

Capitolo 1

Il modello dei dati ADM

Nella metodologia sviluppata nel progetto Araneus, un sito web può essere descritto attraverso uno schema definito utilizzando un opportuno modello logico chiamato Araneus Data Model (ADM). Tale modello consente di descrivere i dati d'interesse in modo indipendente dalla loro locazione e rappresentazione fisica. *ADM* può essere definito *orientato alle pagine* perché mira a descrivere le informazioni contenute nelle singole pagine che costituiscono gli elementi base dell'ipertesto.

Nel seguito vengono presentati i costrutti di ADM, e verranno introdotti molti esempi, ispirati al sito web del Dipartimento di Ingegneria e Fisica dell'Ambiente (*DIFA*) dell'Università di Potenza per migliorare la spiegazione dei concetti presentati.

1.1 ADM Page Scheme

Il modello ADM si basa sul concetto di *page-scheme*, che somiglia al concetto di schema di relazione in una base di dati relazionale o di classe una base di dati object-oriented: una *page-scheme* è una descrizione dettagliata di un insieme di pagine web con caratteristiche comuni. Un'istanza di un *page-scheme* è una pagina web, che può essere considerata al pari di un oggetto: è infatti identificabile univocamente dal proprio *URL* e risulta costituita da un insieme di attributi. Ogni attributo corrisponde a gruppi di informazioni individuabili nella pagina. Gli attributi possono essere rappresentati da testo, immagini, link ad altre pagine, etc. , oppure possono assumere una struttura più complessa come ad esempio liste di elementi, tabelle, etc..

1.1.1 Pagine Uniche

All'interno di un sito esistono però anche pagine uniche, cioè pagine la cui struttura è completamente diversa da quella di tutte le altre, un esempio tipico di pagina unica è proprio la home page.

1.2 Attributi

In ADM il contenuto di una pagina web è descritto per mezzo di *attributi*, questi possono avere una struttura semplice o complessa.

1.2.1 Attributi Semplici

Gli attributi semplici sono *monovalore* (atomici), cioè hanno una struttura semplice, come ad esempio stringhe di testo, immagini o link ad altre pagine. Gli attributi di tipo link sono usati per modellare i links ipertestuali: ogni link è costituito da un'*ancora* e un *riferimento*, quest'ultimo è l'URL della pagina di destinazione, l'*ancora* è un testo (stringa costante o tuple di attributi) o un'immagine.

Analizziamo qualche pagina del sito del Dipartimento di Ingegneria e Fisica dell'Ambiente (*DIFA*) dell'Università di Potenza.

Ad esempio nella Figura 1.1, rappresentante la pagina della ricercatrice Aurelia Sole, è possibile individuare un insieme di attributi semplici: la qualifica ('Ricercatore'), l'indirizzo ('D.I.F.A. - Università della Basilicata Via della Tecnica, 3 85100, Potenza-- Italia'), il numero di telefono ('0971 - 474616_682') e il numero di fax ('0971 - 56537') tutti di tipo *testo* e una lista di link alle aree di ricerca di principale interesse (link che risulta individuabile dal carattere sottolineato).



Figura 1.1: sito del DIFA , pagina della professoressa Aurelia Sole.

Altre pagine del sito sono organizzate con la stessa struttura. Si prenda in considerazione la Figura 1.2 rappresentante la pagina del professor Franco Amatucci. In essa si possono individuare gli stessi attributi presenti nella pagina della ricercatrice Aurelia Sole: la qualifica, l'indirizzo, il numero di telefono, il numero di fax e una lista di link alle aree di ricerca di principale interesse. In questa pagina non è presente l'attributo E-mail di tipo link che però è presente nella pagina della ricercatrice Aurelia Sole, per tale motivo questo attributo deve essere considerato opzionale. Da ciò si intuisce come allora insiemi di pagine che possiedono informazioni simili possono essere descritte attraverso uno stesso page-scheme.



Figura 1.2: sito del DIFA , pagina del professor Franco Amatucci.

