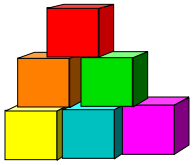
XXIII. Website Design

The Web
Hypertext Data Independence
Data Models for Hypertext Documents
The Araneus Data Model (ADM)
The Navigational Conceptual Model (NCM)
The Araneus Methodology for Website Design



The Web

- The spread of the World-Wide Web (hereafter "Web") technology is one of the most remarkable phenomena of the last few years in all areas of computing and communication technologies.
- The Web (e.g., Web browsers) is becoming a standard interface for the general public to access and exchange information:
  - ✓ The protocol is very simple and public;
  - ✓ The interface is uniform;
  - ✓ The content is extremely rich (both in breadth and in depth);
- The Web is becoming a standard interface for accessing many services, specifically information systems and databases of every type.

Web Features and Open Problems

- The Web is a simple and powerful data integration tool.
- Two basic approaches to Web-based data integration:
  - ✓ Coarse-grain: pages of hypertext;
  - ✓ Fine-grain: unified interface for accessing different (usually similar) information systems available on the Web.
- The Web is built out of semi-structured (HTML/XML) documents, databases contain structured (I.e., tuple/record) data.
- Databases can be queried in a flexible way; hypertext documents are easy to access, but cannot be "queried".
- Web sites are often difficult to explore, use and monitor.
- Web sites are also difficult to design and maintain.

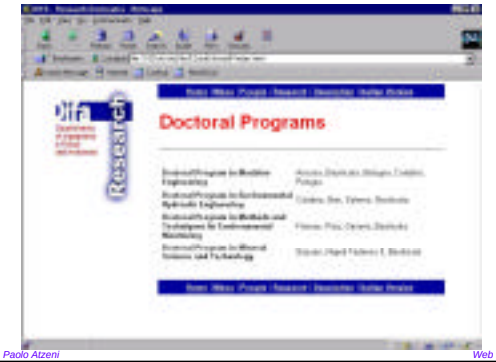
Problems with Large Websites

- Information is often poorly organized and difficult to access.
- It is often unclear what information is available on a given website.
- The access structure of many websites is casual and idiosyncratic, causing frequent dangling references.
- The style of presentation is often heterogeneous.
- Large websites are usually difficult to update, or change their structure.
- It is also difficult to change the presentation structure and/or details.

Data Independence for Hypertext Documents

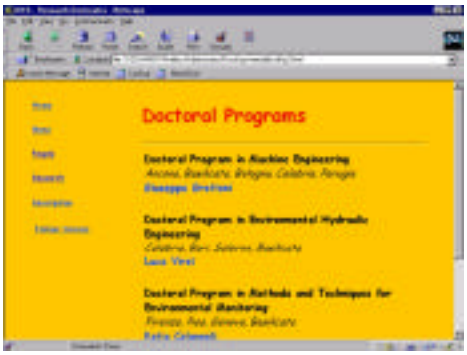
- You might say that there are three facets to the Web:
  - ✓ ***Data*** -- what information is offered through the site and what are the conceptual details and the logical organization.
  - ✓ ***Hypertext*** -- how data is arranged in pages and what navigation links correlate them.
  - ✓ ***Presentation*** -- the appearance of each piece of information on each pages.
- As much as possible, we'd like to decouple the three, so that changes to one affect minimally the other two facets of the Web.

An Example



...Another...

...and Another...



Design Issues

- ***Data*** -- choose the content
- ***Hypertext*** -- choose navigation paths
- ***Presentation*** -- define layout and graphics

Maintenance Issues

- ***Data*** -- changing the content
- ***Hypertext*** -- changing navigation paths
- ***Presentation*** -- changing layout and graphics

Components and Models

Data	**ER and Relational**
Hypertext	
Presentation	**HTML**

What is missing is a model for hypertexts!

Models for Hypertext Documents

- In ***data-intensive*** Websites (and often in general) there are (many) pages with a similar (or even the same) structure.
- Forty years ago people realized that in an application it is often the case that there are many records ***with the same structure***; files and file technology were invented to exploit this fact.
- Likewise, the notion of a ***schema*** for a database was later introduced as an overall description of the content of a database.
- We need something similar for the Web!

