



**UNIVERSITÀ
DI TORINO**

di.unito.it

**DIPARTIMENTO
DI INFORMATICA**

UNIVERSITÀ DI TORINO DIPARTIMENTO DI INFORMATICA
CORSO DI LAUREA TRIENNALE IN INFORMATICA

Editing di ontologie tramite il linguaggio di programmazione funzionale CDuce

Relatore
Professoressa Viviana Bono

Laureando
Davide Camino
Matricola: 897753

ANNO ACCADEMICO 2022-2023

Data di laurea Novembre 21, 2023

A tutti quelli che mi vogliono bene.

*L'informatica riguarda i computer
non più di quanto l'astronomia riguardi i telescopi.
E. W. Dijkstra*

Dichiaro di essere responsabile del contenuto dell'elaborato che presento al fine del conseguimento del titolo, di non avere plagiato in tutto o in parte il lavoro prodotto da altri e di aver citato le fonti originali in modo congruente alle normative vigenti in materia di plagio e di diritto d'autore. Sono inoltre consapevole che nel caso la mia dichiarazione risultasse mendace, potrei incorrere nelle sanzioni previste dalla legge e la mia ammissione alla prova finale potrebbe essere negata.

ABSTRACT

Questo lavoro illustra lo sviluppo di strumenti per l'editing di ontologie tramite il linguaggio di programmazione funzionale CDuce, in particolare la realizzazione di programmi per il refactoring, il merge e la traduzione da tesauro a ontologie. Lo scopo dello studio è quello di valutare se, e in che condizioni, lo sviluppo di strumenti più o meno ad hoc per l'editing di ontologie è vantaggioso rispetto all'uso di strumenti grafici tradizionali come Protégé. Nella valutazione si tengono in considerazione principalmente la difficoltà tecnica e il tempo di sviluppo degli strumenti creati con CDuce e il tempo necessario e la ripetitività di fare le stesse modifiche con strumenti grafici.

Indice

ABSTRACT	iii
1 Concetti di base	1
1.1 Ontologie e tesauroi	1
1.1.1 ontologie	1
1.1.2 Tesauroi	2
1.2 Strumenti di editing	2
1.2.1 Protégé	2
1.2.2 CDuce	2
1.3 Metalinguaggi	3
1.3.1 OWL	3
1.3.2 SKOS	3
2 Strumenti offerti da CDuce	4
2.1 Parsing documenti XML	4
2.1.1 Esempio	4
2.2 Manipolare documenti XML	5
2.2.1 Esempio	6
2.3 Query	6
2.3.1 Esempio	7
3 Trasformare un tesauro in un'ontologia	9
3.1 Vantaggi del passaggio da tesauro ad ontologia	9
3.1.1 Europeana Fashion Thesaurus: capturing imagination	9
3.1.2 Svantaggi del tesauro	10
3.2 Struttura del tesauro	10
3.3 Struttura dell'ontologia	11
3.4 Da concetto SKOS a classe OWL	11
3.4.1 Trasformare gli attributi	12
3.4.2 Trasformare una singola classe	12
3.5 Costruire la nuova ontologia	13
3.6 Versione compatta	14
3.7 Aumentare l'espressività	14
3.8 In Protégé	15
3.9 Conclusioni	15
4 Merge di ontologie	16
4.1 Obiettivo del merge	16
4.1.1 Ontologie di partenza	16
4.1.2 Ontologia di arrivo	17

4.2	Struttura generica di un'ontologia	17
4.3	Merge	18
4.3.1	Funzioni utili	18

Capitolo 1

Concetti di base

Introduzione

Qui illustriamo alcuni dei concetti di base che serviranno per comprendere il resto della discussione, daremo una definizione di ontologia e tesauro, descriveremo brevemente gli strumenti utilizzati e i linguaggi con cui si descrivono le basi di conoscenza che tratteremo.

1.1 Ontologie e tesauri

1.1.1 ontologie

Esempio

Consideriamo una semplice ontologia che rappresenta persone con legami di parentela genitore-figlio; le persone hanno uno o più nomi salvate nel tag `comment`. Modelliamo questa ontologia con una classe `Persone` e una sottoclasse `Genitori` (i cui individui sono `Persone` che realizzano la relazione `genitoreDi`). Creiamo la relazione `genitoreDi`. Infine popoliamo l'ontologia con alcuni individui. Il risultato ottenuto con Protégé è un documento XML di questo tipo:

Listato 1.1: persone.rdf

```
1 <?xml version="1.0"?>
2 <rdf:RDF xmlns="http://www.persone/"
3     xml:base="http://www.persone/"
4     xmlns:owl="http://www.w3.org/2002/07/owl#"
5     xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
6     xmlns:www="http://www.persone#"
7     xmlns:xml="http://www.w3.org/XML/1998/namespace"
8     xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
9     xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
10 <owl:Ontology rdf:about="http://www.persone"/>
11 <!-- http://www.persone#genitoreDi -->
12 <owl:ObjectProperty rdf:about="http://www.persone#genitoreDi"/>
13 <!-- http://www.persone#Genitori -->
14 <owl:Class rdf:about="http://www.persone#Genitori">
```

```

15     <owl:equivalentClass>
16         <owl:Restriction>
17             <owl:onProperty>
18                 ↪ rdf:resource="http://www.persone#genitoreDi"/>
19             <owl:someValuesFrom>
20                 ↪ rdf:resource="http://www.persone#Persone"/>
21             </owl:Restriction>
22         </owl:equivalentClass>
23     <rdfs:subClassOf rdf:resource="http://www.persone#Persone"/>
24 </owl:Class>
25 <!-- http://www.persone#Persone -->
26 <owl:Class rdf:about="http://www.persone#Persone"/>
27 <!-- http://www.persone#Bruto -->
28 <owl:NamedIndividual rdf:about="http://www.persone#Bruto">
29     <rdf:type rdf:resource="http://www.persone#Persone"/>
30     <rdfs:comment>Bruto</rdfs:comment>
31 </owl:NamedIndividual>
32 <!-- http://www.persone#Cesare -->
33 <owl:NamedIndividual rdf:about="http://www.persone#Cesare">
34     <rdf:type rdf:resource="http://www.persone#Genitori"/>
35     <www:genitoreDi rdf:resource="http://www.persone#Bruto"/>
36     <rdfs:comment>Caio</rdfs:comment>
37     <rdfs:comment>Augusto</rdfs:comment>
38     <rdfs:comment>Giulio</rdfs:comment>
39     <rdfs:comment>Cesare</rdfs:comment>
40 </owl:NamedIndividual>
</rdf:RDF>

```

1.1.2 Tesauri

1.2 Strumenti di editing

1.2.1 Protégé

1.2.2 CDuce

Pattern matching

È un'operazione fondamentale in Cduce ed ha la forma:

```

match e with
| p1 -> e1
...
| pn -> en

```

Si cerca di fare il match tra la valutazione di un'espressione *e* e vari pattern *pi*. Il primo pattern che fa il match con *e* attiva la corrispondente espressione sulla destra che può usare le variabili legate dal pattern.