La struttura della pagina può essere descritta dallo schema presentato nella Figura 1.3; in tale Figura si sono indicati:

cinque attributi semplici di tipo testo, denominati Nome, Cognome, Qualifica, Indirizzo, Telefono, Fax;

un attributo semplice di tipo link, denominato Email che permette di passare alla pagina relativa all'indirizzo E-mail del ricercatore;

un attributo di tipo lista *ListaGruppiRicerca*.



Figura 1.3: schema per la pagina in Figura 1.1.

1.2.2 Attributi Composti

Gli attributi composti sono *multivalore* e rappresentano una collezione (ordinata) di oggetti. In ADM gli attributi multivalore sono descritti come liste di ennuple, eventualmente nidificate. Nell'esempio di Figura 1.3 è riportato anche un attributo complesso, che è la lista di aree di ricerca di principale interesse. Questo attributo multivalore è modellato come una lista, i cui elementi sono links, le cui ancore sono rappresentate dalla stringa sottolineata.

1.3 Unione Eterogenea

Il modello ADM fornisce il tipo, *unione eterogenea*, in accordo con la natura eterogenea del Web. Per esempio, si immagina che sia i **Professori** sia i **Ricercatori** possano essere autori di pubblicazioni e che si voglia implementare un percorso di navigazione che, partendo dalla pagina **Pubblicazione** conduce ai rispettivi autori, senza voler sapere se un dato autore è un professore o un ricercatore. Gli autori saranno istanze delle pagine **Professori** e **Ricercatori**, che potrebbero avere una struttura completamente diversa tra loro. Il tipo unione eterogenea è usato per soddisfare questo fine: si parte dalle istanze del page-scheme **Pubblicazione** e si arriva all'unione disgiunta delle istanze delle pagine dei **Professori** e dei **Ricercatori**.

Graficamente tale tipo è rappresentato tramite la lettera “U” racchiusa in un cerchio.

La struttura della pagina di tale esempio può essere descritta dallo schema presentato nella Figura 1.4.

