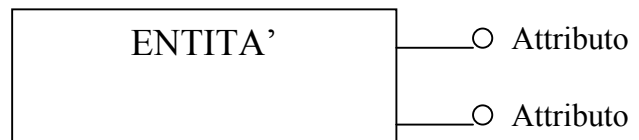
ENTITA'

→ ATTRIBUTO

→ ATTRIBUTO

.....

→ ATTRIBUTO



Un ulteriore livello di gerarchia con radice ATTRIBUTO non ha ragione di esistere. Al progettista è richiesta molta attenzione nell'assegnare ad uno stesso codice sempre la stessa forma base e lo stesso significato.

Un problema nella lettura delle sorgenti si presenta anche al quarto livello delle codifiche, il sistema leggerebbe Ball come attributo di Toy e non lo coinvolgerebbe nell'integrazione, per ovviare a questo inconveniente si è inserito un livello di gerarchia fittizio rappresentato dalla definizione "Cod". "Cod" diventa, quindi, un attributo legato ai codici del quarto livello che fa sì che anche questi vengano considerati come entità.

Si deve ricordare che, nonostante la gerarchia non venga letta nelle sorgenti immesse, le relazioni intra-schema vengono ripristinate da SIMA, anche se al posto di relazioni NT/ BT individua relazioni RT.

## 7.3 Difficoltà legate ai codici

Le codifiche hanno strutture che non favoriscono l'integrazione e che, indipendentemente dalla validità del sistema MOMIS, hanno comportato molte difficoltà prima in fase di individuazione dei frammenti poi in fase di integrazione. L'obiettivo dei paragrafi successivi è analizzare i problemi imposti dalle codifiche.

### 7.3.1 Parziale sfruttamento delle gerarchie

Le gerarchie, che potrebbero costituire un punto di forza per la consistenza e la

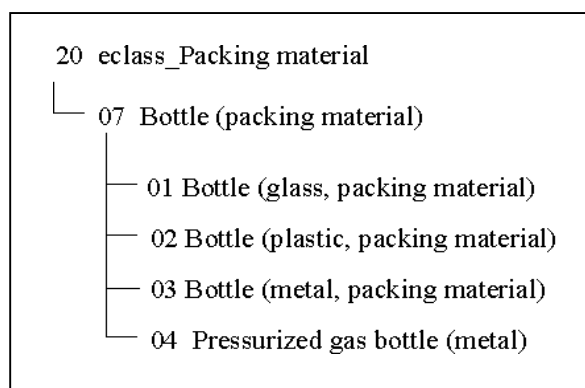


Figura 60: Ripetizioni nella gerarchia

completezza delle codifiche, sono, in molti casi, utilizzate male. Non si può fare riferimento ad uno standard in particolare, infatti, dipende dai frammenti considerati.

In [ecl@ss](#) capita frequentemente che due livelli consecutivi presentino la stessa

descrizione dei codici, con l'aggiunta, al

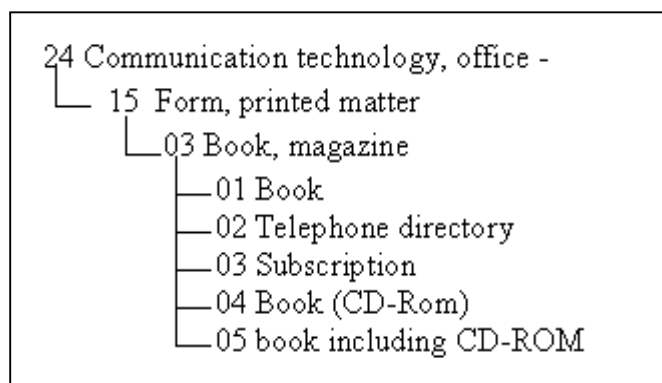


Figura 61: Sfruttamento parziale della gerarchia

livello inferiore, di una specificazione tra parentesi. A livello pratico, ciò non comporta particolari problemi, il problema sorge a livello concettuale, in quanto la gerarchia permetterebbe una gestione più precisa e ordinata dei contenuti. Nel caso del ramo relativo a book in [ecl@ss](#) (riportato in Figura 61), all'ultimo livello si

potrebbero inserire molti più codici in grado di dettagliare in modo approfondito il dominio. L'esempio, riportato in Figura 62, relativo al dominio software è molto

emblematico, in quanto i codici del terzo livello hanno la stessa descrizione di quelli del secondo, con l'aggiunta di una laconica specificazione tra parentesi. Al quarto livello, infine, un codice è associato alla descrizione "Other hardware" che

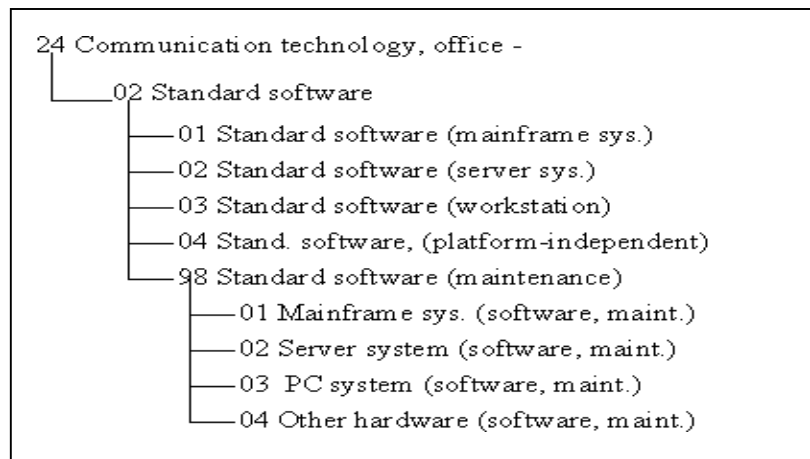


Figura 62: Gerarchia non ottimizzata

lascia molte diverse possibilità interpretative all'utente. In altri casi, in cui la gerarchia fornirebbe ampie possibilità di approfondimento del dominio, la codifica si limita a fornirne una descrizione sommaria e incompleta.

### 7.3.2 Diversi obiettivi alla base delle classificazioni

Le classificazione dei prodotti sono state create indipendentemente l'una dall'altra e, specialmente, con scopi molto diversi. NAICS classifica le attività produttive e di

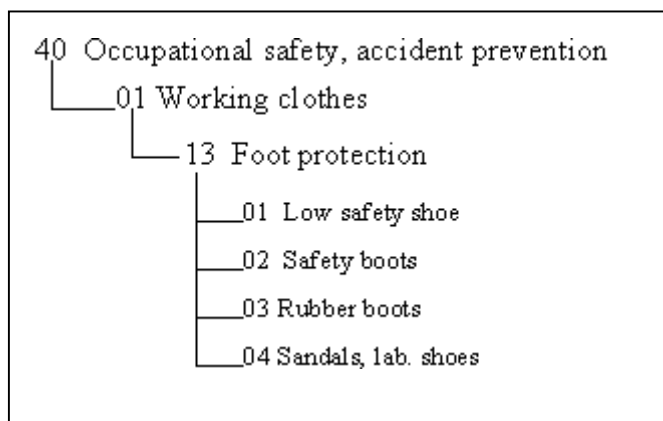


Figura 63: Contesto di classificazione dei prodotti

vendita, quindi, non si può trovare al suo interno nessun codice finalizzato alla descrizione di un prodotto in sé. [Ecl@ss](#) è stato sviluppato con lo scopo di classificare i materiali per le aziende chimiche, ma poi ha

esteso la gamma di prodotti considerati fino ad inglobare

molte altre categorie di prodotti, mantenendo sempre un'orientazione volta alle forniture industriali. La conseguenza è che in tale codifica i prodotti come le scarpe e l'abbigliamento vengono considerati solo dal punto di vista della protezione sul lavoro e dell'antifortunistica. Si intuisce immediatamente che una tale diversità di orientazione in alcuni casi rende l'integrazione priva di senso. UNSPSC è prettamente orientato

verso i prodotti e prende in considerazione tutta la gamma di prodotti in commercio, quindi, in alcuni domini raggiunge livelli di dettaglio inferiori alle altre codifiche, in altri fornisce una descrizione molto più completa.

## 7.4 Esempi di integrazione

Nelle pagine successive saranno riportati alcuni esempi di integrazione svolti per testare il sistema. Si noterà che ci sono due fattori critici nello svolgimento dell'integrazione:

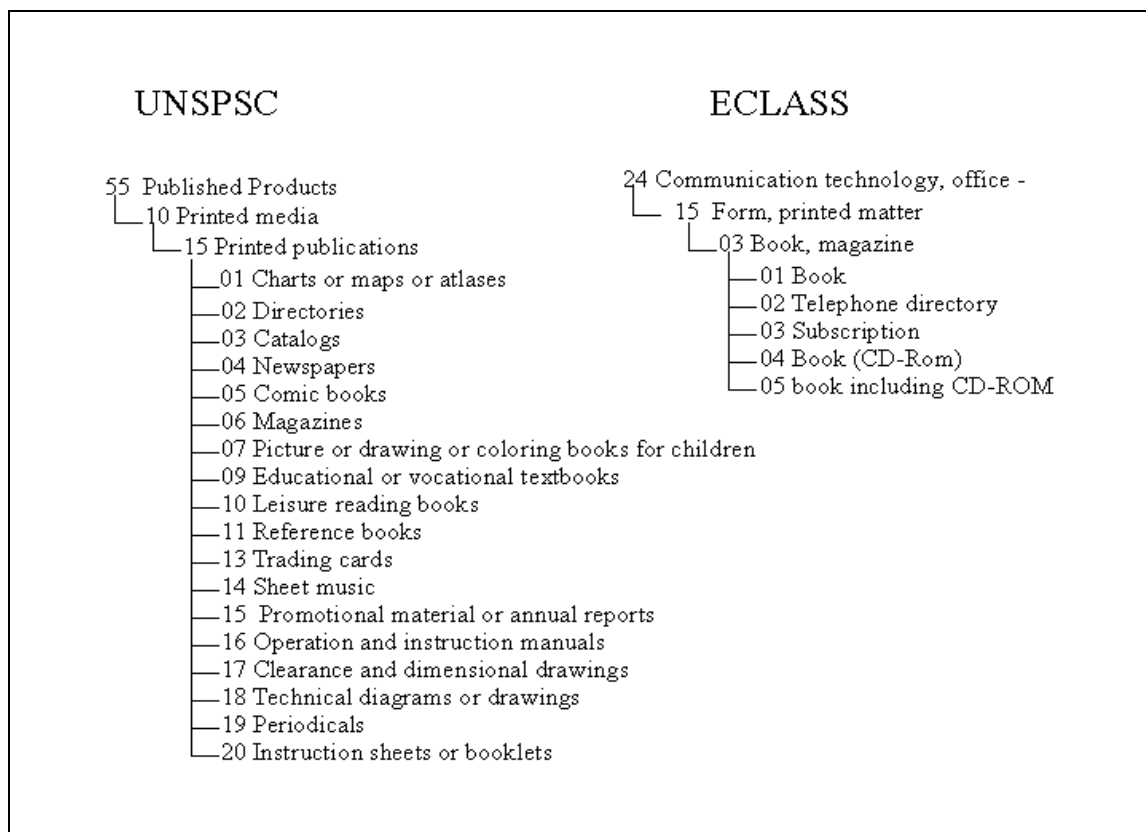
- gli obiettivi che il progettista si pone;
- la struttura delle codifiche.

Il progettista deve, infatti, decidere a priori che tipo di mapping desidera ottenere e orientare tutte le sue scelte successive al raggiungimento di questi obiettivi. Può scegliere di creare il mapping solo quando i codici corrispondono perfettamente, oppure può scegliere di creare mapping più generali tra codici non perfettamente corrispondenti. Le codifiche rappresentano la base per l'integrazione, quindi, la loro struttura ne influenza fortemente i risultati. Quando i frammenti non presentano una buona corrispondenza il risultato dell'integrazione non sarà ottimo, indipendentemente dall'abilità del progettista.

### 7.4.1 Integrazione dei frammenti relativi a Book

I frammenti relativi al dominio Book sono stati scelti con l'approccio Bottom-up, sono stati individuati inserendo nel motore di ricerca delle codifiche il termine "Book". Per UNSPSC ed [ecl@ss](mailto:ecl@ss) la ricerca ha fornito dei buoni risultati che hanno portato all'individuazione dei frammenti rappresentati nell'immagine, per quanto riguarda NAICS ci si è trovati in una situazione molto critica, infatti, i codici individuati sono relativi ad aziende che commercializzano libri o ad aziende che li producono e non ai libri stessi.





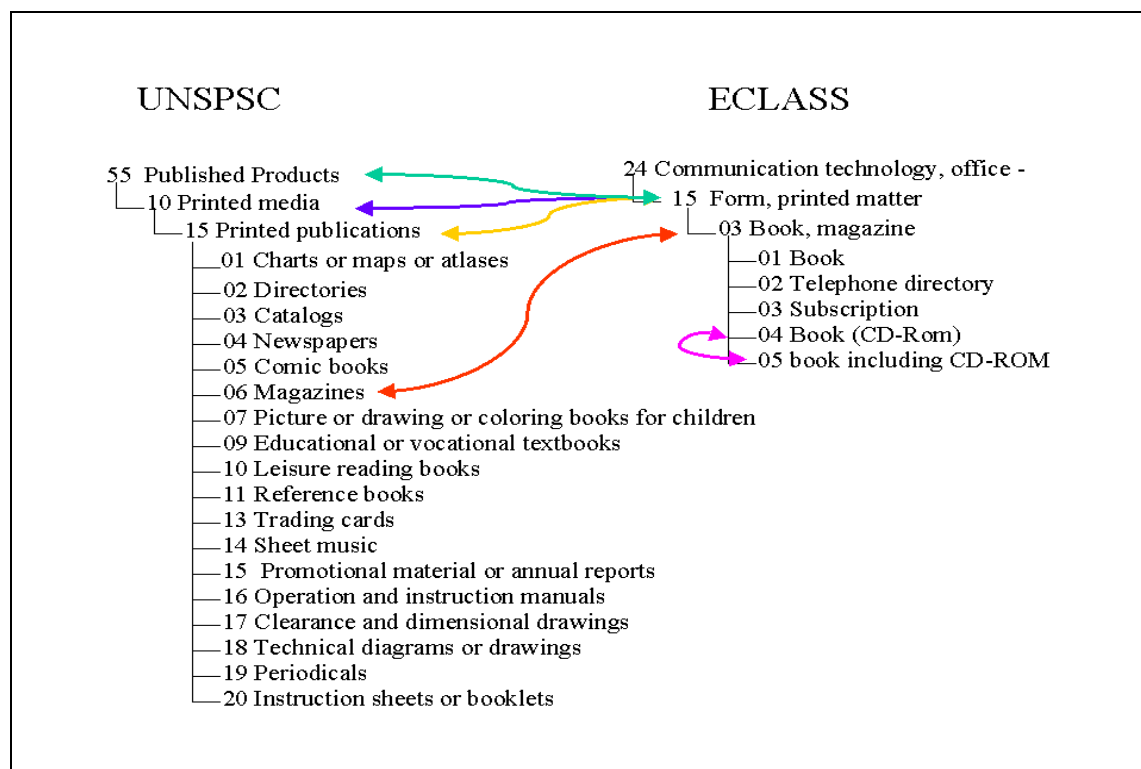
**Figura 64: UNSPSC-ECLASS :BOOK**

Si è scelto di non considerare in questo esempio il frammento di NAICS, perché troppo diverso dagli altri due che classificano espressamente i prodotti e non le attività produttive. Se si ritenesse inevitabile l'integrazione con la codifica NAICS, considerati i diversi punti di vista, si dovrebbe ricorrere ad espedienti che la rendano più omogenea ad [ecl@ss](#) e UNSPSC. Si nota immediatamente la differenza di approfondimento tra le due codifiche considerate e anche il diverso sfruttamento della gerarchia.