1.3 Metalinguaggi

1.3.1 OWL

1.3.2 SKOS

Capitolo 2

Strumenti offerti da CDuce

Introduzione

In questo capitolo analizziamo brevemente gli strumenti principali offerti da CDuce per l'editing di documenti XML. La trattazione vale per un qualsiasi documento XML, ma negli esempi ci concentreremo sulle ontologie vedendo dei primi comandi per trattare ontologie dalla struttura semplice.

2.1 Parsing documenti XML

Un documento XML non è altro che una struttura ad albero: ogni nodo rappresenta un concetto che può essere meglio definito nei figli del nodo stesso.

In CDuce possiamo ricostruire tale struttura definendo dei tipi. La forma generale di un elemento XML è `<(tag) (attr)> content` dove `tag`, `attr` e `content` sono espressioni su cui è possibile fare pattern matching. Si può quindi creare un tipo generale che faccia match con il tag root del documento XML e che come `content` abbia un array eterogeneo che conterrà tutti i figli del tag root. Ad ogni elemento del vettore sarà associato un tipo che avrà la stessa struttura del tipo generale e permetterà di descrivere la struttura del documento XML discendendo fino alle foglie.

2.1.1 Esempio

Torniamo all'esempio dell'ontologia di persone definita nel listato 1.1 e descriviamo in CDuce la sua struttura.

Listato 2.1: persone.cd

```
1 namespace owl="http://www.w3.org/2002/07/owl#"
2 namespace rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3 namespace www="http://www.persone#"
4 namespace xml="http://www.w3.org/XML/1998/namespace"
5 namespace xsd="http://www.w3.org/2001/XMLSchema#"
6 namespace rdfs="http://www.w3.org/2000/01/rdf-schema#";
7
8 type Ontology = <rdf:RDF xml:base=String> [ Thing * ]
9 type Thing = Ont | Property | Class | Individual
```

```

10
11 type Ont = <owl:Ontology rdf:about=String> []
12
13 type Property = <owl:ObjectProperty rdf:about=String> []
14
15 type Class = <owl:Class rdf:about=String> [ ClassAttr * ]
16 type ClassAttr = EqClass | SubClass
17 type EqClass = <owl:equivalentClass> [ <owl:Restriction> [ AnyXml * ] ]
18 type SubClass = <rdfs:subClassOf rdf:resource=String> []
19
20 type Individual = <owl:NamedIndividual rdf:about=String> [ IndAttr * ]
21 type IndAttr = TypeInd | PropInd | Name
22 type Name = <rdfs:comment> String
23 type TypeInd = <rdf:type rdf:resource=String> []
24 type PropInd = <_ rdf:resource=String> [];;
25
26 let ontology :? Ontology = load_xml "persone.rdf";;

```

Si definiscono (linea 1 a 6) i Namespace in modo che CDuce li possa correttamente interpretare nel resto del documento. Si passa poi alla struttura vera e propria del file: tutte le informazioni sono contenute nell'elemento di tipo `Ontology` che è costituito da un tag e da una lista di elementi di tipo `Thing` che a sua volta può rappresentare un elemento di tipo `Ont`, `Property`, `Class` o `Individual`. Ognuno di questi elementi fa il match con un nodo figlio del nodo root del documento XML; Al loro volta questi elementi hanno una struttura interna che può essere più o meno dettagliata a seconda di come ci interessa operare sul documento. È interessante notare che, dato che in questo caso le specifiche proprietà della restrizione di un elemento `EqClass` non ci interessano, abbiamo potuto fare il match di queste con un vettore di lunghezza arbitraria (anche nulla) di generici elementi XML. Se avessimo voluto specificare che il tag restrizione contiene esattamente 2 elementi senza specificare quali avremmo potuto scrivere `<owl>Restriction> [AnyXml AnyXml]`.

Quando andremo a caricare il documento con il comando alla riga 26 CDuce carica il documento ed esegue il controllo di tipo verificando che la struttura del file XML sia effettivamente quella descritta dall'elemento `Ontology`; se il controllo va a buon fine otterremo l'elemento chiamato `ontology` di tipo `Ontology` contenente tutto il file XML.

2.2 Manipolare documenti XML

Una volta definita la struttura del documento possiamo definire delle funzioni che mappano un elemento in un altro elemento. Gli strumenti fondamentali sono:

- Pattern matching come descritto in [1.2.2](#)
- `map`: permette di applicare una funzione a tutti gli elementi di una lista e restituisce una nuova lista di elementi trasformati della stessa lunghezza della lista iniziale;
- `transform`: permette di applicare una funzione a ogni elemento di una lista, restituendo per ogni elemento una lista di lunghezza arbitraria (anche nulla) e concatenando infine il risultato. Grazie alla funzione `transform` si può ottenere una lista di lunghezza diversa rispetto alla lista di partenza.

2.2.1 Esempio

Consideriamo nuovamente l'ontologia padri-figli con struttura definita in 2.1 e costruiamo 2 funzioni, la prima (chiamata **name**) che permetta di estrarre da un individuo tutti i nomi, la seconda (chiamata **names**) che usi la prima per costruire una lista con tutti i nomi contenuti nell'ontologia.