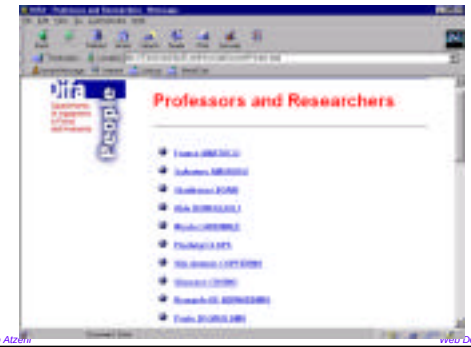
A Web Page

A Page Schema: ProfessorPage

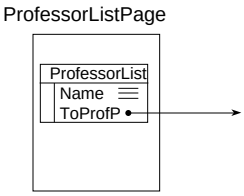
ADM (Araneus Data Model): A Logical Model for Hypertext Documents

- Developed at the University of Rome III (Universita di Roma Tre) by Paolo Atzeni, Paolo Merialdo, Giansalvatore Mecca and colleagues.
- Its features include:
  - ✓ Page schemas
  - ✓ Simple attributes
    - ✓ text, images, ...
    - ✓ link (anchor, URL)
  - ✓ Complex attributes such as lists, possibly nested.
  - ✓ A heterogeneous union operation.
  - ✓ Forms as virtual lists over form fields and links to a result.

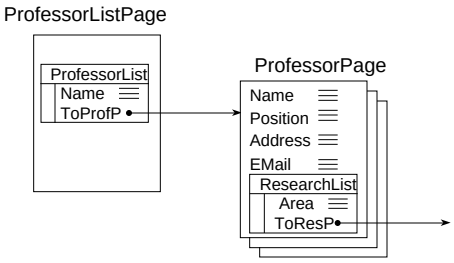
Another Web Page -- Containing a List of Links



A Page Schema for ProfessorListPage

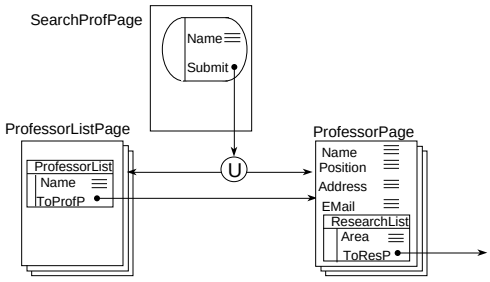


An ADM Schema



Heterogeneous Union and Forms

Heterogeneous Union and Forms



Data Models, Again

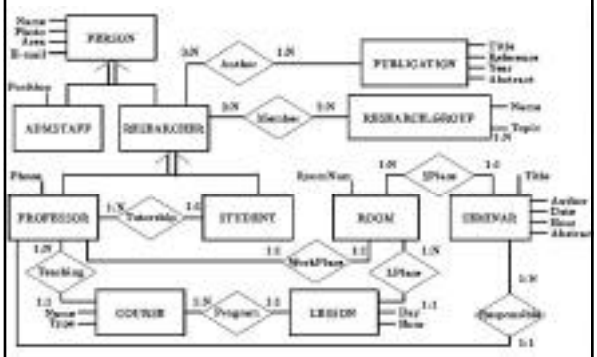
ER — Database Conceptual Schema *(entities - relationships)*

ADM — Hypertext Logical Schema *(page schemas, links)*

There is considerable conceptual distance between the two!

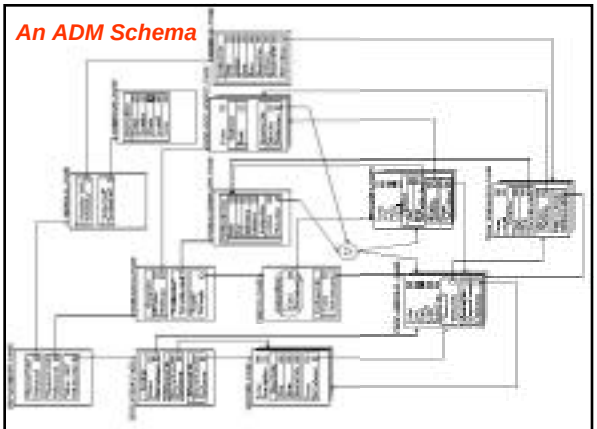
© 2001 Paolo Atzeni

A Simple ER Schema

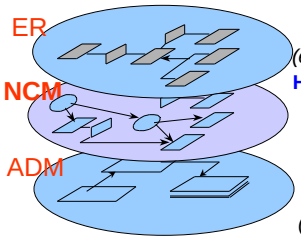


© 2001 Paolo Atzeni

An ADM Schema



The Navigational Conceptual Model (NCM)



Database Conceptual Schema *(entities - relationships)*

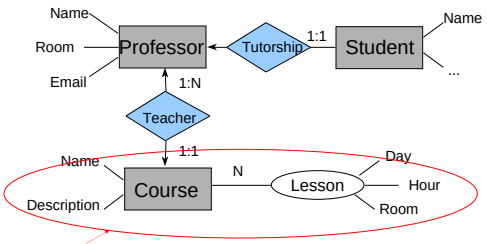
Hypertext Conceptual Schema *(macroentities, directed relationships, aggregations)*