UNSPSC dettaglia molto bene il dominio considerando tutti i tipi di pubblicazioni stampate, [ecl@ss](#), al contrario, lo analizza in maniera sommaria distinguendo libri e giornali in 5 categorie: libri, elenco telefonico, sottoscrizioni, libri su CD-ROM e libri con incluso CD-ROM.

Il progettista deve scegliere che strategia adottare, una possibilità è creare solo i mapping tra i codici che corrispondono perfettamente, l'altra è generalizzare le definizioni in modo da ottenere un numero maggiore di match. Nel primo caso l'obiettivo è mettere in relazione tra loro solo i codici che hanno descrizioni perfettamente corrispondenti e sono ad uno stesso livello di dettaglio nel dominio. Per ottenere un mapping di questo tipo è sufficiente inserire le forme base e i significati

corretti, il sistema individuerà autonomamente tutte le relazioni indispensabili all'integrazione. In Figura 65 è raffigurato il mapping ottenuto.



**Figura 65: Mapping semplice sul dominio Book**

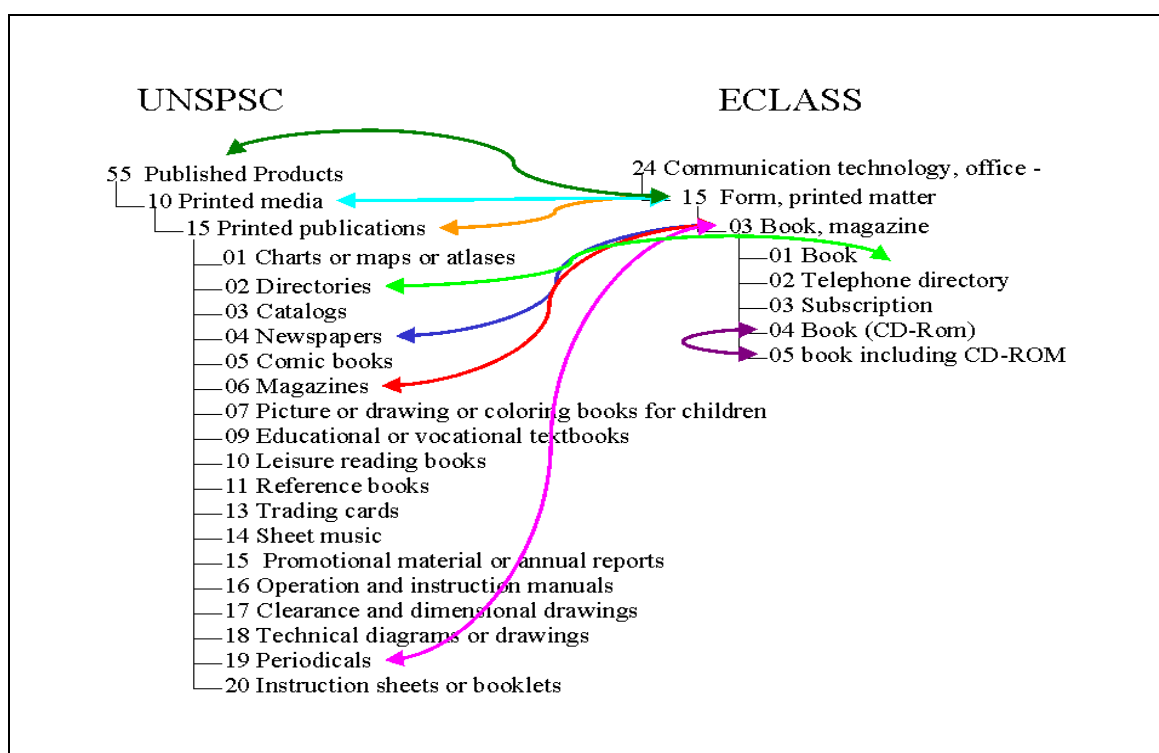
I risultati dell'integrazione non sono buoni, infatti, le poche corrispondenze individuate si trovano a livelli alti della gerarchia. La responsabilità degli scarsi risultati ottenuti non è imputabile a MOMIS, ma alla diversità tra le codifiche che rende l'integrazione infruttuosa.

Per ottenere un mapping più ricco, il progettista deve aggiungere delle relazioni a quelle che il sistema ha autonomamente creato tra le sorgenti. In questo caso, sono state aggiunte delle relazioni molto semplici, che possono essere dedotte tramite la conoscenza del dominio, ma che il sistema non era stato in grado di individuare. Le relazioni inserite manualmente sono riportate in Tabella 16.

Dopo aver inserito le relazioni, è necessario ricostruire i Cluster, in modo che il sistema prenda in considerazione anche queste nuove informazioni. Nel caso in esame le relazioni inserite hanno provocato delle variazioni nelle classi globali e, conseguentemente, nel mapping tra le codifiche. In Figura 66 è raffigurato il mapping ottenuto con l'aggiunte delle relazioni.

Source	Type	Destination
UNSPSC Comic Book	NT	<a href="#">Ecl@ss</a> Book
UNSPSC Leisure Reading Book	NT	<a href="#">Ecl@ss</a> Book
UNSPSC Reference Book	NT	<a href="#">Ecl@ss</a> Book
UNSPSC Newspapers	SYN	<a href="#">Ecl@ss</a> Book magazines
UNSPSC Periodicals	SYN	<a href="#">Ecl@ss</a> Book magazines
UNSPSC Directories	SYN	<a href="#">Ecl@ss</a> Telephone Directory
UNSPSC Reference Books	BT	<a href="#">Ecl@ss</a> Telephone Directory

**Tabella 16: Relazioni inserite manualmente**



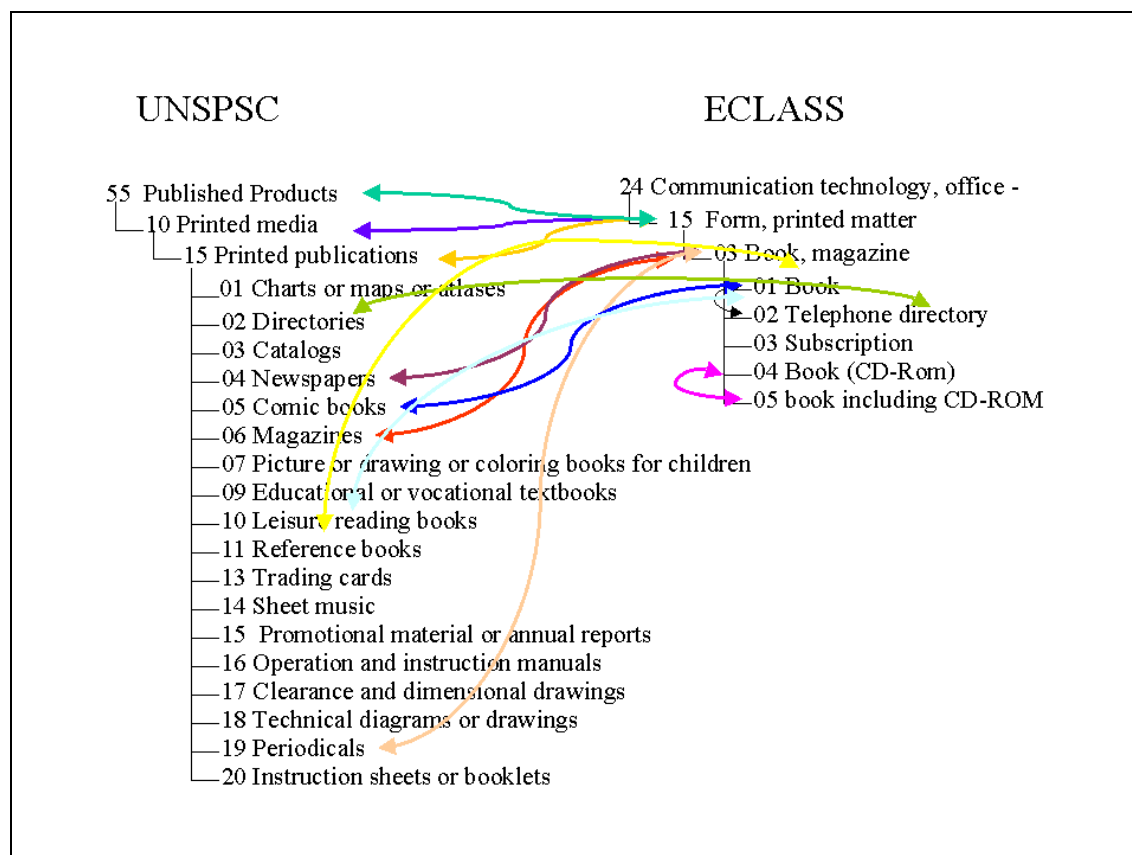
**Figura 66: Mapping sul dominio Book**

La differenza principale del mapping corrente è che lega codici che non hanno necessariamente lo stesso livello di dettaglio, per esempio lega “book and magazine” a “magazines” e a “periodicals”. Questo dipende dal fatto che il progettista ha inserito una relazione di sinonimia tra tali termini, se avesse inserito una relazione di iponimia, come nel caso di “Book” e “Comic Books”, non si sarebbe creato un mapping. Queste scelte non sono discutibili a priori, ma dipendono strettamente dagli obiettivi che il progettista si è posto. A questo punto, si possono modificare le relazioni inserite trasformandole in relazioni di sinonimia:

Sources	Type	Destination
UNSPSC Comic Book	SYN	<a href="#">Ecl@ss</a> Book
UNSPSC Leisure Reading Book	SYN	<a href="#">Ecl@ss</a> Book
UNSPSC Reference Book	SYN	<a href="#">Ecl@ss</a> Book
UNSPSC Reference Books	SYN	<a href="#">Ecl@ss</a> Telephone Directory

**Tabella 17:Relazioni modificate**

L'inserimento di queste relazioni provoca la cancellazione di quelle, inserite nella fase precedente, che coinvolgevano la stessa sorgente e la stessa destinazione. Il mapping ottenuto è più completo dei precedenti, in quanto individua un numero maggiore di match, anche se questi ultimi non implicano corrispondenze perfette. Si può concludere che il sistema è estremamente flessibile, infatti, permette al progettista di ottenere il mapping che più si addice alle sue esigenze, adottando, semplicemente, qualche astuzia nell'assegnazione delle forme base o inserendo qualche relazione che il sistema non aveva individuato autonomamente.



**Figura 67: Mapping ampio sul dominio Book**

## 7.4.2 Integrazione dei frammenti relativi a Pen

I frammenti relativi al dominio Pen sono stati scelti con l'approccio Bottom-up, infatti, sono stati individuati inserendo nel motore di ricerca delle codifiche in termine "Pen". La codifica NAICS non è stata presa in considerazione perché, come ribadito precedenza, la discordanza di punti di vista con le altre codifiche richiede eccessive forzature nell'interpretazione dei codici.

Si è scelto di effettuare l'integrazione dei frammenti di codice che descrivono gli accessori per il disegno e la scrittura, in quanto la terminologia specifica del settore è abbastanza semplice, quindi, dovrebbe essere comprensibile anche per WordNet. Nel caso di UNSPSC, PEN è classificato nel segmento: "Office Equipment and Accessories", invece, nel caso di [ecl@ss](#) appartiene alla classe : "Communication Technology- Office". La definizione del secondo livello è per entrambi "Office Supplies". UNSPSC è, come solito, più ampio di [ecl@ss](#) e dettaglia in modo più approfondito il dominio. Non è, però, possibile valutare quale dei due codici sia più completo, anche se [ecl@ss](#) utilizza meno codici, ci sono codici di [ecl@ss](#) che