Listato 2.2: basic functions

```
1  (*first function*)
2  let fun name (Individual -> [String *])
3    <owl:NamedIndividual ..> [ (n::Name | _) * ] -> map n with
4    <rdfs:comment> s -> s;;
5  (*second function*)
6  let fun names (Ontology -> [String *])
7    <rdf:RDF ..> x -> transform x with
8    | (x & Individual) -> name x
9    | _ -> [];
```

Nella prima funzione, usando il pattern matching, lego la variabile **n** alla sequenza di tutti gli elementi di tipo **Name** associati a quell'individuo, poi uso **map** per estrarre da ogni elemento di questa lista solo la stringa col nome.

Nella seconda funzione uso **transform** per selezionare dalla lista **x** (lista di elementi tipo **Thing**) solo gli elementi **Individual**, a questi applico la funzione **name** sopra definita. La seconda clausola di **transform**, che serve a buttare via tutti gli elementi di cui non si è ancora fatto il match è implicita e può essere quindi omessa.

La prima funzione ci ha permesso di passare da una lista **n** di elementi **Name** ad una lista della stessa lunghezza di stringhe. La seconda ci ha permesso di passare da una lista di elementi **Thing** a una lista di elementi **Individual** di differente lunghezza.

2.3 Query

Un punto di forza di CDuce sono le Query. Le useremo profusamente nei capitoli 3 e 4 per creare liste e selezionare elementi in modo rapido e leggibile al posto della funzione **transform**. La forma generale di una Query è molto simile alla stessa espressa in SQL

```
select e
from p1 in e1,
     p2 in e2,
     :
     pn in en
where c
```

Dove **e** ed **ei** sono espressioni, **pi** sono pattern e **c** è un'espressione booleana. Il risultato finale è la lista di tutti i valori ottenuti valutando **e** nella sequenza di possibili combinazioni in cui le variabili libere di **e** sono legate facendo il match dei pattern **pi** con le espressioni **ei**, a condizione che **c** sia rispettata.

Le Query possono essere simulate con **transform** che permette di creare, grazie a pattern matching successivi l'espressione **e** e di selezionarla solo nel caso in cui l'espressione booleana **b** sia rispettata. Si può riscrivere la generica Query usando la **transform** come segue:

```

transform e1 with p1 ->
  transform e2 with p2 ->
    ...
    transform en with pn ->
      if b then [e] else []

```

La valutazione dei due comandi produce il medesimo risultato.

Anche se **transform** può sembrare più versatile o espressiva delle Query, quando possibile, è sempre vantaggioso usare queste ultime. I vantaggi sono molteplici:

- la struttura delle Query è più elegante e leggibile, permette di capire con più facilità cosa si sta cercando;
- le Query sono automaticamente ottimizzate con le stesse tecniche di ottimizzazione di SQL, questo è molto vantaggioso soprattutto in ontologie o tesauroi particolarmente popolati;
- CDuce fornisce dei controlli a priori sul risultato della Query permettendo di evidenziare errori che porterebbero ad una Query che viene interpretata correttamente ma produce sempre risultati vuoti.

Oltre che con la struttura del **select from** le Query possono essere espresse anche come proiezioni, come mostrato nel listato 2.3

2.3.1 Esempio

Facendo sempre riferimento all'ontologia con struttura definita nel listato 2.1 scriviamo delle Query per ottenere lo stesso risultato della seconda funzione definita nel listato 2.2

Listato 2.3: basic Query

```

1  (*Query*)
2  let sel = select y
3      from x in [ontology]/Individual,
4           y in [x]/Name
5  in
6  map sel with <rdfs:comment> n -> n;;
7
8  (*projection 1*)
9  map [ontology]/Individual/Name with <rdfs:comment> n -> n;;
10
11 (*projection 2*)
12 map [ontology]/Individual/<rdfs:comment> _ with <rdfs:comment> n -> n;;
13
14 (*error 1*)
15 [ontology]/Name;;
16
17 (*error 2*)
18 let sel = select x
19     from x in [ontology]/Individual
20 in [sel]/Name
21

```

```

22  (*correction error 2*)
23  let sel = select x
24      from x in [ontology]/Individual
25  in sel/Name

```

La query crea una lista di elementi di tipo `Name` a cui si assegna il nome `sel`, questa lista viene poi trasformata in una lista di stringhe grazie alla `map`. La proiezione fa esattamente la stessa cosa in modo ancora più conciso. Sia nella forma del `select from` sia nella forma delle proiezioni si possono usare anche espressioni per fare il match oltre che tipi, come si vede nella `projection 2`.

Gli ultimi due esempi mostrano come CDuce possa aiutare nella rilevazione degli errori: se eseguiti restituiscono rispettivamente gli avvertimenti: `Warning: This projection always returns the empty sequence` e `Warning: This branch is not used`, informandoci che il risultato sarà sempre vuoto.

Analizziamo il primo errore. CDuce sa che, affinché il risultato possa contenere dei valori, gli elementi della lista chiamata `ontology` (che sono di tipo `Thing`) devono essere sottotipi di `[<_ ..>[Any* Name Any*] *]`; dato che non è così il risultato sarà sempre vuoto.

Il secondo errore è dovuto al fatto che mettendo le parentesi quadre intorno a `sel` questo viene interpretato come una sequenza (anche se lo è già), ne risulta una sequenza il cui unico elemento è una sequenza (in particolare il tipo attribuito da CDuce è `[[Individual*]]`) anche in questo caso controllando il tipo CDuce è in grado di capire che il risultato sarà sempre vuoto. Perché la Query restituisca dei valori il tipo deve essere `[Individual*]`, cosa che avviene nella Query successiva che in effetti restituisce il risultato atteso.

Conclusioni

Abbiamo visto come si può descrivere la struttura di un documento XML in CDuce, come si può trasformare un elemento di un tipo in elemento di un altro tipo e come creare liste selezionando quali elementi aggiungere. Nei capitoli 3 e 4 applicheremo gli strumenti analizzati ad esempi notevoli che mostrino in che modo e in che contesti CDuce possa essere usato per operare, modificandoli, tesauri e ontologie

Capitolo 3

Trasformare un thesauro in un'ontologia

Introduzione

Qui vediamo come sia possibile usare CDuce per creare ontologia a partire da una rappresentazione della conoscenza già formalizzata in altri modi. Per illustrare il processo di creazione faremo riferimento ad un esempio facilmente replicabile: la trasformazione di un thesauro in un'ontologia (per una trattazione più astratta e formale sulla re-ingegnerizzazione di un thesauro si veda [2]). Commenteremo le scelte effettuate e le strategie adottate. L'esempio permette una discussione lineare e senza eccessivi giri di parole mantenendo comunque una buona generalità qualora si applicassero le stesse scelte e strategie a diversi contesti.