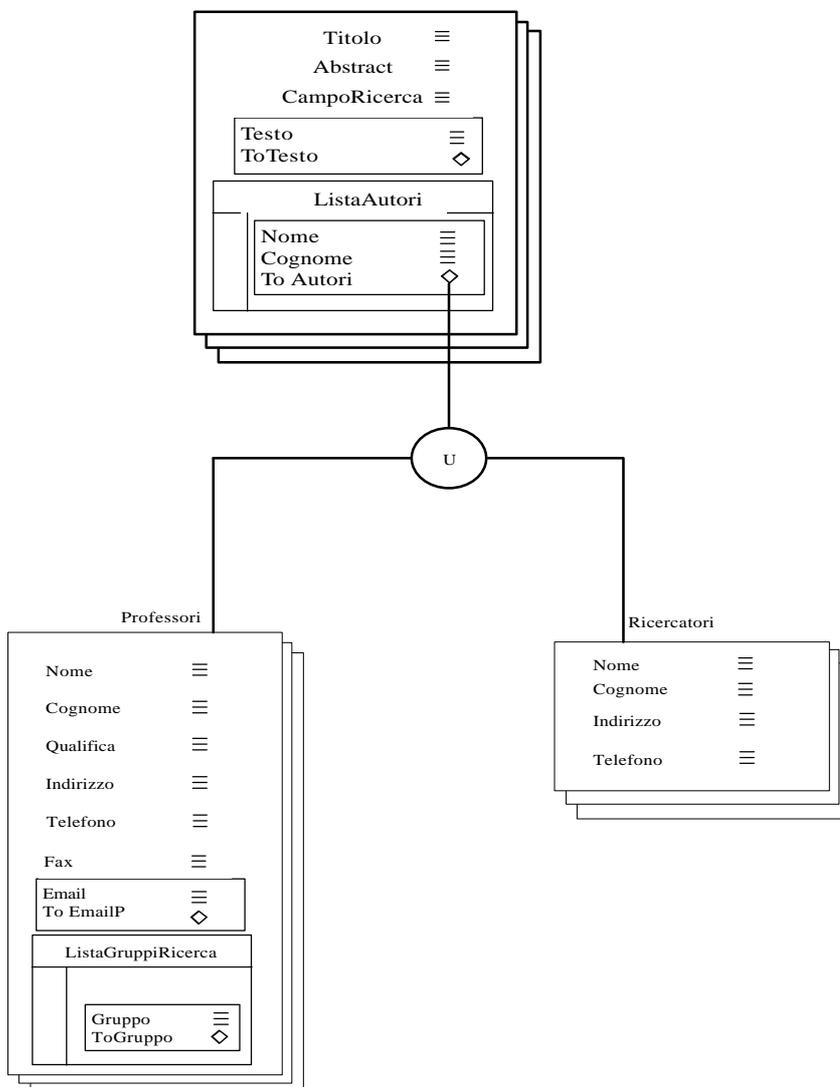


Figure 1.4: Esempio di schema in cui vi è il tipo unione eterogenea

1.4 I costrutti del modello

Il modello ADM fornisce due tipi semplici, testo (*TEXT*) e immagini (*IMAGE*). A partire dai tipi elementari è possibile definire tutti e soli i tipi validi nel modello *ADM*.

Dato un insieme di tipi elementari *I*, un insieme di nomi di attributi *A* e un insieme di nomi di pagine *P*, l'insieme dei tipi validi in ADM I_{ADM} può essere definito ricorsivamente come segue:

Ogni tipo elementare è un tipo *monovalore* valido per *ADM*;

es. Nome: TEXT;

LINK-TO *P* è un tipo *monovalore* valido per *ADM*, per ogni nome di pagina *P*;

es. ToEmail: LINK-TO EmailP;

LIST-OF ($A_1:T_1, A_2:T_2, \dots, A_n:T_n$) è un tipo *multivalore* valido per *ADM*, se A_1, \dots, A_n sono nomi di attributi e T_1, \dots, T_n sono tipi validi per *ADM*.

es. ListaGruppiRicerca: LIST-OF (GruppoRicerca: LINK-TO Gruppo);

Gli attributi A_i possono eventualmente essere etichettati con la parola *optional*; in questo caso è consentito che le istanze degli attributi ammettano valori nulli.

Come si può intuire, i tipi monovalore rappresentano attributi semplici che possono assumere uno e un solo valore alla volta. Al contrario variabili di tipo multivalore possono assumere contemporaneamente una *lista* finita di valori diversi.

A loro volta, questi valori possono essere istanze di attributi monovalore o multivalore, dando luogo a possibili nidificazioni. In questo modo è possibile rappresentare anche le strutture più complesse presenti nell'ipertesto quali liste o tabelle di elementi. Un *page-scheme*, che nel modello *ADM* assume la forma

$PS(A_1:T_1, \dots, A_n:T_n)$, è valido se A_1, \dots, A_n sono nomi di attributi e T_1, \dots, T_n sono tipi validi per *ADM*. Il page-scheme può essere eventualmente etichettato con la parola *unique*, in questo caso l'istanza del page-scheme sarà costituita da una unica pagina web. Uno schema ADM $S(PS_1, \dots, PS_n)$ è un insieme di page-scheme PS. Ad esempio, si prenda in considerazione il frammento del sito Web relativo al Dipartimento di Ingegneria e Fisica dell'Ambiente (*DIFA*) dell'Università di Potenza rappresentato in figura 1.5; nella stessa figura è riportata anche una legenda che descrive il significato dei simboli utilizzati.