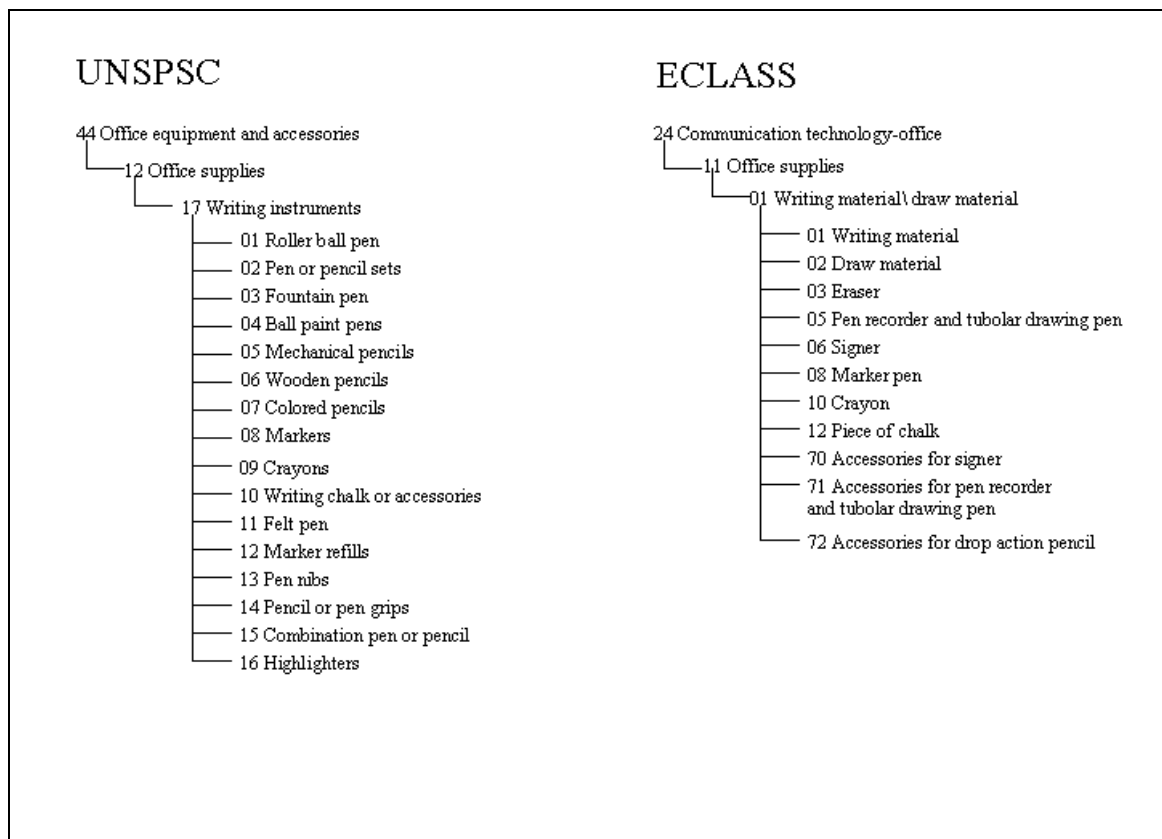


Figura 68:UNSPSC-ECLASS: PEN

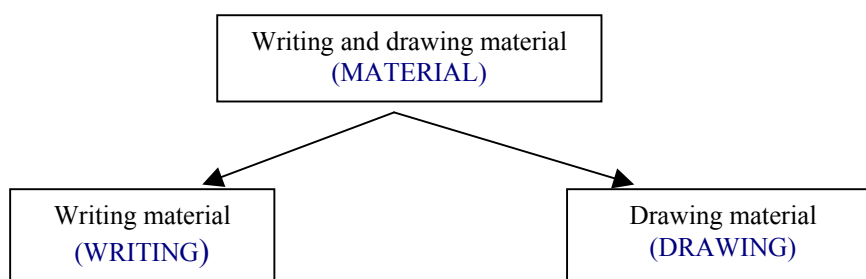
non trovano una corrispondenza in UNSPSC. Allo stesso modo, ci sono codici di UNSPSC che non trovano corrispondenza in [ecl@ss](mailto:ecl@ss).

Si è deciso di analizzare la classificazione fino all'ultimo livello di dettaglio, infatti, i frammenti selezionati si diramano solo in tale livello. Il problema è che la terminologia di questo livello è specifica e, nell'assegnazione delle forme base, sorgono delle difficoltà dovute al fatto che alcuni termini non vengono riconosciuti da WordNet.

Nell'integrazione dei frammenti relativi a "Pen", le maggiori difficoltà si sono incontrate nell'assegnazione di significati e forme base. Infatti, è molto difficile assegnare la forma base ad una descrizione come "Mechanical pencil". Mechanical verrebbe interpretato come aggettivo e non verrebbe automaticamente legato al dominio di pencil, se si assegnasse Pencil si generalizzerebbe eccessivamente con il rischio di arrivare alla creazione di grandi classi globali prive di senso e di utilità.

La stessa cosa accade per "Accessories for signer", infatti, "Signer" non è la forma base corretta, in quanto ha un significato diverso ed "Accessory" è troppo generico.

Ci sono casi in cui, vista la struttura gerarchica del codice, si è, comunque, riusciti ad assegnare un significato pertinente, ad esempio nel caso di:



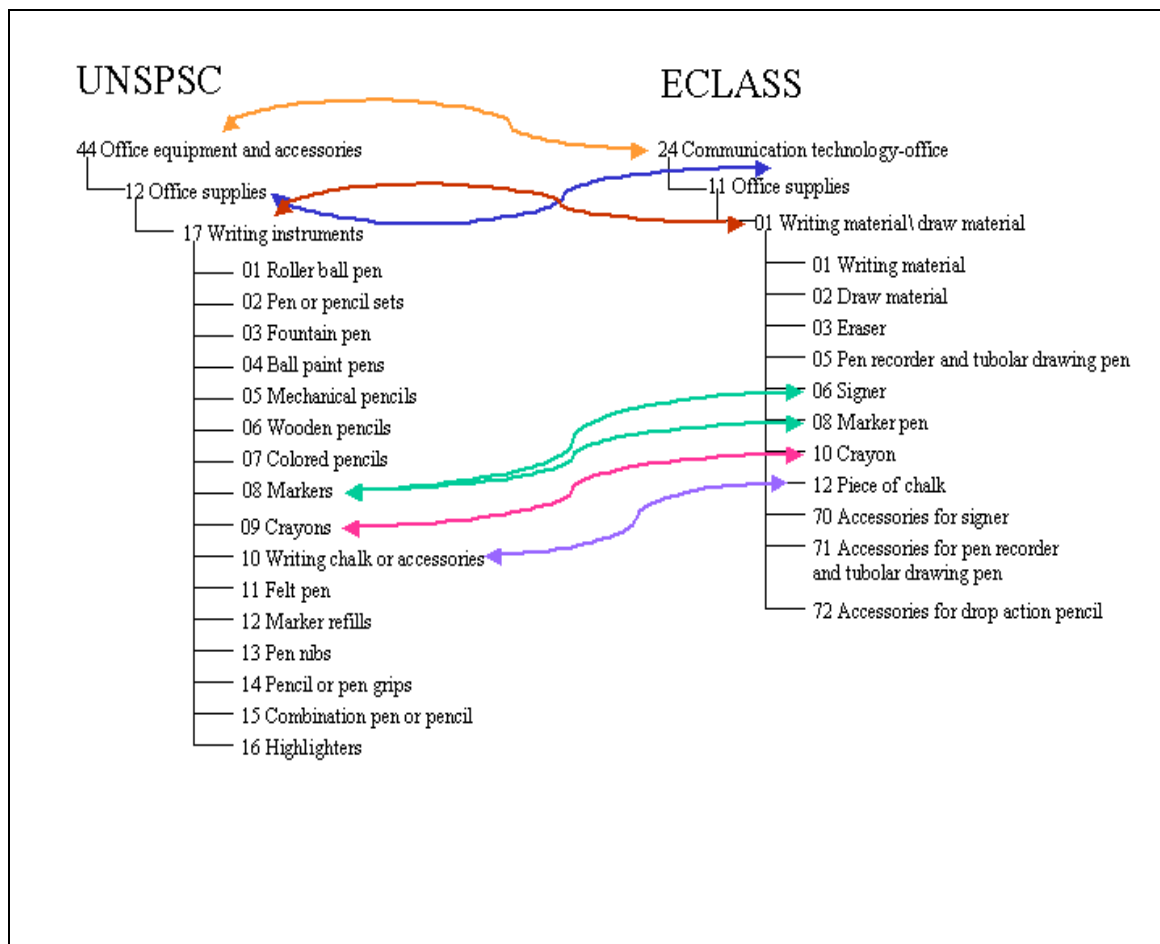
Ci sono casi, come quello di "Signer", in cui WordNet riconosce la forma base ma non possiede un significato appropriato. Oppure il caso di "Pen-nib" in cui non si è riusciti a trovare una forma base riconosciuta da WordNet.

In questa prova, si è scelto di assegnare la forma base ed il significato solo nel caso in cui se ne individuasse uno perfetto per il concetto in questione, in modo da evitare generalizzazioni e, conseguentemente, grandi classi globali prive di utilità. Con questa strategia, si individuano molti meno match, ma quelli individuati costituiscono una corrispondenza 1:1 che delinea un mapping valido tra i due sistemi di classificazione.

Nelle righe successive si riportano le classi globali generate dall'integrazione:

<b>Global Class</b>	<b>Description</b>
0	UNSPSC combination pen or pencil
1	Eclass accessories for signer
2	Eclass pencil or pen grips
3	Eclass piece of chalk UNSPSC writing chalk or accessories
4	Eclass crayon UNSPSC crayons
5	UNSPSC colored pencils
6	UNSPSC pen or pencil sets
7	UNSPSC fountain pen
8	UNSPSC mechanical pencils
9	Eclass office supply UNSPSC office supplies
10	Eclass Communication technology –office UNSPSC Office equipment and accessories
11	UNSPSC Rollerball pen
12	Eclass accessories for pen recorder and tubular drawing pen
13	UNSPSC felt pen
14	Eclass writing material draw material
15	Eclass draw material
16	UNSPSC ball paint pen
17	Eclass writing material UNSPSC writing instruments
18	Eclass eraser
19	UNSPSC pen nibs
20	Eclass marker pen Eclass signer UNSPSC markers
21	UNSPSC marker refills
22	UNSPSC highlighters
23	Eclass pen recorder and tubular drawing pen

24	UNSPSC wooden pencils
25	Eclass accessories for drop action pencil



**Figura 69: Mapping sul dominio PEN**

In Figura 69 sono evidenziate le corrispondenze che MOMIS ha individuato tra i due frammenti esaminati, senza l'introduzione di relazioni aggiuntive da parte del progettista e evitando l'introduzione di forme base e significati non perfettamente calzanti. I mapping individuati nell'ultimo livello sono pochi, quindi, può risultare utile un approfondimento mediante l'introduzione di relazioni da parte del progettista e l'ulteriore ricerca di forme base accettate dal sistema.



### 7.4.3 Integrazione dei frammenti relativi a Food

La scelta di effettuare l'integrazione dei codici di prodotto ([ecl@ss](#), UNSPSC, NAICS) relativi all'argomento Food deriva dalla considerazione che, anche i termini ad un alto livello di specificità, appartenenti a questo dominio, sono termini di uso comune e dovrebbero essere presenti WordNet. Questa integrazione è un esempio di approccio intermedio tra il Bottom-up e il Top down, la scelta di mixare i due approcci è stata imposta dalle caratteristiche del dominio. Per individuare i frammenti dei codici che descrivono questo tipo di prodotti sono stati utilizzati, come nell'approccio Bottom-up, i motori di ricerca specifici dei sistemi di codifica, inserendo vari termini inerenti all'argomento si è riusciti a risalire ai frammenti di codice cercati. I tre frammenti di codice analizzati sono strutturati in modo molto diverso: la posizione assegnata ai prodotti alimentari nella gerarchia è molto differente, allo stesso modo, differiscono le scelte d'aggregazione o di separazione dei prodotti in una o più voci.

In UNSPSC, il frammento di codice interessato al match prende origine da FOOD BEVERAGE AND TOBACCO PRODUCTS (segment 50). Si è scelto di considerare solo i due livelli immediatamente successivi (family e class) trascurando l'ultimo e più dettagliato livello (commodity). Il grado di dettaglio del livello commodity sarebbe stato eccessivo rispetto alle informazioni fornite dagli altri codici, non sarebbe, quindi, stato possibile ottenere un match valido con queste ulteriori voci. L'alto livello di specificità della classificazione avrebbe comportato molte difficoltà nell'assegnazione dei significati con WordNet.

In NAICS settore selezionato è FOOD MANUFACTURING (311). Dal nome si percepisce immediatamente il diverso punto di vista di questo sistema di codifica che, come ripetutamente sottolineato, non è orientato ai prodotti ma alle attività produttive. Per ottenere un match con gli altri codici si è deciso di assegnare significati che privilegiassero il prodotto rispetto all'attività produttiva. Il ramo di codice derivante da 311 non è stato interamente considerato, sono state scelte solo quelle parti che si ritenevano pertinenti con gli altri codici. Come nel caso precedente, non si è scesi al massimo livello di dettaglio, partendo dal livello contrassegnato con 311 si è arrivati fino al livello contrassegnato con 311XX.

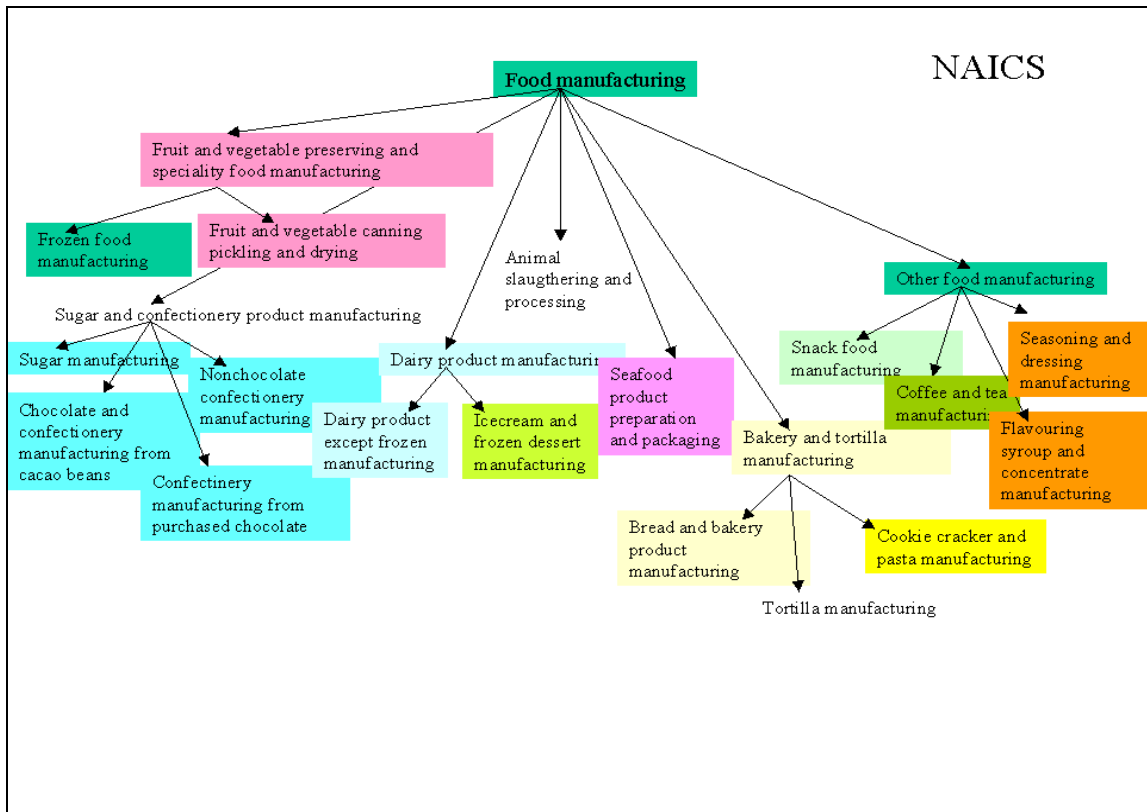
È importante notare che in molti casi il livello successivo fornisce la stessa descrizione del padre: la descrizione di 311XXY è spesso coincidente con quella di 311XX.