Dopo aver trasformato il thesauro in un'ontologia con CDuce vediamo brevemente quali altri strumenti avremmo potuto utilizzare, in particolare cercheremo di ottenere risultati analoghi con Protégé discutendo vantaggi e svantaggi di ciascun approccio.

3.1 Vantaggi del passaggio da thesauro ad ontologia

Potremmo essere interessati a trasformare un thesauro espresso in SKOS in un'ontologia espressa in OWL per varie ragioni: prima fra tutte un'ontologia è una rappresentazione più formale della conoscenza e OWL ha un potere espressivo maggiore di SKOS; potremo quindi effettuare Query più avanzate e utilizzare strumenti di inferenza più potenti.

Per presentare questo argomento prendiamo in considerazione il thesauro: "Europeana Fashion Thesaurus"¹ e tentiamo di trasformarlo in un ontologia. Prima di addentrarci nella parte tecnica forniamo una breve introduzione al thesauro che editeremo.

3.1.1 Europeana Fashion Thesaurus: capturing imagination

L' "Europeana Fashion project"² è un progetto che si è posto come obiettivo quello di organizzare sotto una struttura gerarchica tutto ciò che riguardasse la moda, comprendendo anche sinonimi e contrari. Il thesauro è pubblico può essere scaricato liberamente da chiunque sia interessato. Il thesauro permette di accedere in modo logico, organizzato e strutturato

¹<http://thesaurus.europeanafashion.eu/>

²<http://www.europeanafashion.eu/>

ad una vasta conoscenza. Data la struttura ad albero le relazioni tra oggetti sono chiare e l'aggiunta o la ricerca di informazioni può essere fatta in modo rapido ed efficiente.

3.1.2 Svantaggi del thesauro

Data la sua struttura ad albero il thesauro permette una rappresentazione della conoscenza puramente tassonomica, questo in certi contesti può essere una limitazione, ad esempio nonostante nel thesauro siano presenti sia capi d'abbigliamento che materiali non c'è nessuna possibilità di mettere in relazione i due concetti in modo formale, inoltre se avessimo degli strumenti per fare inferenza non saremmo in grado di dedurre che una bandana è un accessorio per il capo, perché quest'ultima non si trova nel ramo degli accessori della testa.

Il passaggio ad un'ontologia permetterebbe di costruire relazioni più complesse per descrivere in modo più profondo gli oggetti del dominio di interesse.

3.2 Struttura del thesauro

La struttura del thesauro che vogliamo re-ingegnerizzare è semplice, in particolare abbiamo quattro concetti fondamentali da cui partono tutti gli altri rami che rappresentano i concetti derivati. I concetti fondamentali sono:

- oggetti di moda;
- eventi di moda;
- tecniche;
- materiali.

Ogni concetto che sia uno di quelli fondamentali o meno è rappresentato con un tag **Description**, contenente, a sua volta, una lista di attributi tra cui le **label** che contengono il nome (nelle varie lingue) del concetto, il tag **broader** che nei concetti derivati permette di risalire al nodo padre, le **scopeNote** che contengono una descrizione del concetto, e il tag **exactMatch** che rimanda “The Getty Vocabularies”³.

Analizzata la struttura del thesauro possiamo descriverla formalmente con CDuce, in modo da fare il parsing del documento per poter definire funzioni che permettano di trasformare i concetti descritti dal tag **Description** in classi nel linguaggio OWL. Una possibile descrizione del thesauro in CDuce è la seguente:

Listato 3.1: thesaurus_europeana.cd

```
1 type Thesaurus = <rdf:RDF>[Desc *]
2 type Desc = <rdf:Description rdf:about=String>[DescAtt *]
3
4 type DescAtt = AltLabel | InScheme | PrefLabel | ScopeNote | ExactMatch |
   ↳ Broader | Type
5 type AltLabel = <skos:altLabel xml:lang=String> String
6 type InScheme = <skos:inScheme rdf:resource=String> []
7 type PrefLabel = <skos:prefLabel xml:lang=String> String
8 type ScopeNote = <skos:scopeNote xml:lang=String> String
9 type ExactMatch = <skos:exactMatch rdf:resource=String> []
```

³<http://vocab.getty.edu/>

```

10 type Broader = <skos:broader rdf:resource=String> []
11 type Type = <rdf:type rdf:resource=String> [];;

```

3.3 Struttura dell'ontologia

L'ontologia che vogliamo creare a partire da questo tesoro, almeno inizialmente, non potrà contenere più informazioni di quelle già presenti, ci limitiamo quindi a costruire una tassonomia, che potrà poi essere arricchita passando da un albero ad un grafo, con relazioni più ricche tra le classi.

Una volta creata l'ossatura dell'ontologia potremo andare a definire delle relazioni sugli individui ad esempio per specificare che un dato capo d'abbigliamento è prodotto con un determinato tessuto.

La struttura dell'ontologia, almeno per quello che ci serve per riportare tutte le informazioni contenute nel tesoro, può essere formalizzata in CDuce nel seguente modo:

Listato 3.2: ontology_europeana.cd

```

1 type Ontology = <rdf:RDF xml:base=String> [ Thing * ]
2 type Thing = Ont | Class
3
4 type Ont = <owl:Ontology rdf:about=String> []
5
6 type Class = <owl:Class rdf:about=String> [ ClassAtt * ]
7 type ClassAtt = SubClass | Label | Note | Dictionary
8 type SubClass = <rdfs:subClassOf rdf:resource=String> []
9 type Label = <rdfs:label xml:lang=String> String
10 type Note = <skos:scopeNote xml:lang=String> String
11 type Dictionary = <skos:exactMatch rdf:resource=String> []

```

Notiamo subito come la struttura dell'ontologia si sia molto semplificata rispetto alla struttura del listato 2.1, questo perché adesso ci interessa solo ricostruire con classi OWL i concetti SKOS.

È interessante come si possano integrare tag SKOS direttamente nel linguaggio OWL, in questo caso li usiamo per aggiungere informazioni umanamente leggibili (le note) e per mantenere il riferimento al dizionario (per una discussione dettagliata sull'interazione tra OWL e SKOS si veda [1]).