Lo schema è costituito da un insieme di page-scheme organizzati come segue:

- all'interno della home page del sito, ovvero della pagina, possono essere individuati una serie di attributi di tipo link; tra questi, l'attributo `ToPersonale` permette di accedere alla pagina relativa al personale del DIFA, mentre l'attributo `ToRicerca` permette di accedere alla pagina relativa alle attività di ricerca;
- dalla pagina `PersonaleP` è possibile accedere alla pagina `ListaPersonaleDocenteP` attraverso l'attributo di tipo link `ToListaPersonaleDocente`;
- nella pagina `ListaPersonaleDocenteP` è riportata una lista di link uno per ogni ricercatore o professore, attraverso questi link si può accedere all'insieme di pagine relative ai docenti;
- nella pagina `DocentiRicercaP` sono presenti i seguenti attributi: `Nome`, `Cognome`, `Qualifica`, `Indirizzo`, `Telefono` e `Fax` di tipo testo, `Email` di tipo link e `ListaGruppiRicerca` di tipo lista, quest'ultima è composta da una serie di link che permettono di accedere alla pagina `GruppoRicercaP`;
- la pagina `GruppoRicercaP` contiene una breve descrizione del gruppo, un attributo di tipo link che accede alle pagine relative alle aree di ricerca, due liste di link di cui una relativa ai docenti e l'altra relativa agli studenti afferenti al gruppo. A tale pagina si accede anche attraverso la pagina `ListaGruppiRicercaP` nella quale è riportata una lista di link;
- l'ultima pagina rimasta è `RicercaP` nella quale ci sono tre link.