All'interno del sistema di codifica [ecl@ss](#), i prodotti alimentari sono classificati a partire dal secondo livello; il gruppo principale che li riguarda è FOOD AND LUXURY GOODS (29.04.00.00). Tutte le voci considerate appartengono al terzo livello (gruppo), il livello successivo (classe dei beni) offre la stessa descrizione del gruppo, di conseguenza non si è ritenuto opportuno il suo inserimento.

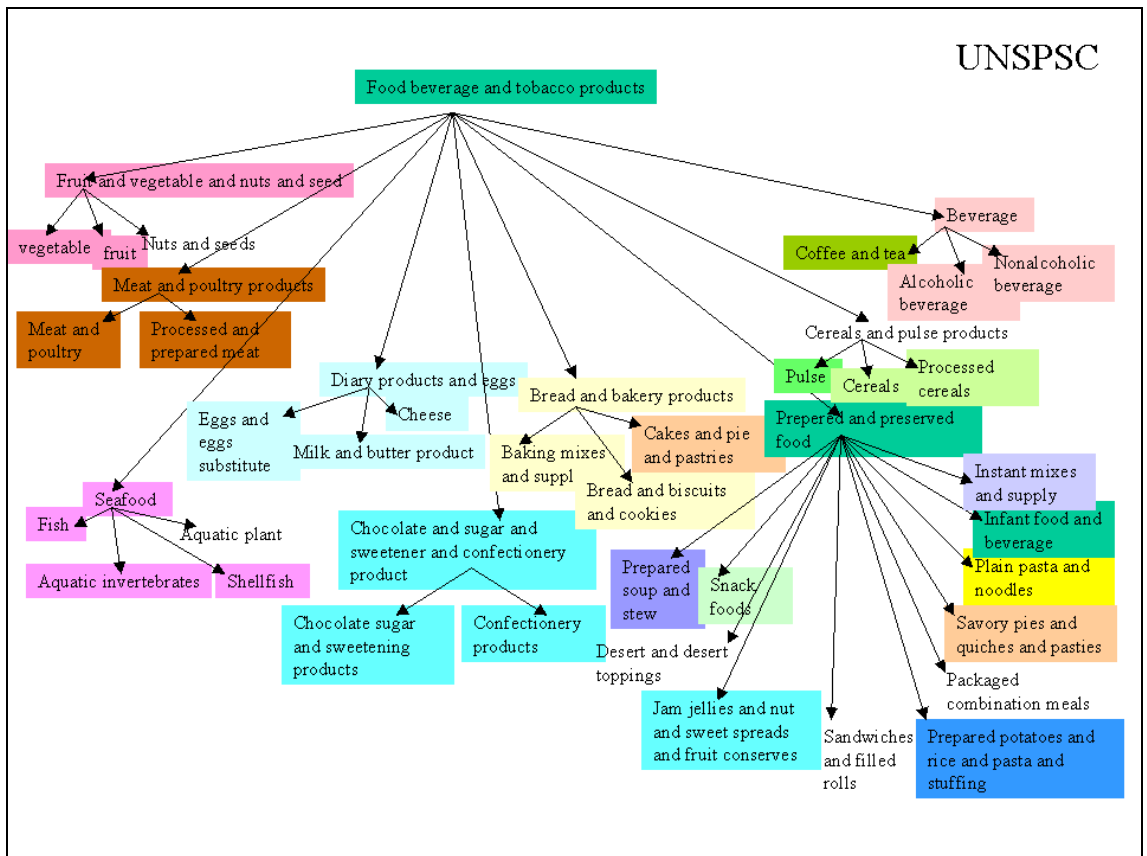
Nell'assegnare le forme base alle varie voci, sono state fatte delle scelte che hanno fortemente condizionato il risultato finale. Per esempio, in alcuni casi era presente la categoria "fruit and vegetable", in altri "fruit" e in altri "vegetable"; WordNet non permette l'inserimento di due termini per assegnare significato ad una voce, quindi, "fruit and vegetable" dovrebbe essere descritto individuando un termine che comprenda sia "fruit" che "vegetable". L'alternativa è inserire un termine comune per tutti i codici che riguardano "fruit", "vegetable" o entrambi. In questo caso si è deciso di mettere come termine comune fruit, quindi, può capitare di trovare assegnato alla voce vegetable il significato "fruit".

La conseguenza di questa scelta è che le voci "fruit", "vegetable" e "fruit and vegetable" vengono considerate come sinonimi, visto che è stata loro assegnata la stessa forma base e, conseguentemente, lo stesso significato. La scelta è molto discutibile, ma trova giustificazione nel fatto che in altro modo difficilmente si sarebbe ottenuto un mapping completo. L'unica alternativa poteva essere l'inserimento manuale delle relazioni ma un'operazione di questo tipo è molto impegnativa e porterebbe all'ottenimento degli stessi risultati.

Il diverso livello di dettaglio dei sistemi di codifica impone all'utente la generalizzazione dei concetti che non sono dettagliati in modo analogo nei vari sistemi di codifica. È il caso della categoria "fish" che viene dettagliata da [ecl@ss](#) e da UNSPSC in due modi molto differenti che non lasciano molte possibilità di match. La soluzione adottata in questo caso è ridurre il livello di dettaglio e assegnare, a tutte le voci di questo dominio, la forma generica "fish". Ciò fa sì che si costruisca un'unica classe globale relativa a "fish", invece di ottenere tante classi contenenti ognuna una sola voce. Nelle pagine successive vengono riportati gli schemi dei frammenti di codice integrati, gli elementi uniti dal mapping sono stati evidenziati con lo stesso colore in tutti gli schemi. Si nota che la maggior parte delle voci è stato trovato un corrispondente negli altri frammenti, anche se tale risultato ha richiesto, in molti casi, una eccessiva generalizzazione il mapping è complessivamente buono.



**Figura 70: NAICS –FOOD**



**Figura 71: UNSPSC-FOOD**

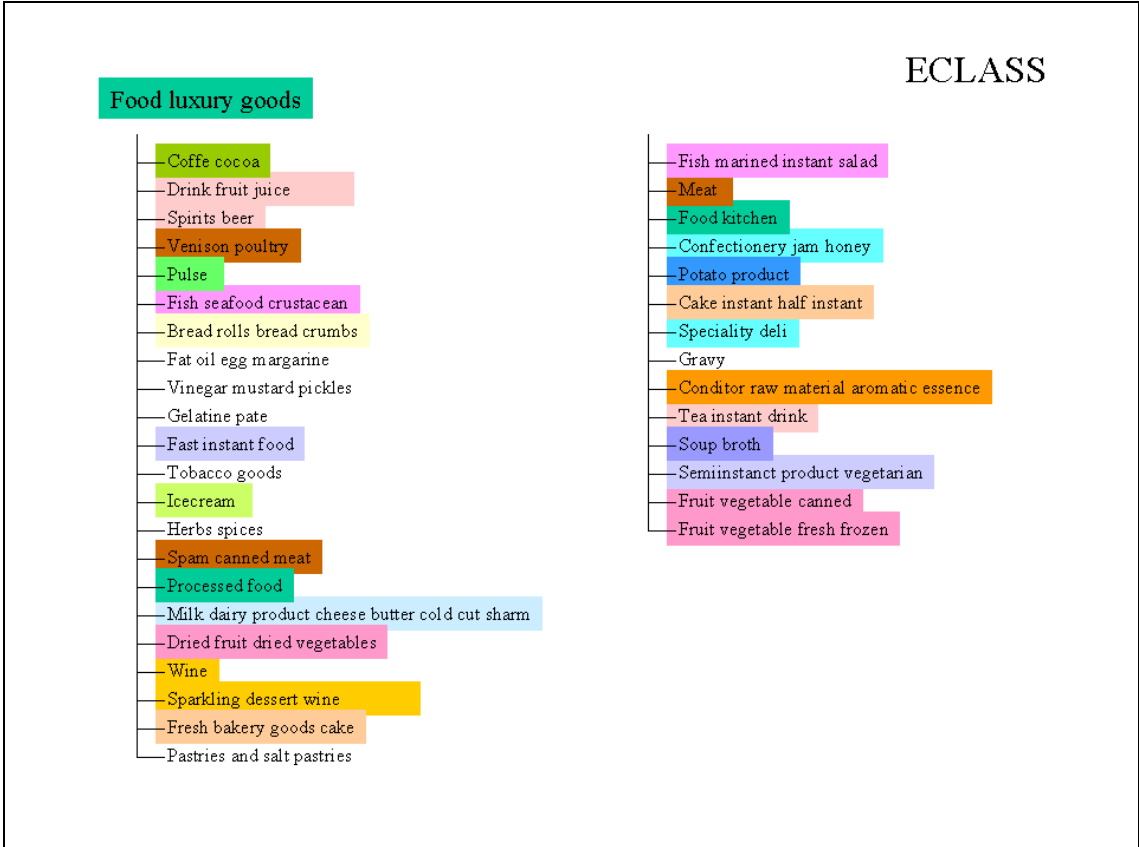


Figura 72: ecl@ss -FOOD

#### 7.4.4 Integrazione dei frammenti relativi a Farm

Questo mapping è stato effettuato su frammenti di codice relativi alle attività agricole e di allevamento. La scelta dell'argomento è stata guidata dal fatto che sia NAICS che UNSPSC lo classificano al primo livello della gerarchia e la similitudine tra le due codifiche è parsa abbastanza evidente. Si deve sottolineare che [ecl@ss](mailto:ecl@ss) non classifica questo dominio di prodotti \ attività, o meglio, anche se alcuni di questi prodotti vengono classificati, il punto di vista è molto diverso, per cui non sono raggruppati in uno stesso ramo di codice e sono attribuiti a vari domini. L'approccio di selezione dei frammenti utilizzato è, quindi, di tipo Top-down.

I punti di vista dei due sistemi di classificazione sono molto diversi, infatti, NAICS è orientato verso le attività produttive mentre UNSPSC verso i prodotti. S'intuisce immediatamente che "l'allevamento di animali" e gli "animali" sono concetti molto diversi ed è molto difficile individuare un match tra loro. Procedendo con una lettura rigorosa dei codici non si sarebbe riusciti a effettuare nessuna integrazione tra i frammenti. Si è, quindi, cercato di abbandonare i punti di vista iniziali dei sistemi di classificazione per ottenere il mapping. Sono stati esaminati i primi tre livelli di codifica, in quanto un ulteriore approfondimento sarebbe stato difficoltoso a causa dell'imponente mole di dati e delle difficoltà nella gestione di termini molto specifici. Scendere di un ulteriore livello nella gerarchia avrebbe sicuramente portato ad individuare qualche corrispondenza aggiuntiva, ma avrebbe, allo stesso tempo, apportato un notevole numero di voci di difficile integrazione.

Il settore di NAICS selezionato è classificato come:

**11\_Agriculture, forestry, fishing and hunting.**

Il segmento di UNSPSC selezionato è:

**10\_Live plant, animal material and accessories and supplies.**

Come in tutte le altre prove si nota (Figure 73, 74) la differenza del livello di dettaglio dei due frammenti e dei sistemi di codifica in generale. UNSPSC, infatti, è molto più approfondito e completo di NAICS, si coglie subito la sua predisposizione alla gestione delle forniture, mentre NAICS è, evidentemente, più appropriato per studi di settore e analisi statistiche.

Il dominio in questione è molto ampio, ci sono vari argomenti che vengono considerati da uno solo dei due sistemi di codifica, in questo caso non è possibile ottenere un match.

Si è tentato di assegnare ad ogni termine il significato più aderente possibile a quello reale, il risultato non è stato soddisfacente, data la diversità tra i codici, infatti, si sono ottenuti pochissimi match. È ovvio che non è un problema del sistema MOMIS ma insito nei sistemi di classificazione. A favore di un'integrazione più consistente, si è scelto di assegnare alle varie voci dei significati più generici, in grado di delineare dei sotto-domini. Ad esempio, sono stati assegnati dei significati come floriculture, farming, crop, seed, animal a tutte le voci di codice inerenti con queste attività.

Ciò comporta la parziale perdita di ricchezza delle codifiche, ma permette di creare delle classi globali per ognuno dei sotto-argomenti che raggruppano vari codici provenienti dai due standard di classificazione.

Il risultato che si ottiene in questo modo si discosta dallo scopo iniziale, cioè trovare per ogni codice di UNSPSC un codice analogo in NAICS o [ecl@ss](#), ma porta, comunque, ad un legame tra i due sistemi di codifica. Si ottengono quindici Global Classes, ognuna delle quali raggruppa i codici relativi ad un argomento.

Nelle immagini riportate in seguito, le diverse classi globali sono evidenziate con colori differenti, la prima cosa che si nota è che ci sono classi globali composte interamente da codici appartenenti alla stessa codifica. Queste classi non sono utili al fine di individuare un mapping, sono concettualmente corrette, in quanto conseguenza delle generalizzazioni introdotte, ma praticamente non necessarie.

In casi come quello in esame, il contributo del sistema nell'integrazione è minimo, lo sforzo maggiore è del progettista che deve avere una visione globale dei frammenti coi quali sta lavorando e individuare dei potenziali collegamenti tra le codifiche. Se il progettista non tenesse sempre in considerazione la struttura e il contenuto delle codifiche, non potrebbe intervenire per favorire la formazione di classi globali complesse come quelle in questione.