Hypertext Logical Schema *(page schemas, links)*

NCM fills the gap between the two

© 2001 Paolo Atzeni

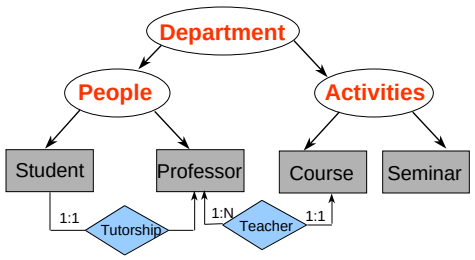
Macroentities and Directed Relationships



Macroentity (was two entities)

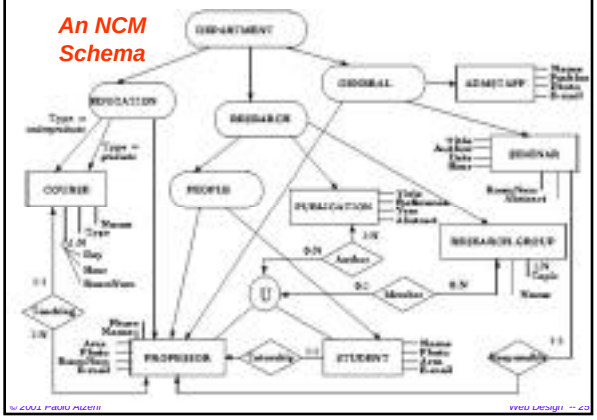
© 2001 Paolo Atzeni

Aggregation Nodes

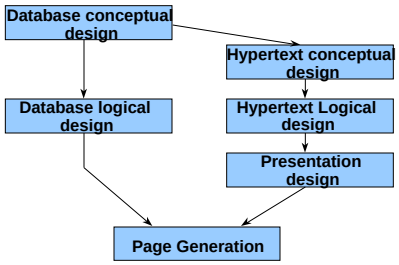


© 2001 Paolo Atzeni

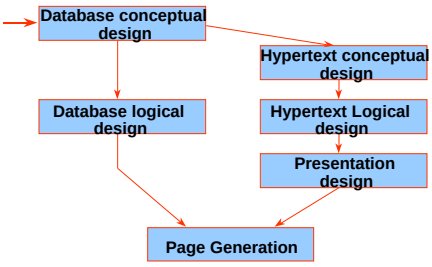
An NCM Schema



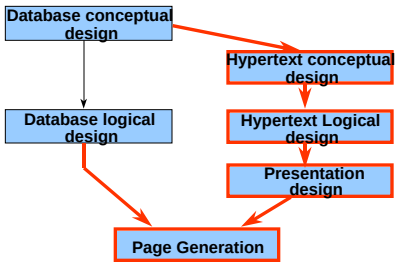
The Araneus Methodology



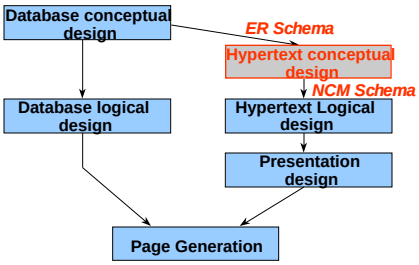
The Araneus Methodology: Design from Scratch



The Araneus Methodology: Design from an Existing Database

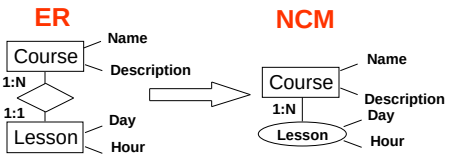


Hypertext Conceptual Design: From an ER schema to a NCM Schema



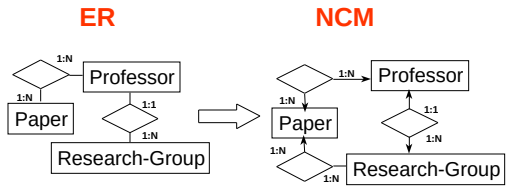
Hypertext Conceptual Design: Step 1

Choose and describe macroentities; design views over the input ER schema.