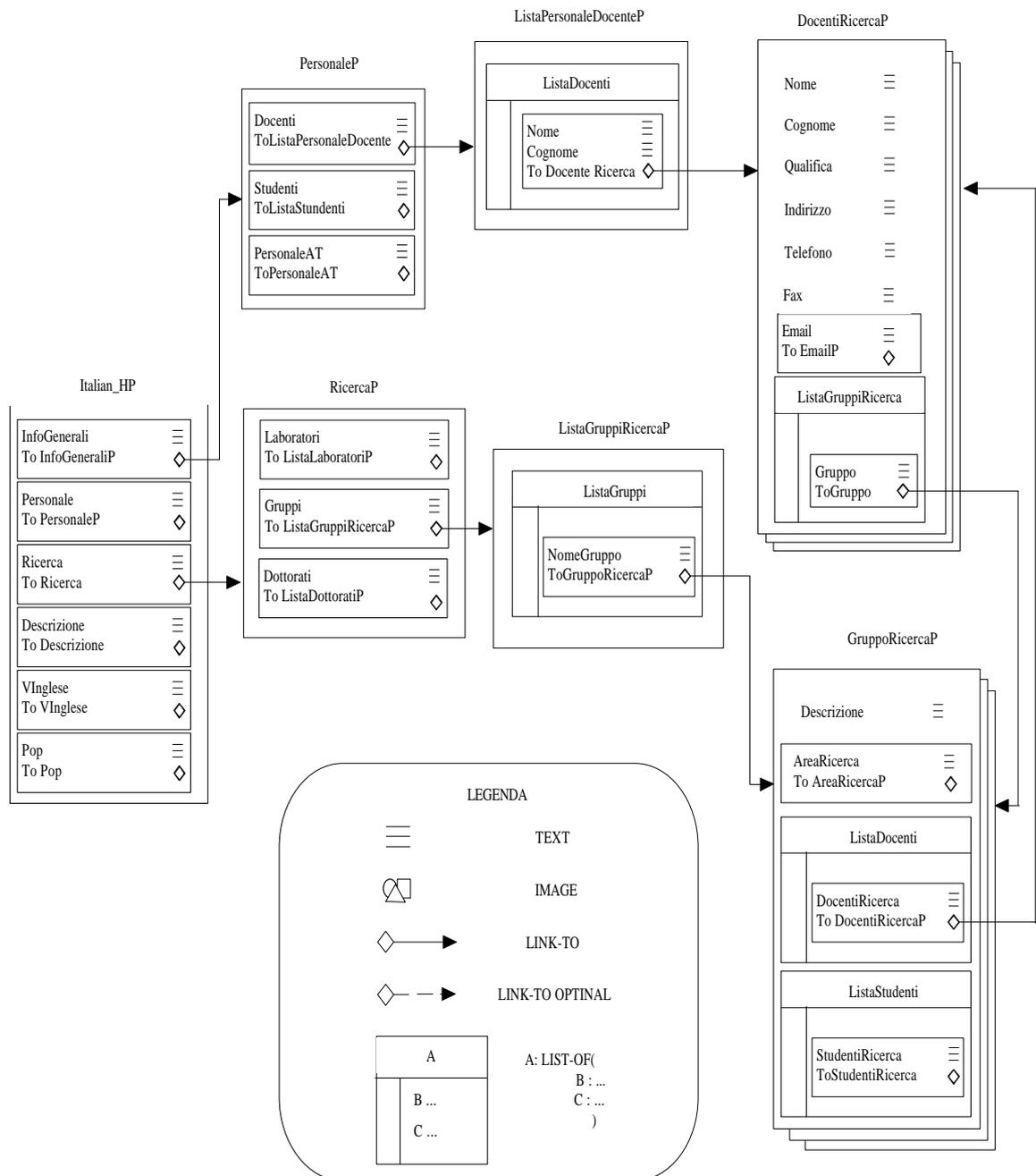


Figura 1.5: Schema del dipartimento di Ingegneria e Fisica dell’Ambiente (DIFA) dell’Università di Potenza.

Ad ogni pagina è quindi possibile associare uno *schema di pagina*, o *page-scheme* in grado di descrivere le informazioni d’interesse in essa contenute nella pagina.

Lo schema descrive l'organizzazione di tutte le pagine del sito aventi la struttura corrispondente.

Un sito Web viene quindi modellato mediante uno schema composto da un insieme finito di page-scheme. Lo schema costituisce una vista strutturata del sito in grado di definirne l'organizzazione dei dati. Per esempio una rappresentazione grafica di una porzione del sito del DIFA è riportata, congiuntamente alla legenda, in Figura 1.5. Ogni rettangolo presente nella figura corrisponde a un page-scheme (unico se è singolo); il nome che compare in grassetto, appena al di sopra del rettangolo, indica il nome del page-scheme stesso.

All'interno del singolo rettangolo sono presenti i nomi degli attributi definiti nel corrispondente page-scheme. Vicino a ciascun attributo monovalore è posto un simbolo che ne indica il tipo (TEXT, IMAGE, LINK-TO). Gli attributi multivalore (LIST-OF) sono rappresentati tramite una tabella che contiene il nome dell'attributo e tante righe quanti sono gli attributi presenti nella definizione del tipo multivalore. Inoltre sono tracciate delle linee per evidenziare i link presenti tra i page-scheme corrispondenti.

Possiamo notare che nella descrizione della struttura di ogni pagina ci sono elementi che non abbiamo preso in considerazione. Ad esempio nella pagina `PersonaleP` ci sono anche due menu di servizio, uno nell'header e l'altro nel footer, un'immagine e alcune stringhe costanti, poiché però tutti questi elementi fanno parte della presentazione della pagina e non della sua struttura, non vengono riportati nello schema *ADM*. Lo stesso discorso è valido per tutte le altre pagine del sito.

1.5 Il linguaggio di definizione degli schemi

Per la descrizione degli schemi è stato introdotto un opportuno linguaggio di definizione dei dati (DDL). La sintassi del DDL è estremamente semplice, tanto da poterne dare una definizione sintetica per mezzo della seguente grammatica:

Schema → **SCHEME** NomeSchema BaseURL Pages **END-SCHEME**

Pages → **PAGE-SCHEME** NomePagina Tipo ListaAttributi **END** Pages | ε

Tipo → **UNIQUE** RelativeURL | ε

ListaAttributi → NomeAttributo : TipoAttributo Optional ; ListaAttributi | ε

TipoAttributo → **IMAGE** | **TEXT** | **LINK-TO** NomePagina | **LIST-OF** NomePagina

Optional → **OPTIONAL** | ε

La definizione di uno schema deve iniziare con la parola chiave **SCHEME** seguita dal nome dello schema, dall'indirizzo del sito, dalla lista dei page-scheme ed infine conclusa da **END-SCHEME**. L'indirizzo deve indicare il protocollo di connessione e la locazione logica o fisica del sito. Il protocollo può assumere due forme diverse:

1. <http://HostName/HostDir> nel caso si tratti di un sito Web; ad esempio <http://www.difa.unibas.it/>;
2. file:/FileSystemDir nel caso s'intenda utilizzare il mirror del sito memorizzato sul file system; ad esempio file:/afs/.../db/ditomass/mirror/DIFA/.