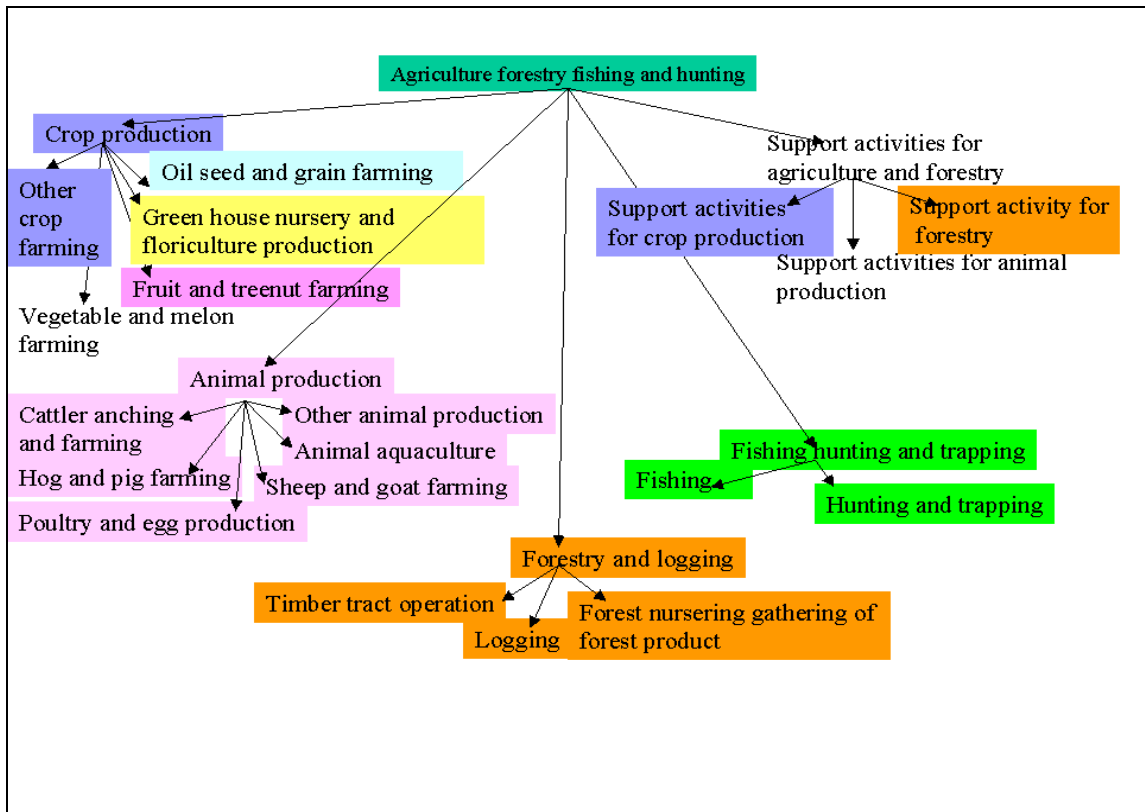


Figura 73: NAICS-FARM

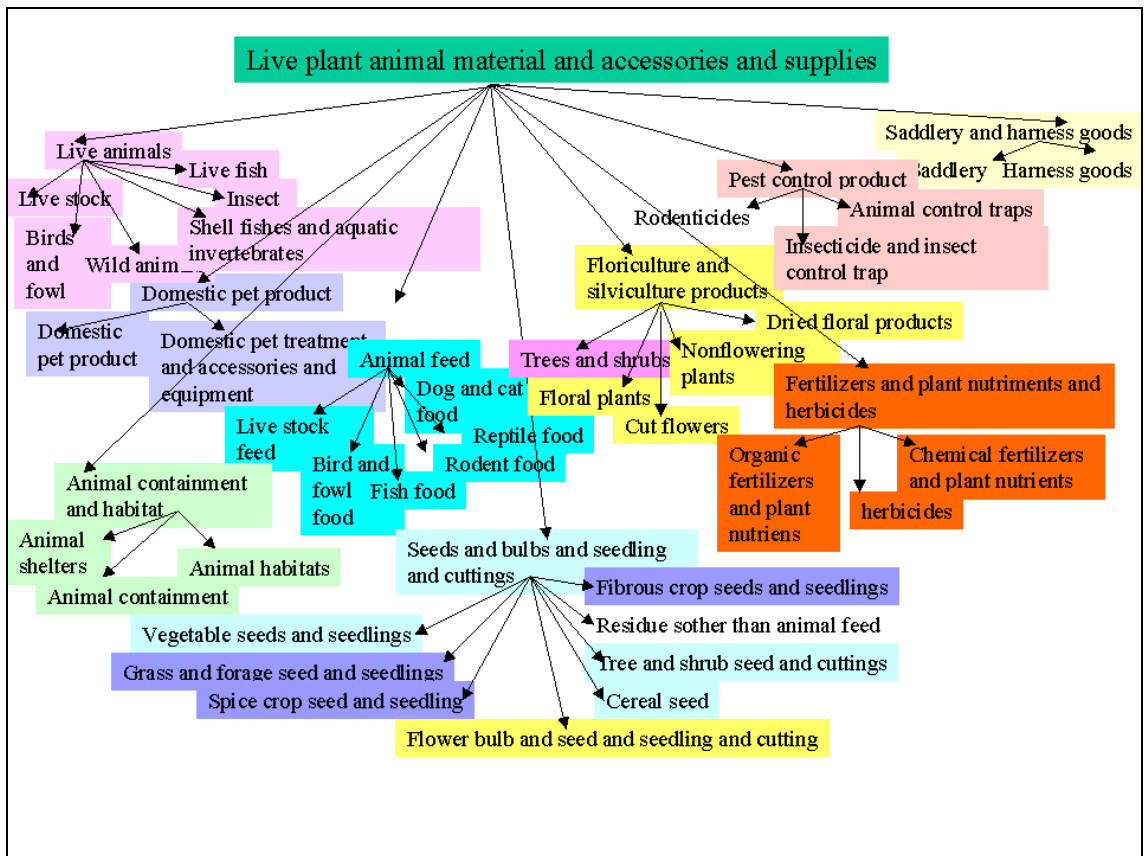


Figura 74: UNSPSC- FARM





## Conclusioni

L'argomento trattato nella tesi si colloca nell'ambito dell'integrazione degli standard di classificazione dei prodotti per l'e-commerce; l'obiettivo è stato quello di svolgere un'analisi critica della validità di un sistema a mediatore, quale il sistema MOMIS, in tale attività. Si è partiti con un'analisi delle peculiarità dell'e-commerce e in particolare degli e-marketplace, che, mediando tra tanti diversi compratori e venditori, si trovano quotidianamente di fronte alla necessità di far comunicare utenti che utilizzano linguaggi diversi. L'analisi di tali realtà ha fornito uno spunto valido per affrontare il problema dell'integrazione in modo più realistico. Il problema affrontato è fondamentale per il funzionamento degli e-marketplace, ma non ha mai ricevuto molta attenzione. Ci sono molti studi focalizzati sull'integrazione dei documenti per l'e-commerce e pochissimi che si occupano delle problematiche dell'integrazione degli standard di classificazione dei prodotti.

Sono state esaminate in particolare tre codifiche: UNSPSC, [Ecl@ss](#) e NAICS. I tre sistemi di classificazione sono estremamente eterogenei e hanno comportato molte difficoltà. È stato utilizzato il sistema MOMIS, un sistema a mediatore preposto all'integrazione di sorgenti di dati eterogenee.

L'analisi effettuata ha messo in evidenza che, nel contesto dell'integrazione degli standard di classificazione, le attività che si possono svolgere in maniera automatica sono limitate, e, quindi, è preponderante il ruolo del progettista nel corso dell'integrazione. Il progettista ha, infatti, la possibilità di decidere che risultati vuole ottenere:

- Corrispondenze perfette tra i codici, quindi, solo match che sottendono una relazione di sinonimia.
- Corrispondenze generiche tra i codici che, quindi, includono anche match che sottendono una relazione BT/NT.

Una volta individuato il suo obiettivo il progettista farà delle scelte coerenti con esso e l'integrazione risulterà soddisfacente. Ci sono due circostanze in cui il progettista ha l'opportunità di pilotare il mapping:

- assegnazione delle forme base, tramite la quale si assegna un significato ai termini utilizzati negli standard di classificazione;
- inserimento delle relazioni, tramite la quale si assegna esplicitamente un mapping tra termini utilizzati negli standard di classificazione.

In questi due momenti si possono, infatti, effettuare scelte in grado di influenzare fortemente i risultati dell'integrazione. Da un lato le forme base influenzano le relazioni che il sistema crea autonomamente tra le sorgenti immesse; mentre l'immissione di relazioni da parte del progettista va ad influenzare direttamente lo schema globale che il sistema creerà. Dall'analisi dei risultati ottenuti si può concludere che il sistema è estremamente flessibile e risponde perfettamente alle esigenze del progettista. La maggior parte dei problemi incontrati, riconducibili a mancanze del sistema, derivano dalla genericità di WordNet.

I test hanno mostrato che molte delle difficoltà che si incontrano in fase di integrazione, si possono attribuire alla struttura delle codifiche. Gli standard, infatti, sono estremamente eterogenei ed è stato molto difficoltoso anche solo individuare dei frammenti per lo svolgimento delle prove. Le codifiche sono progettate con scopi molto diversi, quindi non è possibile intuirne una perfetta sovrapposizione. Se si analizzano attentamente, ci si rende conto che, oltre ad essere incomplete e approssimative, fanno un uso parziale delle potenzialità offerte dalla struttura gerarchica che le caratterizza.

In conclusione, l'analisi effettuata ha individuato le funzionalità del sistema MOMIS che devono essere modificate per adattarlo a questa tipologia di integrazioni, allo scopo di ottenere dei miglioramenti sia dal punto di vista della rapidità che della modalità di svolgimento delle operazioni fondamentali.

# Appendice A

Nelle pagine successive sono riportati i primi livelli degli standard di classificazione di prodotto che sono oggetto della tesi.

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	Semifinished products, materials
36000000	Machine, apparatus

37000000	Pipeline technology
38000000	Chemical product (inorganic)
39000000	Chemical product (organic)
40000000	Occupational safety, accident prevention
41000000	Advertising

UNSPSC v.5:

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	Defense 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 Jewelry 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 Defense and Public Order and Security and Safety Services
93000000	Politics and Civic Affairs Services
94000000	Organizations and Clubs

NAICS:

<b>Class</b>	<b>Preferred Name</b>
11	Agriculture, Forestry, Fishing and Hunting
21	Mining
22	Utilities
23	Construction

---

31-33	Manufacturing
42	Wholesale Trade
44-45	Retail Trade
48-49	Transportation and Warehousing
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

---





## Appendice B

Si riporta il file XML relativo allo stato salvato dell'integrazione sul dominio "Toy", questo file permette di ripristinare il lavoro svolto mediante l'interfaccia SI-Designer o di apportare modifiche. In particolare sono evidenziati i frammenti di codice richiamati nel capitolo 7.

```
<?xml version="1.0" ?>
- <!-- Build file for application -->
- <Schemaname="">
-
<SourceName="etoy_xml" type="semistructured" wrapperName="etoy" hostName="sparc20.ing.unimo.it" portNumber="1050">
- <InterfaceName="stuffedanimal" persistent="false">
  <extent name="stuffedanimal" />
- <AttributeName="cod" type="string">
- <AttributeAdditionalInfo>
- <SlimNodeFormaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeFormaBase="stuffed">
  <SlimNodeSenso numero="1" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
- <InterfaceName="ball" persistent="false">
  <extent name="ball" />
- <AttributeName="cod" type="string">
- <AttributeAdditionalInfo>
- <SlimNodeFormaBase="cod">
  <SlimNodeSenso numero="-2" />
```

```

</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="ball">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="balloon"persistent="false">
  <extent name="balloon" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="ball">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="toy"persistent="false">
  <extent name="toy" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="toy">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="buildingblock"persistent="false">
  <extent name="buildingblock" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="building_block">
  <SlimNodeSenso numero="2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="promotionalarticle"persistent="false">
  <extent name="promotionalarticle" />

```

```

-
<Attributename="promgiftschildrenplayingandfun"type="promgiftsch
ildrenplayingandfun">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="gift">
  <SlimNodeSenso numero="1" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="promotion">
  <SlimNodeSenso numero="1" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
- <Interfacename="drawingpens"persistent="false">
  <extent name="drawingpens" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="pen">
  <SlimNodeSenso numero="1" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
- <Interfacename="advertising"persistent="false">
  <extent name="advertising" />
- <Attributename="promotionalarticle"type="promotionalarticle">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="promotionalarticle">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="advertising">
  <SlimNodeSenso numero="1" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
-
<Interfacename="promgiftschildrenplayingandfun"persistent="false
">
  <extent name="promgiftschildrenplayingandfun" />
- <Attributename="stuffedanimal"type="stuffedanimal">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="stuffed">
  <SlimNodeSenso numero="1" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>

```

```

- <Attributename="ball"type="ball">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="ball">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="balloon"type="balloon">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="ball">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="toy"type="toy">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toy">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="buildingblock"type="buildingblock">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="building_block">
  <SlimNodeSenso numero="2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="drawingpens"type="drawingpens">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="pen">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="gift">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
</Source>
-
<Sourcename="utoy_xml"type="semistructured"wrapperName="utoy"hostName="sparc20.ing.unimo.it"portNumer="1050">
- <Interfacename="toysandgames"persistent="false">
  <extent name="toysandgames" />
- <Attributename="toys"type="toys">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toy">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="game">

```

```

    <SlimNodeSenso numero="1" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="dollhouses"persistent="false">
  <extent name="dollhouses" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="doll">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
- <Interfacename="buildingblocks"persistent="false">
  <extent name="buildingblocks" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="building_block">
  <SlimNodeSenso numero="2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
- <Interfacename="pulltoys"persistent="false">
  <extent name="pulltoys" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="pull_toy">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
</Interface>
- <Interfacename="stuffedanimalsorpuppets"persistent="false">
  <extent name="stuffedanimalsorpuppets" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>

```

```

    </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="stuffed">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="toys"persistent="false">
  <extent name="toys" />
- <Attributename="toyballoonsorballs"type="toyballoonsorballs">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="ball">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="dollhouses"type="dollhouses">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="doll_houses">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
-
<Attributename="childrenssciencekits"type="childrenssciencekits"
>
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="childrenssciencekits">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="inflatabletoys"type="inflatabletoys">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="inflatabletoys">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="buildingblocks"type="buildingblocks">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="building_block">
  <SlimNodeSenso numero="2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="ridingtoys"type="ridingtoys">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="ridingtoys">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="pulltoys"type="pulltoys">
- <AttributeAdditionalInfo>

```

```

- <SlimNodeformaBase="pulltoys">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="toytrains"type="toytrains">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toy_train">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
-
<Attributename="stuffedanimalsorpuppets"type="stuffedanimalsorpu
ppets">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="stuffed">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="educationaltoys"type="educationaltoys">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="educationaltoys">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="dolls"type="dolls">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="doll">
  <SlimNodeSenso numero="1" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="playhouses"type="playhouses">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="playhouses">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="musicaltoys"type="musicaltoys">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="musicaltoys">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <Attributename="toyvehicles"type="toyvehicles">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toyvehicles">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>

```

```

- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="toy">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="toyballoonsorballs"persistent="false">
  <extent name="toyballoonsorballs" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="ball">
  <SlimNodeSenso numero="1" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="childrenssciencekits"persistent="false">
  <extent name="childrenssciencekits" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="childrenssciencekits">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="inflatabletoys"persistent="false">
  <extent name="inflatabletoys" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="inflatabletoys">
  <SlimNodeSenso numero="-2" />
</SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="ridingtoys"persistent="false">
  <extent name="ridingtoys" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">

```



```

    <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="ridingtoys">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="toytrains"persistent="false">
  <extent name="toytrains" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="toy_train">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="educationaltoys"persistent="false">
  <extent name="educationaltoys" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="educationaltoys">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>
- <Interfacename="dolls"persistent="false">
  <extent name="dolls" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="doll">
  <SlimNodeSenso numero="1" />
  </SlimNode>
</InterfaceAdditionalInfo>
</Interface>