Si potrebbero anche eliminare completamente i tag SKOS mappandoli adeguatamente in altri tag (ad esempio trasformando `scopeNote` in `comment`), ma questo non porterebbe nessun vantaggio dal punto di vista della formalità dell'ontologia e renderebbe più complesse le funzioni di trasformazione.

3.4 Da concetto SKOS a classe OWL

Il nostro obiettivo è trasformare tutti i concetti del tesoro in classi di un'ontologia, mantenendo la gerarchia esprimendola come relazione di sottoclasse. Per raggiungere questo scopo definiamo una funzione per mappare gli elementi di tipo `DescAttr` in elementi di tipo `ClassAttr`. Poi usiamo queste funzioni per definirne una che ci permetta di passare da un

intero concetto del tesoro ad una classe dell'ontologia; infine, usando la funzione `map` (2.2), possiamo applicare questa funzione a tutti i concetti del tesoro per ottenere le classi che popoleranno l'ontologia.

3.4.1 Trasformare gli attributi

Iniziamo definendo le funzioni per mappare i tag del tesoro in tag dell'ontologia, questo è il primo esempio in cui si possono apprezzare i vantaggi dell'uso di un linguaggio funzionale. Esprimeremo le trasformazioni in modo semplice ed elegante senza perdere in leggibilità, inoltre avremo la garanzia che la trasformazione restituisca esattamente il tipo dichiarato nell'interfaccia della funzione e il controllo di tipo e l'inferenza del tipo di un'espressione ci aiuteranno a trovare e correggere eventuali errori. Definiamo una funzione per ciascun tag che desideriamo esportare, in particolare tralasciamo il tag `type` che nel tesoro assume solo 2 valori (`Concept` e `ConceptScheme`). Le funzioni possono essere scritte in C'Duce come segue:

Listato 3.3: SKOS_to_OWL.cd

```

1  let fun transformNote (ScopeNote -> Note)
2    x -> x
3
4  let fun transformDictionary (ExactMatch -> Dictionary)
5    x -> x
6
7  let fun transformLabel (PrefLabel -> Label ; AltLabel -> Label)
8    | <skos:prefLabel xml:lang=l> lab -> <rdfs:label xml:lang=l> lab
9    | <skos:altLabel xml:lang=l> lab -> <rdfs:label xml:lang=l> lab;;
10
11 let fun transformSubClass (Broader -> SubClass)
12   <skos:broader rdf:resource=res> [] -> <rdfs:subClassOf rdf:resource=res>
    ↳ [];;

```

Le prime due funzioni sono banali: confrontando i listati 3.1 e 3.2 notiamo che gli elementi di tipo `ScopeNote` e `Note` hanno la stessa struttura, così come gli elementi di tipo `ExactMatch` e `Dictionary`. Per questi tag è sufficiente la funzione identità.

Per quanto riguarda le `label` nel tesoro ci sono due tipi di elementi, nell'ontologia abbiamo deciso di usarne solo uno, vediamo quindi un esempio di overloading della funzione. Per quanto riguarda la trasformazione vera e propria usiamo il pattern matching per legare la variabile `l` e `lab` rispettivamente alla lingua e al testo della `label`; una volta legate queste variabili le usiamo per costruire il nuovo elemento di tipo `Label`.

La funzione più importante è certamente quella che lavora sull'elemento tipo `Broader`, questa si occupa infatti di trasformarlo in un elemento di tipo `SubClass` in modo da mantenere le relazioni gerarchiche del tesoro. Dal punto di vista della trasformazione però la funzione è molto semplice, leghiamo un'unica variabile `res` alla stringa che identifica il nodo padre e con questa costruiamo il tag `subClassOf`.

3.4.2 Trasformare una singola classe

Per trasformare un singolo concetto espresso in SKOS in una classe OWL definiamo due funzioni, la prima trasforma un attributo del concetto (`DescAttr`) in un vettore di attributi di una classe (`[ClassAttr *]`) per fare questo usiamo le funzioni definite prima (3.3). La

ragione per cui il risultato deve essere un vettore è che se l'attributo ci interessa restituiamo un vettore con un elemento, se l'attributo va scartato restituiamo un vettore vuoto. La seconda funzione costruisce l'involucro esterno della classe e usa la prima per trasformare tutti gli attributi del concetto in attributi della classe. L'implementazione di queste funzioni potrebbe essere la seguente:

Listato 3.4: concept_to_class.cd

```

1  let fun transformAtt (DescAtt -> [ClassAtt*])
2    | x & PrefLabel   -> [(transformLabel x)]
3    | x & AltLabel    -> [(transformLabel x)]
4    | x & ScopeNote   -> [(transformNote x)]
5    | x & ExactMatch  -> [(transformDictionary x)]
6    | x & Broader     -> [(transformSubClass x)]
7    | Any             -> [];
8
9  let fun transformClass (Desc -> Class)
10    <rdf:Description rdf:about=ab> [ (descAtt :: DescAtt)* ] ->
11      let classAtt = flatten ( map descAtt with x -> transformAtt x) in
12      <owl:Class rdf:about=ab> classAtt;;

```

Le funzioni sono abbastanza semplici, descriviamo brevemente la seconda: usiamo il pattern matching per legare la variabile `ab` alla stringa che identifica il concetto (useremo questo identificatore anche per la classe); leghiamo poi la variabile `descAtt` al vettore di attributi del concetto SKOS, usando poi la `map` per trasformare questo vettore in attributi di una classe OWL. Siccome la funzione `transformAtt` prende un elemento e restituisce un vettore, al termine della `map` avremo un vettore di vettori, per appiattirne la struttura usiamo la funzione già definita in CDuce `flatten`. A questo punto abbiamo tutti gli elementi per definire la nuova classe che costruiamo assemblando il tag `Class` con la stringa identificativa e il vettore di attributi.