La definizione di ogni page-scheme deve iniziare con la parola chiave **PAGE-SCHEME** seguita dal nome della pagina, dalla lista degli attributi in esso presenti e infine conclusa da **END**. Se si tratta di una pagina unica, il nome della pagina è seguito dalla parola chiave **UNIQUE** e dall'indirizzo della pagina che può essere assoluto o relativo rispetto all'indirizzo del sito, già precedentemente specificato. La definizione di ciascun attributo avviene attraverso l'indicazione del nome e del tipo, eventualmente seguita dalla parola chiave **OPTIONAL** e conclusa dal ';'. La dicitura **OPTIONAL** indica che l'attributo può assumere valori nulli. Il tipo di attributo può

essere semplice (TEXT, IMAGE, LINK-TO) oppure complesso (LIST-OF). In questo secondo caso la parola chiave è seguita dalla lista degli attributi, costituenti il tipo complesso, racchiusa tra le parentesi tonde.

Le dichiarazioni corrispondente al sito del DIFA, il cui schema è mostrato in Figura 1.5, è riportata nelle pagine che seguono.

Sito Web del DIFA

SCHEME Difa <http://www.unibas.it/>

```
DEFINE PAGE ItalianHP UNIQUE
  ToInfoGenerali : LINK-TO InfoGeneraliP;
  ToPersonale : LINK-TO PersonaleP;
  ToRicerca : LINK-TO RicercaP;
  ToDescrizione : LINK-TO WebMasterP;
  ToVInglese : LINK-TO IngleseHP;
  ToPop : LINK-TO PopP;
END

DEFINE PAGE PersonaleP UNIQUE
  ToDocenti : LINK-TO ListaPersonaleDocenteP ;
  ToStudenti : LINK-TO ListaStudentiP;
  ToPersonaleAT : LINK-TO ListaPersonaleATP;
END

DEFINE PAGE RicercaP UNIQUE
  ToLaboratori : LINK-TO ListaLaboratoriP ;
  ToGruppi : LINK-TO ListaGruppiRicercaP ;
  ToDottorati : LINK-TO ListaDottoratiP ;
END

DEFINE PAGE ListaPersonaleDocenteP UNIQUE
  ListaDocenti : LIST-OF (
```

```
ToDocenteRicerca : LINK-TO DocentiRicercaP
                                );
```

```
END
```

```
DEFINE PAGE ListaGruppiRicercaP UNIQUE
  ListaGruppi : LIST-OF (
    ToGruppoRicerca : LINK-TO GruppoRicercaP
  );
```

```
END
```

```
DEFINE PAGE DocentiRicercaP
  Nome : TEXT ;
  Cognome : TEXT ;
  Qualifica : TEXT ;
  Indirizzo: TEXT;
  Telefono : TEXT;
  Fax: TEXT;
  ToEmail : LINK-TO ExternalP
  ListaGruppiRicerca : LIST-OF (
                                ToGruppo : LINK-TO GruppoRicercaP
                                );
```

```
END
```

```
DEFINE PAGE GruppoRicercaP
  Descrizione : TEXT = <Descrizione>;
  ToAreeRicerca : LINK-TO AreeRicercaP ;
  ListaDocenti : LIST-OF (
    ToDocenteRicerca : LINK-TO DocentiRicercaP ;
  );
  ListaStudenti : LIST-OF (
    ToStudentiRicerca : LINK-TO StudentiP;
  );
```

```
END
```