```

```

-
<Interfacename="musicalinstrumentsandrecreationalequipmentandsup
pliesandaccessories"persistent="false">
  <extent
name="musicalinstrumentsandrecreationalequipmentandsuppliesandac
cessories" />
- <Attributename="toysandgames"type="toysandgames">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="toysandgames">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
-
<SlimNodeformaBase="musicalinstrumentsandrecreationalequipmentan
dsuppliesandaccessories">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
- <Interfacename="playhouses"persistent="false">
  <extent name="playhouses" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="playhouses">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
- <Interfacename="musicaltoys"persistent="false">
  <extent name="musicaltoys" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </AttributeAdditionalInfo>
  </Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="musicaltoys">
  <SlimNodeSenso numero="-2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
- <Interfacename="toyvehicles"persistent="false">
  <extent name="toyvehicles" />
- <Attributename="cod"type="string">
- <AttributeAdditionalInfo>
- <SlimNodeformaBase="cod">

```

```

    <SlimNodeSensio numero="-2" />
  </SlimNode>
</AttributeAdditionalInfo>
</Attribute>
- <InterfaceAdditionalInfo>
- <SlimNodeformaBase="toyvehicles">
  <SlimNodeSensio numero="-2" />
  </SlimNode>
  </InterfaceAdditionalInfo>
  </Interface>
  </Source>
- <ThesRelations>
  <TRInterfaceRel
attributel="utoy_xml.toys"relType="rt"attribute2="utoy_xml.toysa
ndgames"validated="true"extensional="false"class="InterfaceRel"p
roducerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.buildingblocks"relType="rt"attribute2="utoy
_xml.toys"validated="true"extensional="false"class="InterfaceRel
"producerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.childrenssciencekits"relType="rt"attribute2
="utoy_xml.toys"validated="true"extensional="false"class="Interf
aceRel"producerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.dollhouses"relType="rt"attribute2="utoy_xml
.toys"validated="true"extensional="false"class="InterfaceRel"pro
ducerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.educationaltoys"relType="rt"attribute2="uto
y_xml.toys"validated="true"extensional="false"class="InterfaceRe
l"producerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.inflatabletoys"relType="rt"attribute2="utoy
_xml.toys"validated="true"extensional="false"class="InterfaceRel
"producerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.musicaltoys"relType="rt"attribute2="utoy_xm
l.toys"validated="true"extensional="false"class="InterfaceRel"pr
oducerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.playhouses"relType="rt"attribute2="utoy_xml
.toys"validated="true"extensional="false"class="InterfaceRel"pro
ducerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.pulltoys"relType="rt"attribute2="utoy_xml.t
oys"validated="true"extensional="false"class="InterfaceRel"produ
cerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.ridingtoys"relType="rt"attribute2="utoy_xml
.toys"validated="true"extensional="false"class="InterfaceRel"pro
ducerId="900" />
  <TRInterfaceRel
attributel="utoy_xml.stuffedanimalsorpuppets"relType="rt"attribu
te2="utoy_xml.toys"validated="true"extensional="false"class="Int
erfaceRel"producerId="900" />

```

```

    <TRInterfaceRel
    attributel="utoy_xml.toyballoonsorballs"relType="rt"attribute2="
    utoy_xml.toys"validated="true"extensional="false"class="Interfac
    eRel"producerId="900" />
    <TRInterfaceRel
    attributel="utoy_xml.toytrains"relType="rt"attribute2="utoy_xml.
    toys"validated="true"extensional="false"class="InterfaceRel"prod
    ucerId="900" />
    <TRInterfaceRel
    attributel="utoy_xml.toyvehicles"relType="rt"attribute2="utoy_xm
    l.toys"validated="true"extensional="false"class="InterfaceRel"pr
    oducerId="900" />
    <TRInterfaceRel
    attributel="utoy_xml.toysandgames"relType="rt"attribute2="utoy_x
    ml.musicalinstrumentsandrecreationalequipmentandsuppliesandacces
    sories"validated="true"extensional="false"class="InterfaceRel"pr
    oducerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.promgiftschildrenplayingandfun"relType="rt"
    attribute2="etoy_xml.promotionalarticle"validated="true"extensio
    nal="false"class="InterfaceRel"producerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.ball"relType="rt"attribute2="etoy_xml.promg
    iftschildrenplayingandfun"validated="true"extensional="false"cla
    ss="InterfaceRel"producerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.balloon"relType="rt"attribute2="etoy_xml.pr
    omgiftschildrenplayingandfun"validated="true"extensional="false"
    class="InterfaceRel"producerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.buildingblock"relType="rt"attribute2="etoy_
    xml.promgiftschildrenplayingandfun"validated="true"extensional="
    false"class="InterfaceRel"producerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.drawingpens"relType="rt"attribute2="etoy_xm
    l.promgiftschildrenplayingandfun"validated="true"extensional="fa
    lse"class="InterfaceRel"producerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.stuffedanimal"relType="rt"attribute2="etoy_
    xml.promgiftschildrenplayingandfun"validated="true"extensional="
    false"class="InterfaceRel"producerId="900" />
    <TRInterfaceRel
    attributel="etoy_xml.toy"relType="rt"attribute2="etoy_xml.promgi
    ftschildrenplayingandfun"validated="true"extensional="false"clas
    s="InterfaceRel"producerId="900" />
    <TRAttrIntRel
    attributel="etoy_xml.stuffedanimal"relType="syn"attribute2="etoy
    _xml.promgiftschildrenplayingandfun.stuffedanimal"validated="tru
    e"extensional="false"class="AttrIntRel"producerId="740" />
    <TRInterfaceRel
    attributel="etoy_xml.stuffedanimal"relType="syn"attribute2="utoy
    _xml.stuffedanimalorsoppuppets"validated="true"extensional="false"
    class="InterfaceRel"producerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.stuffedanimal"relType="syn"attribute2="utoy

```

```

_xml.toy.stuffedanimalsorpuppets"validated="true"extensional="false"
class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attribute1="utoy_xml.stuffedanimalsorpuppets"relType="syn"attribute2="etoy_xml.promgiftschildrenplayingandfun.stuffedanimal"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRAttributeRel
attribute1="etoy_xml.promgiftschildrenplayingandfun.stuffedanimal"relType="syn"attribute2="utoy_xml.toy.stuffedanimalsorpuppets"validated="false"extensional="false"class="AttributeRel"producerId="740" />
  <TRAttrIntRel
attribute1="utoy_xml.stuffedanimalsorpuppets"relType="syn"attribute2="utoy_xml.toy.stuffedanimalsorpuppets"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attribute1="utoy_xml.dolls"relType="syn"attribute2="utoy_xml.toy.s.dolls"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attribute1="etoy_xml.toy"relType="bt"attribute2="utoy_xml.toy.dolls"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRAttributeRel
attribute1="utoy_xml.toy.dolls"relType="nt"attribute2="etoy_xml.promgiftschildrenplayingandfun.toy"validated="false"extensional="false"class="AttributeRel"producerId="740" />
  <TRAttributeRel
attribute1="utoy_xml.toy.dolls"relType="nt"attribute2="utoy_xml.toy.sandgames.toy"validated="false"extensional="false"class="AttributeRel"producerId="740" />
  <TRAttrIntRel
attribute1="utoy_xml.toy"relType="bt"attribute2="utoy_xml.toy.dolls"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRInterfaceRel
attribute1="utoy_xml.dolls"relType="nt"attribute2="etoy_xml.toy"validated="true"extensional="false"class="InterfaceRel"producerId="740" />
  <TRAttrIntRel
attribute1="utoy_xml.dolls"relType="nt"attribute2="etoy_xml.promgiftschildrenplayingandfun.toy"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attribute1="utoy_xml.dolls"relType="nt"attribute2="utoy_xml.toy.sandgames.toy"validated="true"extensional="false"class="AttrIntRel"producerId="740" />
  <TRInterfaceRel
attribute1="utoy_xml.dolls"relType="nt"attribute2="utoy_xml.toy"validated="true"extensional="false"class="InterfaceRel"producerId="740" />
  <TRAttrIntRel
attribute1="etoy_xml.toy"relType="syn"attribute2="etoy_xml.promgiftschildrenplayingandfun.toy"validated="true"extensional="false"class="AttrIntRel"producerId="740" />

```

```

    <TRAttrIntRel
    attributel="etoy_xml.toy"relType="syn"attribute2="utoy_xml.toysa
    ndgames.toys"validated="true"extensional="false"class="AttrIntRe
    l"producerId="740" />
    <TRInterfaceRel
    attributel="etoy_xml.toy"relType="syn"attribute2="utoy_xml.toys"
    validated="true"extensional="false"class="InterfaceRel"producerI
    d="740" />
    <TRAttributeRel
    attributel="etoy_xml.promgiftschildrenplayingandfun.toy"relType=
    "syn"attribute2="utoy_xml.toysandgames.toys"validated="false"ext
    ensional="false"class="AttributeRel"producerId="740" />
    <TRAttrIntRel
    attributel="utoy_xml.toys"relType="syn"attribute2="etoy_xml.prom
    giftschildrenplayingandfun.toy"validated="true"extensional="fals
    e"class="AttrIntRel"producerId="740" />
    <TRAttrIntRel
    attributel="utoy_xml.toys"relType="syn"attribute2="utoy_xml.toys
    andgames.toys"validated="true"extensional="false"class="AttrIntR
    el"producerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.promgiftschildrenplayingandfun"relType="syn
    "attribute2="etoy_xml.promotionalarticle.promgiftschildrenplayin
    gandfun"validated="true"extensional="false"class="AttrIntRel"pro
    ducerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.drawingpens"relType="syn"attribute2="etoy_x
    ml.promgiftschildrenplayingandfun.drawingpens"validated="true"ex
    tentional="false"class="AttrIntRel"producerId="740" />
    <TRInterfaceRel
    attributel="etoy_xml.advertising"relType="nt"attribute2="etoy_xm
    l.promotionalarticle"validated="true"extensional="false"class="I
    nterfaceRel"producerId="740" />
    <TRInterfaceRel
    attributel="etoy_xml.ball"relType="syn"attribute2="etoy_xml.ball
    oon"validated="true"extensional="false"class="InterfaceRel"produ
    cerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.ball"relType="syn"attribute2="etoy_xml.prom
    giftschildrenplayingandfun.ball"validated="true"extensional="fal
    se"class="AttrIntRel"producerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.ball"relType="syn"attribute2="etoy_xml.prom
    giftschildrenplayingandfun.balloon"validated="true"extensional="
    false"class="AttrIntRel"producerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.ball"relType="syn"attribute2="utoy_xml.toys
    .toyballoonsorballs"validated="true"extensional="false"class="At
    trIntRel"producerId="740" />
    <TRInterfaceRel
    attributel="etoy_xml.ball"relType="syn"attribute2="utoy_xml.toyb
    alloonsorballs"validated="true"extensional="false"class="Interfa
    ceRel"producerId="740" />
    <TRAttrIntRel
    attributel="etoy_xml.balloon"relType="syn"attribute2="etoy_xml.p

```

```

romgiftschildrenplayingandfun.ball"validated="true"extensional="
false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attributel="etoy_xml.balloon"relType="syn"attribute2="etoy_xml.p
romgiftschildrenplayingandfun.balloon"validated="true"extensiona
l="false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attributel="etoy_xml.balloon"relType="syn"attribute2="utoy_xml.t
oys.toyballoonsorballs"validated="true"extensional="false"class=
"AttrIntRel"producerId="740" />
  <TRInterfaceRel
attributel="etoy_xml.balloon"relType="syn"attribute2="utoy_xml.t
oyballoonsorballs"validated="true"extensional="false"class="Inte
rfaceRel"producerId="740" />
  <TRAttributeRel
attributel="etoy_xml.promgiftschildrenplayingandfun.ball"relType
="syn"attribute2="etoy_xml.promgiftschildrenplayingandfun.balloo
n"validated="false"extensional="false"class="AttributeRel"produc
erId="740" />
  <TRAttributeRel
attributel="etoy_xml.promgiftschildrenplayingandfun.ball"relType
="syn"attribute2="utoy_xml.toys.toyballoonsorballs"validated="fa
lse"extensional="false"class="AttributeRel"producerId="740" />
  <TRAttrIntRel
attributel="utoy_xml.toyballoonsorballs"relType="syn"attribute2=
"etoy_xml.promgiftschildrenplayingandfun.ball"validated="true"ex
tensional="false"class="AttrIntRel"producerId="740" />
  <TRAttributeRel
attributel="etoy_xml.promgiftschildrenplayingandfun.balloon"relT
ype="syn"attribute2="utoy_xml.toys.toyballoonsorballs"validated=
"false"extensional="false"class="AttributeRel"producerId="740"
/>
  <TRAttrIntRel
attributel="utoy_xml.toyballoonsorballs"relType="syn"attribute2=
"etoy_xml.promgiftschildrenplayingandfun.balloon"validated="true
"extensional="false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attributel="utoy_xml.toys.toyballoonsorballs"relType="syn"attribu
te2="utoy_xml.toys.toyballoonsorballs"validated="true"extensional
="false"class="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attributel="etoy_xml.buildingblock"relType="syn"attribute2="etoy
_xml.promgiftschildrenplayingandfun.buildingblock"validated="tru
e"extensional="false"class="AttrIntRel"producerId="740" />
  <TRInterfaceRel
attributel="etoy_xml.buildingblock"relType="syn"attribute2="utoy
_xml.buildingblocks"validated="true"extensional="false"class="Int
erfaceRel"producerId="740" />
  <TRAttrIntRel
attributel="etoy_xml.buildingblock"relType="syn"attribute2="utoy
_xml.toys.buildingblocks"validated="true"extensional="false"clas
s="AttrIntRel"producerId="740" />
  <TRAttrIntRel
attributel="utoy_xml.buildingblocks"relType="syn"attribute2="eto
y_xml.promgiftschildrenplayingandfun.buildingblock"validated="tr
ue"extensional="false"class="AttrIntRel"producerId="740" />