3.5 Costruire la nuova ontologia

Ora che abbiamo una funzione per trasformare ogni concetto SKOS in una classe OWL possiamo applicarla a tutti i concetti del tesaurus per costruire un'ontologia. Possiamo completare la trasformazione in CDuce in questo modo:

Listato 3.5: thesaurus_to_ontology.cd

```

1  let fashion :? Thesaurus = load_xml "thesaurus.rdf";
2
3  let newClass = map [fashion]/Desc with x -> transformClass x
4  in
5  let newOnt : Ontology = <rdf:RDF
6    ↪ xml:base="http://www.semanticweb.org/OntEur"> ( [ <owl:Ontology
7    ↪ rdf:about="OntEur"> [ ] ] @ newClass )
8  in
9  dump_to_file_utf8 "Ontologia.rdf" (print_xml_utf8 newOnt);;

```

Alla linea 1, facendo riferimento alla struttura del tesauruso definita in 3.1, carichiamo il tesauruso; alla riga 2 usiamo una proiezione per estrarre un vettore con tutti gli elementi di tipo `Desc` (che sono i concetti del tesauruso), usiamo la `map` per applicare ad ognuno di questi elementi la funzione `transformClass` (3.4), infine diamo un nome al vettore di classi appena creato in modo da poterlo usare nel seguito della trasformazione.

Alla linea 5 creiamo l'ontologia vera e propria aggiungendo tutte le classi appena create, infine facciamo il dump su file generando un documento XML che potrà essere visualizzato con qualsiasi altro strumento incluso Protégè.

3.6 Versione compatta

Per trasformare il tesauruso in un'ontologia abbiamo definito varie funzioni che ci hanno permesso di trasformare pezzo per pezzo i tag SKOS nei rispetti OWL. Assemblando progressivamente i pezzi abbiamo costruito l'ontologia. Proviamo ora a sfruttare tutti i costrutti forniti da CDuce per riscrivere la trasformazione in una forma più compatta

Listato 3.6: `thesaurus.to_ontology_compact.cd`

```

1  let fun thesaurusToOntology (Thesaurus -> Ontology)
2    <rdf:RDF>[ (concepts :: Desc) * ] ->
3    let newClasses = map concepts with <rdf:Description rdf:about=ab> [
4      ↪ (descAttr :: DescAtt)* ] ->
5      let classAttr = transform descAttr with
6        | x & ScopeNote -> [x]
7        | x & ExactMatch -> [x]
8        | <skos:prefLabel xml:lang=1> lab -> [<rdfs:label xml:lang=1> lab]
9        | <skos:altLabel xml:lang=1> lab -> [<rdfs:label xml:lang=1> lab]
10       | <skos:broader rdf:resource=res> [] -> [<rdfs:subClassOf
11         ↪ rdf:resource=res> []]
12     in
13     <owl:Class rdf:about=ab> classAttr
14   in
15   <rdf:RDF xml:base="http://www.semanticweb.org/OntEur"> ( [ <owl:Ontology
16     ↪ rdf:about="OntEur"> [ ] ] @ newClasses );;
```

Questa funzione permette di passare direttamente da un elemento di tipo `Thesaurus` ad un elemento di tipo `Ontology`. Rispetto alla definizione più estesa abbiamo perso il controllo puntuale sul tipo degli attributi di tipo `ClassAttr`, infatti in questo caso CDuce ci assicura solo che siamo passati correttamente da attributi di un concetto ad attributi di una classe, senza controlli più specifici.

Nonostante questo minore controllo sui tipi questa nuova definizione è molto più compatta rimanendo comunque leggibile: iniziamo con una `map` per trasformare il tag `Description` in un tag `Class`, dentro la `map` innestiamo una `transform` per trasformare gli attributi da SKOS a OWL per ogni elemento estratto con la funzione `map`.

3.7 Aumentare l'espressività

Ora che abbiamo ottenuto un'ontologia possiamo aggiungere relazioni tra gli elementi, ad esempio possiamo mettere in relazione gli oggetti di moda con i materiali o i colori e le

tecniche con i materiali. Possiamo inoltre definire delle relazioni di sottoclasse più complesse di quella puramente tassonomica. In seguito (??) presenteremo un'estensione della struttura dell'ontologia definita qui 3.2 in cui andremo ad arricchire con nuove informazioni l'ontologia qui creata ed useremo la nova versione per sviluppare selezioni e trasformazioni più sofisticate di quelle che avremo potuto definire su un tesoro.

3.8 In Protégé

Provando ad editare il tesoro in Protégé, vediamo che vengono riconosciute come classi quella dei **Concepts** e quella dei **ConceptScheme**, tutti gli elementi del tesoro sono individui. In particolare i 4 concetti fondamentali appartengono alla classe **ConceptScheme**, tutti gli altri alla classe **Concept**. Tra gli strumenti base di refactor di Protégé non ve ne sono per trasformare degli individui in classi, tanto meno mantenendo la struttura gerarchica specificata nel tesoro. Senza voler operare in modo puntuale su ogni individuo per trasformarlo manualmente in una classe (e senza usare i plug-in che permettono di scrivere programmi, ad esempio in java⁴, per operare delle trasformazioni simili a quelle fatte con CDuce) ciò che si può fare è comunque definire le stesse relazioni che abbiamo accennato nel paragrafo precedente (3.7) che avranno in questo caso dominio e range coincidenti (individui di classe **Concepts**). In questo modo possiamo comunque aumentare l'espressività del tesoro, ma con due grandi vincoli:

- non avremo il supporto dei reasoner per verificare la che le relazioni che valorizzeremo siano corrette (ad esempio dato dominio e range della relazione "è fatto di" potremo liberamente creare la relazione: velluto è fatto di alluminio);
- se volessimo definire delle relazioni di sottogenere più complesse dovremo farlo tramite la definizione di relazioni, al posto che con la più naturale definizione di sottoclasse.

3.9 Conclusioni

Nel contesto della re-ingegnerizzazione di un tesoro CDuce si dimostra molto più versatile di Protégé, questo è dovuto principalmente a due fattori:

- come si legge in [2] una parte importante del processo di trasformazione è la conversione sintattica dei tag, in questo CDuce è uno strumento molto efficace
- Protégé è uno strumento per editare ontologie espresse in OWL [3] e, nonostante sia molto potente e versatile, non è lo strumento più indicato per operare su un tesoro.

In questo frangente l'uso di un linguaggio di programmazione che ci permettesse di definire esattamente come manipolare i dati è stato essenziale per ottenere un buon risultato di traduzione. Una volta acquisite delle competenze di base abbastanza solide con CDuce e avendo familiarità con gli strumenti che mette a disposizione, la scrittura di un programma che esegue la traduzione da tesoro ad ontologia non risulta particolarmente complessa o lunga in termini di tempo (come testimonia la funzione 3.6).