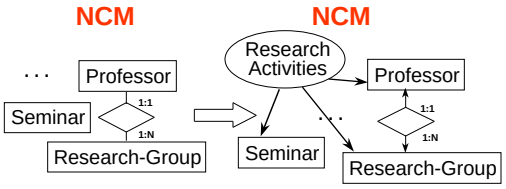
Hypertext Conceptual Design: Step 2

Choose navigation paths

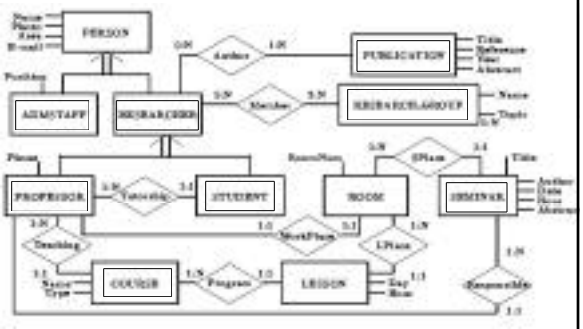


Hypertext Conceptual Design: Step 3

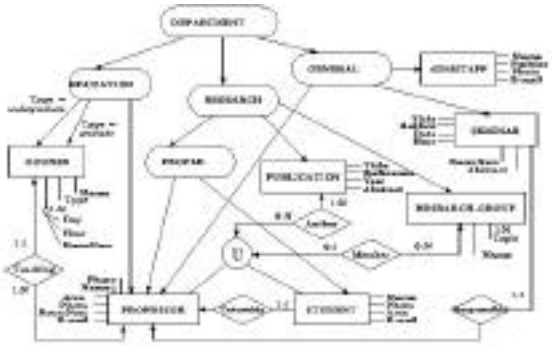
Shape the hypertext access structure



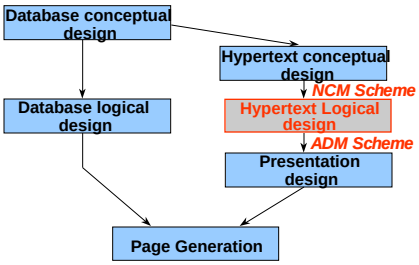
The Input ER Schema



The Output NCM Schema

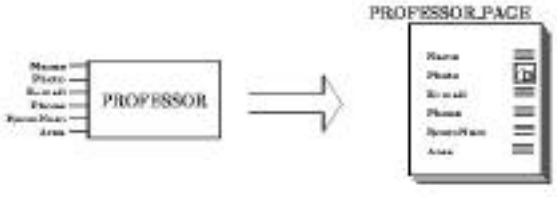


Hypertext Logical Design: From an NCM Schema to an ADM Schema

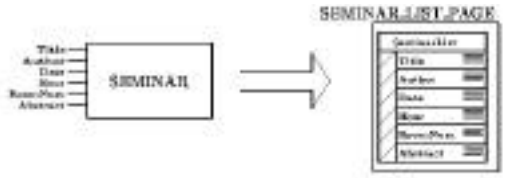


Hypertext Logical Design: Step 1

Map each macroentity into either a page schema or a list element inside a page schema



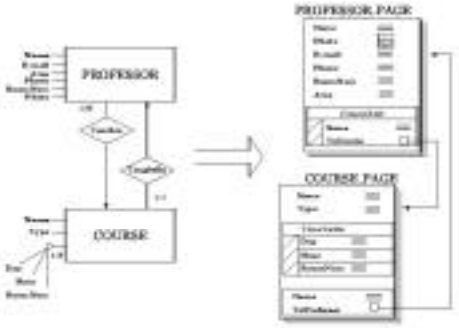
Another Example for Step 1



© 2001 Paolo Atzeni

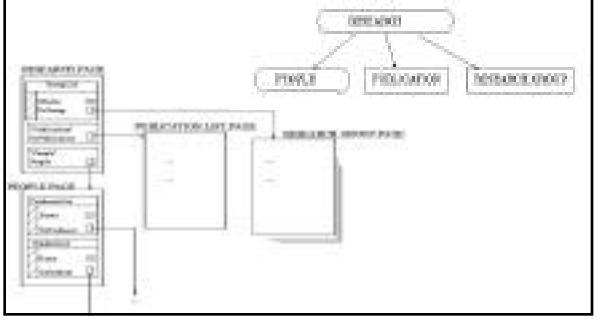
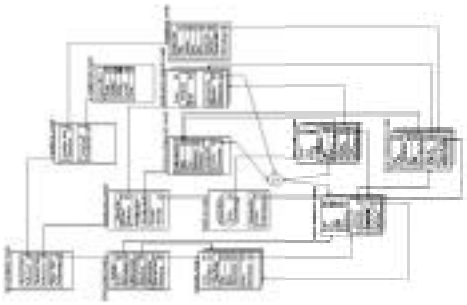
Hypertext Logical Design: Step 2

Map each directed relationship into a (list of) link attribute(s)



Hypertext Logical Design: Step 3

Map each aggregation into a unique page schema with link attributes to the target page schemas



Resulting ADM Schema ...Sideways...