```

```

    <TRAttributeRel
attribute1="etoy_xml.promgiftschildrenplayingandfun.buildingbloc
k"relType="syn"attribute2="utoy_xml.toys.buildingblocks"validate
d="false"extensional="false"class="AttributeRel"producerId="740"
/>
    <TRAttrIntRel
attribute1="utoy_xml.buildingblocks"relType="syn"attribute2="uto
y_xml.toys.buildingblocks"validated="true"extensional="false"cla
ss="AttrIntRel"producerId="740" />
</TheRelations>
- <GlobalClassname="Global0"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.ridingtoys">
  <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global1"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.inflatabletoys">
  <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global10"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.toytrains">
  <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global11"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="etoy_xml.drawingpens">
  <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global12"ifFused="true">
- <GlobalAttributename="ball"domain="complex"complex="false">
-
<LocalAttributelocalClass="etoy_xml.promgiftschildrenplayingandf
un">
  <SimpleMapping attributeName="ball" />
</LocalAttribute>
</GlobalAttribute>
- <GlobalAttributename="balloon"domain="complex"complex="false">
-
<LocalAttributelocalClass="etoy_xml.promgiftschildrenplayingandf
un">
  <SimpleMapping attributeName="balloon" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttributename="buildingblock"domain="complex"complex="fal
se">

```



```

-
<LocalAttribute localClass="etoy_xml.promgiftschildrenplayingandfun">
  <SimpleMapping attributeName="buildingblock" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="drawingpens" domain="complex" complex="false">
-
<LocalAttribute localClass="etoy_xml.promgiftschildrenplayingandfun">
  <SimpleMapping attributeName="drawingpens" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="stuffedanimal" domain="complex" complex="false">
-
<LocalAttribute localClass="etoy_xml.promgiftschildrenplayingandfun">
  <SimpleMapping attributeName="stuffedanimal" />
</LocalAttribute>
</GlobalAttribute>
- <GlobalAttribute name="toy" domain="complex" complex="false">
-
<LocalAttribute localClass="etoy_xml.promgiftschildrenplayingandfun">
  <SimpleMapping attributeName="toy" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClass name="Global13" ifFused="true">
- <GlobalAttribute name="toys" domain="complex" complex="false">
- <LocalAttribute localClass="utoy_xml.toysandgames">
  <SimpleMapping attributeName="toys" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClass name="Global14" ifFused="true">
-
<GlobalAttribute name="toysandgames" domain="complex" complex="false">
-
<LocalAttribute localClass="utoy_xml.musicalinstrumentsandrecreationalequipmentandsuppliesandaccessories">
  <SimpleMapping attributeName="toysandgames" />
</LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClass name="Global15" ifFused="true">
- <GlobalAttribute name="cod" domain="string" complex="false">
- <LocalAttribute localClass="utoy_xml.toyvehicles">
  <SimpleMapping attributeName="cod" />
</LocalAttribute>
</GlobalAttribute>

```

```

</GlobalClass>
- <GlobalClassname="Global16"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.musicaltoys">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global17"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.dollhouses">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global18"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="etoy_xml.ball">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
- <LocalAttributelocalClass="etoy_xml.balloon">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
- <LocalAttributelocalClass="utoy_xml.toyballoonsorballs">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global19"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.dolls">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global2"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.educationaltoys">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global20"ifFused="true">
-
<GlobalAttributename="buildingblocks"domain="complex"complex="false">
- <LocalAttributelocalClass="etoy_xml.toy">
  <NullMapping />
  </LocalAttribute>
- <LocalAttributelocalClass="utoy_xml.toys">
  <SimpleMapping attributeName="buildingblocks" />
  </LocalAttribute>
</GlobalAttribute>
-
<GlobalAttributename="childrenssciencekits"domain="complex"complex="false">

```

```

- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="childrenssciencekits" />
</LocalAttribute>
</GlobalAttribute>
- <GlobalAttribute name="cod" domain="string" complex="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <SimpleMapping attributeName="cod" />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <NullMapping />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="dollhouses" domain="complex" complex="false">
>
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="dollhouses" />
</LocalAttribute>
</GlobalAttribute>
- <GlobalAttribute name="dolls" domain="complex" complex="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="dolls" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="educationaltoys" domain="complex" complex="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="educationaltoys" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="inflatabletoys" domain="complex" complex="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="inflatabletoys" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="musicaltoys" domain="complex" complex="false">
">

```

```

- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="musicaltoys" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="playhouses" domain="complex" complex="false"
>
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="playhouses" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="pulltoys" domain="complex" complex="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="pulltoys" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="ridingtoys" domain="complex" complex="false"
>
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="ridingtoys" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="stuffedanimalsorpuppets" domain="complex" co
mplex="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="stuffedanimalsorpuppets" />
</LocalAttribute>
</GlobalAttribute>
-
<GlobalAttribute name="toyballoonsorballs" domain="complex" complex
="false">
- <LocalAttribute localClass="etoy_xml.toy">
  <NullMapping />
</LocalAttribute>
- <LocalAttribute localClass="utoy_xml.toys">
  <SimpleMapping attributeName="toyballoonsorballs" />
</LocalAttribute>
</GlobalAttribute>

```

```

-
<GlobalAttributename="toytrains"domain="complex"complex="false">
- <LocalAttributelocalClass="etoy_xml.toy">
  <NullMapping />
  </LocalAttribute>
- <LocalAttributelocalClass="utoy_xml.toys">
  <SimpleMapping attributeName="toytrains" />
  </LocalAttribute>
  </GlobalAttribute>
-
<GlobalAttributename="toyvehicles"domain="complex"complex="false
">
- <LocalAttributelocalClass="etoy_xml.toy">
  <NullMapping />
  </LocalAttribute>
- <LocalAttributelocalClass="utoy_xml.toys">
  <SimpleMapping attributeName="toyvehicles" />
  </LocalAttribute>
  </GlobalAttribute>
  </GlobalClass>
- <GlobalClassname="Global3"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.childrensciencekits">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  </GlobalAttribute>
  </GlobalClass>
- <GlobalClassname="Global4"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.playhouses">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
  </GlobalAttribute>
  </GlobalClass>
- <GlobalClassname="Global5"ifFused="true">
-
<GlobalAttributename="promgiftschildrenplayingandfun"domain="com
plex"complex="false">
- <LocalAttributelocalClass="etoy_xml.promotionalarticle">
  <SimpleMapping attributeName="promgiftschildrenplayingandfun"
/>
  </LocalAttribute>
  </GlobalAttribute>
  </GlobalClass>
- <GlobalClassname="Global6"ifFused="true">
-
<GlobalAttributename="promotionalarticle"domain="complex"complex
="false">
- <LocalAttributelocalClass="etoy_xml.advertising">
  <SimpleMapping attributeName="promotionalarticle" />
  </LocalAttribute>
  </GlobalAttribute>
  </GlobalClass>
- <GlobalClassname="Global7"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="etoy_xml.buildingblock">

```

```

    <SimpleMapping attributeName="cod" />
  </LocalAttribute>
- <LocalAttributelocalClass="utoy_xml.buildingblocks">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global8"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="utoy_xml.pulltoys">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
- <GlobalClassname="Global9"ifFused="true">
- <GlobalAttributename="cod"domain="string"complex="false">
- <LocalAttributelocalClass="etoy_xml.stuffedanimal">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
- <LocalAttributelocalClass="utoy_xml.stuffedanimalsorpuppets">
  <SimpleMapping attributeName="cod" />
  </LocalAttribute>
</GlobalAttribute>
</GlobalClass>
</Schema>

```

# Appendice C

Nelle pagine successive si riportano a scopo esemplificativo il listato del file XML e del corrispondente file DTD.

DTD: UNSPSC- “FOOD”

```
<!ELEMENT FoodBeverageandTobaccoProducts (
FruitsandVegetablesandNutsandSeeds | MeatandPoultryProducts | Seafood
| DairyProductsandEggs |
ChocolateandSugarsandSweetenersandConfectionaryProducts |
BreadandBakeryProducts | PreparedandPreservedFoods |
PreparedandPreservedFoods | CerealandPulseProducts | Beverages)>
<!ELEMENT FruitsandVegetablesandNutsandSeeds (Vegetables | Fruits |
NutsandSeeds)>
<!ELEMENT MeatandPoultryProducts ( MeatandPoultry |
ProcessedandPreparedMeats)>
<!ELEMENT Seafood ( Fish | Shellfish | AquaticInvertebrates |
AquaticPlants)>
<!ELEMENT DairyProductsandEggs ( EggsandEggSubstitutes |
MilkandButterProducts | Cheese)>
<!ELEMENT ChocolateandSugarsandSweetenersandConfectionaryProducts
(ChocolateandSugarsandSweeteningProducts | ConfectionaryProducts)>
<!ELEMENT BreadandBakeryProducts ( BakingMixesandSupplies |
BreadandBiscuitsandCookies | CakesandPiesandPastries)>
<!ELEMENT PreparedandPreservedFoods ( PreparedSoupsandStews |
SnackFoods | DesertsandDesertToppings |
JamsandJelliesandNutandSweetSpreadsandFruitConserves |
```

SandwichesandFilledRolls | PreparedPotatoesandRiceandPastaandStuffing  
| PackagedCombinationMeals | SavoryPiesandQuichesandPasties |  
PlainPastaandNoodles | InfantFoodsandBeverages |  
InstantMixesandSupplies)>  
<!ELEMENT Beverages ( CoffeeandTea | AlcoholicBeverages |  
NonAlcoholicBeverages)>  
<!ELEMENT CerealandPulseProducts ( Pulses | Cereals |  
ProcessedCereals)>  
<!ELEMENT Vegetables (cod)>  
<!ELEMENT Fruits (cod)>  
<!ELEMENT NutsandSeeds (cod)>  
<!ELEMENT MeatandPoultry (cod)>  
<!ELEMENT ProcessedandPreparedMeats (cod)>  
<!ELEMENT Fish (cod)>  
<!ELEMENT Shellfish (cod)>  
<!ELEMENT AquaticInvertebrates (cod)>  
<!ELEMENT AquaticPlants (cod)>  
<!ELEMENT EggsandEggSubstitutes (cod)>  
<!ELEMENT MilkandButterProducts (cod)>  
<!ELEMENT Cheese (cod)>  
<!ELEMENT ChocolateandSugarsandSweeteningProducts (cod)>  
<!ELEMENT ConfectionaryProducts (cod)>  
<!ELEMENT BakingMixesandSupplies (cod)>  
<!ELEMENT BreadandBiscuitsandCookies (cod)>  
<!ELEMENT CakesandPiesandPastries (cod)>  
<!ELEMENT PreparedSoupsandStews (cod)>  
<!ELEMENT SnackFoods (cod)>  
<!ELEMENT DesertsandDesertToppings (cod)>  
<!ELEMENT JamsandJelliesandNutandSweetSpreadsandFruitConserves (cod)>  
<!ELEMENT SandwichesandFilledRolls (cod)>  
<!ELEMENT PreparedPotatoesandRiceandPastaandStuffing (cod)>  
<!ELEMENT PackagedCombinationMeals (cod)>  
<!ELEMENT SavoryPiesandQuichesandPasties (cod)>  
<!ELEMENT PlainPastaandNoodles (cod)>  
<!ELEMENT InfantFoodsandBeverages (cod)>  
<!ELEMENT InstantMixesandSupplies (cod)>  
<!ELEMENT CoffeeandTea (cod)>  
<!ELEMENT AlcoholicBeverages (cod)>  
<!ELEMENT NonAlcoholicBeverages (cod)>  
<!ELEMENT Pulses (cod)>  
<!ELEMENT Cereals (cod)>



```
<!ELEMENT ProcessedCereals (cod)>
<!ELEMENT cod (#PCDATA)>
```