Dal punto di vista funzionale abbiamo sfruttato la possibilità di definire funzioni avendo un sofisticato controllo sul tipo delle espressioni che stavamo trattando, questo ha permesso di evitare subito degli errori che con linguaggi imperativi non sarebbero venuti a galla fino al momento dell'esecuzione, in particolare dovendo lavorare alternativamente su vettori ed elementi singoli è capitato spesso che il tipo inferito fosse un vettore, mentre il tipo richiesto fosse un elemento o viceversa (in un linguaggio come C con l'uso dei puntatori un errore del genere avrebbe potenzialmente richiesto molto tempo per essere individuato e risolto).

⁴<https://www.java.com/>

Capitolo 4

Merge di ontologie

introduzione

In questo capitolo vediamo come sia possibile usare CDuce per fare il merge di ontologie, cercando di mettere in risalto quelli che possono essere i vantaggi di usare un linguaggio funzionale rispetto ad uno strumento grafico. Alla fine del capitolo paragoneremo l'approccio con CDuce a quello con Protégé. Per mettere in evidenza i vantaggi di CDuce presentiamo un esempio un po' più complesso di quelli visti in precedenza in modo da far risaltare le potenzialità di CDuce e contemporaneamente dare un'idea di funzioni più complesse.

4.1 Obiettivo del merge

Adesso che abbiamo creato un'ontologia a partire dall'“European Fashion Thesaurus” e abbiamo descritto una semplice ontologia per descrivere persone vorremmo fonderle in una nuova ontologia che possa parlare degli usi e costumi della società, magari riferita ad un ben preciso periodo storico. Vogliamo creare un'ontologia in cui sia possibile rappresentare delle persone, coi rispettivi legami di parentela e nella quale sia possibile associare uno status sociale, un lavoro e dell'abbigliamento alle persone.

4.1.1 Ontologie di partenza

Per creare questa ontologia consideriamo 3 ontologie di partenza:

- **society**: è lo scheletro dell'ontologia che vogliamo ottenere, è comodo crearla a priori in modo da non dover creare ex novo tutta la struttura in CDuce, inizialmente rappresenta delle persone sulle quali definiamo le relazioni **born_in**, **is** e **work_as** che associano una persona con la città di nascita, con il suo status sociale e con il proprio lavoro;
- **people**: rappresenta gli individui con le loro parentele, rispetto all'ontologia presentata in 1.1 abbiamo aggiunto la relazione simmetrica **marry**, e la relazione **son_of** come inversa di **parent_of**, aggiungiamo anche l'anno di nascita;
- **fashion**: l'ontologia che abbiamo creato nel capitolo 3 arricchita con le relazioni **made_of**, **crafted_with**, **color**, che indicano rispettivamente i materiali costituenti, le tecniche realizzative e i colori di un capo d'abbigliamento.

4.1.2 Ontologia di arrivo

Vogliamo modificare **society** in modo da descrivere gli usi e i costumi della società del XII secolo, per fare questo estrapoliamo dall'ontologia **fashion** tutti i vestiti che non siano costituiti da fibre artificiali e scremiamo dall'ontologia **people** tutte le persone che sono vissute in un'epoca che non ci interessa. Vogliamo poi che la classe **People** definita in **society** sia equivalente alla classe **People** definita in **people**, senza perdere tutte le relazioni di parentela, e potendo contemporaneamente sfruttare le nuove relazioni definite in **society**.

4.2 Struttura generica di un'ontologia

Per fare il merge dobbiamo poter caricare le 3 ontologie in CDuce, quindi descriviamo una struttura generale per fare il parsing di un'ontologia generica:

Listato 4.1: general structure

```
1 type Ontology = <rdf:RDF xml:base=String> [ Thing * ]
2 type Thing = Ont | AnnProperty | ObjProperty | DataProperty | Class |
   ↳ Individual
3
4 type Ont = <owl:Ontology rdf:about=String> []
5
6 type AnnProperty = <owl:AnnotationProperty rdf:about=String> []
7
8 type DataProperty = <owl:DatatypeProperty rdf:about=String> [ PropAttr* ]
9 type ObjProperty = <owl:ObjectProperty rdf:about=String> [ PropAttr * ]
10
11 type PropAttr = Inverse | Domain | Range | PropType
12 type Inverse = <owl:inverseOf rdf:resource=String> []
13 type Domain = <rdfs:domain rdf:resource=String> []
14 type Range = <rdfs:range rdf:resource=String> []
15 type PropType = <rdf:type rdf:resource=String> []
16
17 type Class = <owl:Class rdf:about=String> [ ClassAtt * ]
18 type ClassAtt = SubClass | EqClass | Label | Note | Dictionary
19 type SubClass = <rdfs:subClassOf rdf:resource=String> []
20 type EqClass = <owl:equivalentClass> [ EqAttr ] | <owl:equivalentClass
   ↳ rdf:resource=String> []
21 type EqAttr = <owl:Restriction> [ AnyXml* ]
22 type Label = <rdfs:label xml:lang=String> String
23 type Note = <skos:scopeNote xml:lang=String> String
24 type Dictionary = <skos:exactMatch rdf:resource=String> []
25
26 type Individual = <owl:NamedIndividual rdf:about=String> [ IndAttr * ]
27 type IndAttr = IndClass | IndProp | DataProp
28 type IndClass = <rdf:type rdf:resource=String> []
29 type IndProp = <_ rdf:resource=String> []
30 type DataProp = <_ rdf:datatype=String> String
```

Rispetto alle strutture definite in precedenza notiamo varie aggiunte:

- si specifica meglio l'elemento `EqClass`, questo infatti può essere una restrizione (un genitore è una persona con dei figli) oppure un'equivalenza senza condizioni (ci serve per rendere uguali i concetti di persona definiti nelle 2 ontologie);
- definiamo il tipo `DataProperty`, questo rappresenta una proprietà degli individui, nel nostro caso la usiamo per specificare l'anno di nascita di una persona. Gli individui nella loro lista di attributi potranno ora averne uno di tipo `DataProp`;
- il tipo `AnnProperty` serve perché avendo editato l'ontologia `fashion` in protégé per aggiungere le relazioni non presenti nel thesaurus i tag SKOS sono stati correttamente riconosciuti [1] e classificati come tag `owl:AnnotationProperty`;