## XML: UNSPSC- “FOOD”

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE FoodBeverageandTobaccoProducts SYSTEM "ufood.dtd">
<FoodBeverageandTobaccoProducts>
  <FruitsandVegetablesandNutsandSeeds>
    <Vegetables>
      <cod/>
    </Vegetables>
  </FruitsandVegetablesandNutsandSeeds>
</FoodBeverageandTobaccoProducts>
```

Si nota immediatamente che la DTD definisce la struttura per un set molto più ampio di elementi di quanti vengano messi in evidenza dal file XML. La struttura di tale file è dovuta all’uso, nella DTD, del “|” che viene letto come un OR, mentre la virgola verrebbe letta come un AND. Per questo motivo il file XML descrive solo un ramo dell’albero descritto dal DTD.



# Glossario

**Agente** : strumento che realizza un servizio, sia per il suo proprietario, sia per un cliente del suo proprietario.

**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.

**API: Application Program Interface** è un metodo specifico prescritto dal sistema operativo o da un programma applicativo, con il quale un programmatore mentre scrive 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 back-end e front-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.

**Architettura**: insieme di componenti.

**ASP: Application Service Provider**. 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.

**Client/ Sever**: 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

**CORBA: Common Object Request Broker Architecture** 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.

**EDI:** Electronic Data Interchange è un formato standard per lo scambio di dati nelle attività di business.

**eProcurement:** Programmi di gestione dell'approvvigionamento realizzati per funzionare tramite Web.

**Eterogeneità:** incompatibilità trovate tra risorse e servizi sviluppati autonomamente, che vanno dalla piattaforma utilizzata, sistema operativo, modello dei dati, alla semantica, ontologia.

**FTP:** è il metodo più comune per effettuare il download e l'upload di files attraverso una connessione Internet. Il File Transfer Protocol è un metodo standardizzato per connettere computer in modo che i files possano essere condivisi.

**Gerarchia:** struttura di un modello che assegna ogni oggetto ad un livello, e definisce per ogni oggetto l'oggetto da cui deriva.

**HTML:** Hypertext Markup Language 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.

**Interoperare :** combinare sorgenti e domini multipli.

**GUI (Graphical User Interface):** 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 files e il trasferimento di informazioni.

**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 una interrogazione.

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

**OLE: Object Linking and Embedding** è un'interfaccia API per accedere a diverse sorgenti di dati.

**ODBC: Open Database Connectivity** è un'interfaccia per la programmazione di applicazioni (Application Program Interface - API) dalle quali accedere ai database. ODBC è uno standard di tipo open source.

**RDF: Resource Description Framework** 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.

**SQL (Structured Query Language)** è un linguaggio interattivo per ottenere informazioni da un database e aggiornarlo. SQL è uno standard ANSI e ISO.

**SSH: Secure Shell** 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 format gerarchico, i differenti componenti del documento.

**TCP/IP:** Transmission Control Protocol/Internet Protocol è il protocollo standard di comunicazione richiesto da i 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

- [Acom, 2002]“B-to-B eMarketplaces: If you build them will they come?” Acom Solutions, [www.acom.com](http://www.acom.com).
- [Beneventano, 1997a] 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, 1997b] D. Beneventano, Sonia Bergamaschi, Claudio Sartori, and Maurizio Vincini. “ODB-QOPTIMIZER: A tool for semantic query optimisation in oodb”. In Int. Conference on Data Engineering - ICDE97, 1997
- [Beneventano, 2000] 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
- [Bergamaschi, 1998] S. Bergamaschi, D.Beneventano, S. Castano and M. Vincini. “Integrazione di informazione: il linguaggio ODLI3 e la logica descrittiva OLCD”, Technical report T3-R03, 16 Luglio 1998.
- [Bergamaschi, 1999] S. Bergamaschi, C. Sartori, M. Vincini and D. Beneventano. “Semantic Integration of Heterogeneous Information Sources”, in Journal of Data and Knowledge Engineering, 1999.
- [Bergamaschi, 2002] 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, 1999] T.Berners-Lee, “Weaving the Web” Orion Business Books, London, 1999.
- [Bernstein, 01] P.A.Bernstein, E.Rahm, “A survey of approaches to automatic schema matching”, Springer-Verlang 2001.

- [Bianco, 2000] D.Bianco “SIWEB:un interfaccia Internet per il sistema MOMIS” Tesi di Laurea, Università di Modena, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 2000.
- [Castano, 2001] S.Castano, V. De Antonellis, S.De Capitani Di Vimercati “Global Viewing of Heterogeneous Data Sources”, IEEE Transaction on Knowledge and Data Engeneering, 2001.
- [Catarci,1993] 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.
- [Catarri, 1993] T.Catarri and M. Lenirei. “Representing and using interschema knowledge in cooperative information systems”. *Journal of Intelligent and Cooperative Information Systems*, volume 2/4, pages 375-398, 1993.
- [Cattell, 1994] R.G.G.Cattell.“ The Object Database Standard: ODMG93”. Morgan Kaufmann Publishers, San Mateo, CA, 1994.
- [Cattell, 1997] R.G.G. Cattell and others. “The Object Data Standard: ODMG 2.0”, Morgan Kaufmann Publishers, San Francisco, CA, 1997.
- [Clark, 1999] J. Clark: “XSL Transformations (XSLT)” Version 1.0, W3C Working Draft, August 1999. See <http://www.w3.org/TR/WD-xslt>
- [Clark & DeRose, 1999] J.Clarck, S.DeRose, “XML Path Language(Xpath)” Version1.0, W3C Recommendation, November 1999.
- [Connolly, 1997] D. Connolly: “XML: Principles, Tools, and Techniques”, O’Reilly, 1997.
- [Ding, 2002] Y. Ding, D. Fensel, M. Klein, B. Omelayenko, “The Semantic Web: Yet Another Hip”, in Data and Knowledge Engineering, 2002.
- [Dun&Bradstreet] Dun & Bradstreet “UN/SPSC: the first open global commodity classification system”.
- [Duschka, 1996] Oliver M.Duschka and Micheal R.Genesereth. Infomaster – “An Information Integration Toolkit”. Technical report, Department of Computer Science.Stanford University, 1996.
- [Eclass, 2000] Eclass, “White paper”, Versione 0.6, Luglio 2000.
- [Etzioni, 1997] O. Etzioni: “Moving Up the Information Food Chain”, *AI Magazine*,18(2), 1997.
- [Fensel] Dieter Fensel “Ontologies and Electronic Commerce”.
- [Fensel, 2001a] D. Fensel, “Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce”, Springer-Verlang, Berlin 2001.



- [Fensel, 2001b] Dieter Fensel, Ying Ding, Boris Omelayenko, Ellen Schulten, Guy Botquin, Mike Brown, Alan Flett, “Product Data Integration in B2B E-Commerce”, *IEEE Intelligent Systems*, vol. 16, no. 4, 2001, pp. 54-59.
- [Fensel, 2001c] Dieter Fensel, Borys Omelayenko “Ontologies: “An Analysis of Integration Problems of XML-Based Catalogues for B2B” E-commerce”, In Proceedings of the 9<sup>th</sup> IFIP 2.6 Working Conference on Database, Semantic Issues in e-commerce Systems, Hong Kong, April 2001.
- [Fensel, 2001d] B.Omelayenko, D.Fensel, “Ontologies: A Two-Layered Integration Approach for Product Information in B2B E-commerce” Second International Conference on Electronic Commerce and Web Technologies, EC-Web 2001 Munich, Germany, September 4-6, 2001, Proceedings. Lecture Notes in Computer Science 2115 Springer 2001.
- [Ferrari, 2000] 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.
- [Flett] Alan Flett, Mike Brown “ Enterprise-standard Ontology environments for intelligent e-business”.
- [Franceschi, 2000] 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, 1997] N. Fridman-Noy and C.D. Hafner: “The State of the Art in Ontology Design”, *AI Magazine*, 18(3):53-74, 1997.
- [Granada Research, 2002] Granada Research “Using the UNSPSC United Nations Standard Products and Service Code: Why Coding and Classifying Product is Critical to Success in Electronic Commerce”, 2002.
- [Greenwald & Kephart, 1999] A. R. Greenwald and J. O. Kephart: “Shopbots and Pricebots”. In *Proceedings of the 16th International Joint Conference on AI (IJCAI-99)*, Stockholm, Sweden, July 31 - August 6, 1999.
- [Gruber, 1993] T. R. Gruber: “A Translation Approach to Portable Ontology Specifications”, *Knowledge Acquisition*, 5:199—220, 1993.
- [Guerra, 2000] 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, 2000] 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.

- [Hogan, 2001] Mike Hogan “ The supplier’s perspective in B2B E-Commerce”, [www.ecomworld.com](http://www.ecomworld.com), 2001.
- [Hull, 1995] R. Hull and R. King et al. Arpa I<sup>3a</sup> r eference architecture, 1995. Available at <http://www.isse.gmu.edu/I3 Arch/index.html>.
- [Klein, 2001a] M. Klein: “Combining and relating ontologies: an analysis of problems and solutions”. In A. Gómez-Pérez, M. Gruninger, H. Stuckenschmidt, and M. Uschold, editors, *Workshop on Ontologies and Information Sharing, IJCAI’01*, Seattle, USA, 2001.
- [Klein, 2001b] M. Klein and D. Fensel: “Ontology versioning on the Semantic Web”, In *Proceedings of the the 1st Semantic Web Working Symposium*, Stanford, CA, USA, July 30th-August 1st, 2001.
- [Lassila & Swick, 1999] O. Lassila and R. Swick. “Resource Description Framework (RDF)”. W3C proposed Recommendation, January 1999. <http://www.w3c.org/TR/WD-rdf-syntax>.
- [Malvezzi, 2000] G. Malvezzi. “Estrazione di relazioni lessicali con WordNet nel sistema MOMIS” Tesi di Laurea, Università di Modena, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 2000.
- [McGuinness, 1999] D. L. McGuinness: “Ontologies for Electronic Commerce”. In *Proceedings of the AAAI ’99 Artificial Intelligence for Electronic Commerce Workshop*, Orlando, Florida, July, 1999.
- [Melnik&Decker, 2000] S. Melnik, S. Decker, “A layered approach to information modeling and interoperability on the Web”, in ECDL 2000 Workshop on the Semantic Web.
- [Miller, 1993] A.G. Miller, A. Beckwith, C. Fellbaum, D. Gross, K. Miller “Introduction to WordNet: An online lexical database”, August 1993.
- [Miller, 1995] A.G. Miller. “Wordnet: A lexical database for english”. *Communications of the ACM*, 38 (11):39–41, 1995.
- [Miller, 1998] E. Miller: “An Introduction to the Resource Description Framework”, *Dlib Magazine*, May 1998.
- [Montanari 1998] 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, 1999] 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
- [Omelayenko] Borys Omelayenko, Dieter Fensel, “A Layered Integration Approach for Product Descriptions in B2B e-commerce”.

- [Omelayenko, 2001a] 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, 2001b] Borys Omelayenko “Preliminary Ontology Modeling for B2B Content Integration”, In *Proceedings of the First International Workshop on Electronic Business Hubs at the Twelfth International Conference on Database and Expert Systems Applications (DEXA-2001)*, IEEE CS Press, Munich, Germany, September 2001.
- [Omelayenko, 2001c] Borys Omelayenko “Ontology Integration Tasks in Business-to-Business E-commerce” *Engineering of Intelligent Systems, Proceedings of the Fourteenth International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (IEA/AIE-2001)*, LNAI 2070, Budapest, Hungary, June 4-7,2001, pp. 119-124. (short paper) © Springer-Verlag.
- [Omelayenko, 2001d] Borys Omelayenko “ Syntactic-Level Ontology Integration Rules for E-commerce” *Proceedings of The 14th International FLAIRS Conference (FLAIRS-2001)*, Key West, FL, May 21-23, 2001, AAAI Press.
- [Patel-Schneider, 2002] P.F. Patel-Schneider, Dieter Fensel, “Layering the Semantic Web: Problems and Directions”, Horrocks and J.Hendler (Eds):*ISWC 2002*, LNCS 2342, pp.16-29, 2002, Springer-Verlag Berlin Heidelberg 2002.
- [QL, 1998] *Proceedings of W3C Query Language Workshop (QL'98)* - The Query Languages Workshop, Boston, Massachussets, December 3-4, 1998. <http://www.w3.org/TandS/QL/QL98/>.
- [Ricciardi, 1999] 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.
- [Saltor, 1997] F.Saltor and E. Rodriguez. “On intelligent access to heterogeneous information” In *Proceedings of the 4th KRDB Workshop*, Athens, Greece, August 1997.
- [Schneider] Peter F. Patel-Schneider, Dieter Fensel “Layering the Semantic Web:Problems and Directions”.
- [Schulten, 2001] 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.
- [UNCTAD, 2001] “E-Commerce and development report 2001”, United Nation Conference on Trade and Development, 2001.

- [Venuta, 2000] F. Venuta. "Trattamento della Conoscenza Estensionale nel sistema MOMIS" Tesi di Laurea, Università di Modena, Facoltà di Ingegneria, corso di laurea in Ingegneria Informatica, 2000.
- [Weibel et al., 1995] S. Weibel, J. Gridby, and E. Miller: "OCLC/NCSA" Metadata Workshop Report, Dublin, EUA, 1995. [http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin\\_core\\_report.html](http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html).
- [Wiederhold, 1992] G. Wiederhold: "Mediators in the Architecture of Future Information Systems", *IEEE Computer*, 25(3):38-49, 1992.
- [Wiederhold, 1996] G. Wiederhold et al. "Integrating Artificial Intelligence and Database Technology", volume 2/3. *Journal of Intelligent Information Systems*, Giugno 1996.
- [Wiederhold & Genesereth et al., 1997] G. Wiederhold and M. Genesereth: "The Conceptual Basis for Mediation Services". *IEEE Expert*, September/October, pp. 38-47, 1997.
- [Wiederhold et al., 1997] G. Wiederhold: "Value-added Mediation in Large-Scale Information Systems". In R. Meersman and Leo Mark (eds.): *Database Application Semantics*, Chapman and Hall, 1997, pages 34-56.
- [Zaccaria, 1998] 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 di riferimento:***

[ecommerce.about.com/mbody.htm](http://ecommerce.about.com/mbody.htm): sito all'interno del quale è possibile trovare informazioni di carattere generico sull'e-commerce.

[www.ecomworld.com](http://www.ecomworld.com): portale verticale che approfondisce le tematiche del B2B e-commerce.

[www.commerceone.com](http://www.commerceone.com): sito ufficiale di Commerce One, leader mondiale nella fornitura di soluzioni che connettono e ottimizzano le interazioni tra compratori e venditori

[www.esteel.com](http://www.esteel.com): sito in realzione con NewView Technologies Inc, fornitore di Business Interactions Management Solutions

[www.newview.com](http://www.newview.com): NewView si occupa del mondo delle interazioni fornendo soluzioni che agevolano le operazioni di complicate reti di commercio.

[www.eclass.de](http://www.eclass.de): sito ufficiale dell'omonima classificazione.

[www.unspsc.com](http://www.unspsc.com): sito gestito da UNDP che attualmente gestisce e distribuisce la codifica UNSPSC.

[www.eccma.org](http://www.eccma.org) : sito ufficiale di Electronic Commerce Code Management Association (ECCMA), associazione che fino all'ottobre 2002 ha gestito distribuito la codifica UNSPSC.

[www.ntis.gov/product/naics.htm](http://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](http://www.census.gov/epcd/www/naics.html): sito gestito dal Census Bureau che fornisce informazioni relative alla codifica NAICS.

[www.naics.com/search.htm](http://www.naics.com/search.htm): Sito 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](http://www.statcan.ca/english/Subjects/Standard): sito ufficiale di NAICS Canada.

[www.siccode.com](http://www.siccode.com): sito che fornisce informazioni sulle codifiche SIC e NAICS e permette di cercare codici al loro interno.

[www.bls.gov/bls/naics.htm](http://www.bls.gov/bls/naics.htm): sito che fornisce informazioni relative alle codifiche NAICS e SIC, gestito da U.S. Department of Labor.

[www.cogsci.princeton.edu/~wn/](http://www.cogsci.princeton.edu/~wn/): sito gestito da 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.

[www.dbgroup.unimo.it](http://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.acom.com](http://www.acom.com): sito ufficiale di ACOM sviluppa componenti software che estendono le potenzialità dei sistemi di B2B e-commerce con EDI/XML, e-payments ed e-forms solutions.

[www.ontoweb.org](http://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.rosettanet.org](http://www.rosettanet.org): sito ufficiale di Rosettanet, consorzio delle maggiori compagnie di Information Technology, Electronic Components e Semiconductor Manufacturing che si occupa della creazione e dell'implementazione standard per i processi di business.

[www.granadaresearch.com](http://www.granadaresearch.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](http://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](http://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](http://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.

[www.w3c.org](http://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.buyusa.com](http://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](http://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](http://www.italbiz.com) : Italbiz, aggrega la più grande comunità di aziende straniere e italiane che operano nel mercato internazionale, e ti aiuta ad aumentare le esportazioni, a trovare nuovi fornitori e ad aprirti nuovi mercati esteri.

[www.ceramicandmore.com](http://www.ceramicandmore.com): marketplace verticale realizzato in italiano e inglese, contiene un data base 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 tutti coloro che hanno contribuito alla realizzazione della presente tesi, in particolare il Professor Domenico Beneventano, l'Ing. Maurizio Vincini e l'Ing. Francesco Guerra.*

*Un ringraziamento speciale ai miei genitori che mi sono sempre stati vicini e con la loro stima mi hanno dato il più grande incoraggiamento.*

*Ringrazio in modo particolare Viviana, che ha lavorato con me alla realizzazione di questo progetto, per aver reso questi mesi assolutamente piacevoli ed indimenticabili.*

*Ringrazio i miei compagni di corso che in questi anni sono stati un supporto prezioso di cui non avrei saputo fare a meno.*

*In modo particolare voglio ringraziare Prisca, un'amica speciale con la quale in questi anni ho condiviso tutto e a cui devo molto.*

*Infine, un ringraziamento dal profondo del mio cuore a tutti i miei amici, che nei modi più strani e impensabili mi sono stati vicini.*

---