4.3 Merge

4.3.1 Funzioni utili

Prima di passare alla costruzione della nuova ontologia vediamo alcune funzioni utili che si possono applicare a qualsiasi ontologia e che sono servite per costruire le funzioni specifiche adatte a manipolare la particolare ontologia di interesse

Listato 4.2: usefull function

```
1 let fun loadOntology (Latin1 -> Ontology)
2   x -> load_xml x :? Ontology;;
3
4 let fun head ([ Any* ] -> Any)
5   | ([ x ] @ _) -> x
6   | [] -> "error empty list";;
7
8 let fun subClasses (Class -> Ontology -> [ Class* ])
9   <owl:Class rdf:about=ab> [ _* ] -> fun (Ontology -> [ Class* ])
10  ont ->
11    select x from
12      x in [ont]/Class,
13      y in [x]/SubClass
14    where (y = <rdfs:subClassOf rdf:resource=ab> []);;
15
16 let fun subClassesRec ( Class -> Ontology -> [ Class* ])
17   cl -> fun (Ontology -> [ Class* ])
18   ont ->
19     let subCl = subClasses cl ont in
20     subCl @ flatten (map subCl with y -> subClassesRec y ont)
21
22 let fun andArray ( [ Bool* ] -> Bool )
23   | ([ x ] @ y) -> (x && (andArray y))
24   | [] -> `true;;
25
26 let fun orArray ( [ Bool* ] -> Bool )
27   | ([ x ] @ y) -> (x || (orArray y))
```

```

28 | [] -> `false;;
29
30 let fun classOf (Individual -> Ontology -> [Class*])
31   <owl:NamedIndividual ..> [ (tp::IndClass | _) *] -> fun (Ontology ->
32     ↳ [Class*])
33   ont ->
34     transform tp with <rdf:type rdf:resource=str> [] ->
35     transform [ont]/Class with x ->
36       match x with <owl:Class rdf:about=a>[ _* ] -> if a = str then [x]
37       ↳ else [];;
38
39 let fun contains (Any -> [Any*] -> Bool)
40   obj -> fun ([Any*] -> Bool)
41   lst ->
42     let intersect = select x from x in lst
43     where (x = obj) in
44     match intersect with
45     | [] -> `false
46     | [ Any* ] -> `true;;
47
48 let fun isInClasses (Individual -> [Class*] -> Ontology -> Bool)
49   ind -> fun ([Class*] -> Ontology -> Bool)
50   classes -> fun (Ontology -> Bool)
51   ont ->
52     let c1 = classOf ind ont in
53     let res = map c1 with x -> contains x classes in
54     andArray res;;

```

Per la prima volta abbiamo fatto ricorso a applicazioni parziali (curried functions¹), questo è utile per poter parametrizzare anche l'ontologia di riferimento per l'operazione; dato che in questo caso ne manipoliamo contemporaneamente 3 è importante poter specificare volta per volta a quale ci riferiamo.

Vediamo per la prima volta un esempio di funzione ricorsiva: `subClassRec` richiama ricorsivamente se stessa per costruire l'intero albero di sottoclassi a partire da una classe data (e dall'ontologia di riferimento). Ovviamente perché questa funzione possa terminare la struttura delle classi deve essere un grafo aciclico (questa non è una grossa limitazione, infatti se la classe `c2` è contemporaneamente sopraclasse e sottoclasse di `c1` allora `c1` e `c2` sono equivalenti e possono essere accorpate per eliminare i cicli, discorso analogo vale per cicli più lunghi).

Per verificare se un individuo si trova in un certo albero di classi si potrebbe operare al contrario rispetto a `subClassRec`, risalendo l'albero fino a quando non si trova il padre desiderato (restituendo `true`) oppure la radice delle classi (restituendo `false`). Questo approccio, però, presenta delle problematiche:

- un individuo può appartenere a più classi, quindi bisognerebbe risalire n alberi dove n è il numero di classi a cui appartiene l'oggetto;
- ogni classe può essere sottoclasse di più classi diramando così ulteriormente la ricerca.

Facendo alcuni test si nota che un approccio di questo genere, oltre ad essere impegnativo dal punto di vista implementativo è anche poco efficiente dal punto di vista prestazionale.

¹<https://en.wikipedia.org/wiki/Currying>

Infatti come vedremo dopo per ogni capo d'abbigliamento dovremo chiederci se è costituito da materiali artificiali e questo rallenta il processo di merge; per evitare questo problema costruiamo solo una volta la lista di materiali artificiali e quando dobbiamo stabilire se un materiale è naturale o meno verifichiamo se appartiene alla lista dei materiali artificiali. Questo approccio è vantaggioso in quanto la lista di materiali artificiali andrebbe creata in ogni caso per andare ad aggiungere alla nuova ontologia tutti i materiali che non lo sono (conviene creare la lista dei materiali artificiali piuttosto che quella dei materiali naturali perché la prima contiene molti meno elementi, di conseguenza è più veloce da creare).

Elenco listati

1.1	persone.rdf	1
2.1	persone.cd	4
2.2	basic functions	6
2.3	basic Query	7
3.1	thesaurus_europeana.cd	10
3.2	ontology_europeana.cd	11
3.3	SKOS_to_OWL.cd	12
3.4	concept_to_class.cd	13
3.5	thesaurus_to_ontology.cd	13
3.6	thesaurus_to_ontology_compact.cd	14
4.1	general structure	17
4.2	usefull function	18

Bibliografia

- [1] Sean Bechhofer e Alistair Miles. *Using OWL and SKOS*. 2008. URL: <https://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html>.
- [2] Daniel KLESS, Ludger JANSEN, Jutta LINDENTHAL e Jens WIEBENSOHN. *A method for re-engineering a thesaurus into an ontology*. 2012. URL: https://jansenludger.github.io/home/Texte/Kless%20et%20al_A%20method%20for%20reengineering%20a%20thesaurus%20into%20an%20ontology%20_F0IS%202012_Preprint.pdf.
- [3] *Protégé 5 Documentation*. URL: <https://protegeproject.github.io/protege/getting-started/>.