

## Sistemi Informativi e WWW

### Premesse

- WWW: introduce un nuovo paradigma di diffusione (per i fornitori) e acquisizione (per gli utilizzatori) delle informazioni, con facilità d'uso, flessibilità ed economicità indiscutibili.
  - WWW: nato come interfaccia per l'accesso a informazioni distribuite su più siti ma specificatamente predisposte (pagine HTML)
  - WWW: tende a diventare una piattaforma generale per sistemi informativi di ogni genere.
- ⇒ **Sistemi informativi su WEB (WIS: web information systems)**

### WIS vs IS tradizionali affinità

- Problematiche comuni:
  - capacità di supportare attività di vario genere
  - e di gestire i relativi dati
- Gestione dati gestita da DBMS
  - molti dati anche in Web
  - gestione sempre più affidata a DBMS anche nei siti web.

### WIS vs. IS tradizionali specificità

- Web nato per la gestione di informazioni testuali e multimediali
- IS tradizionali gestiscono informazioni rappresentabili per mezzo di dati con struttura (relativamente) semplice
- Storicamente le informazioni sono memorizzate nei due sistemi in modo diverso
- Pur rese uniformi dai protocolli e dalla sostanziale standardizzazione dei browser è diverso l'interfaccia per una internet ad alte velocità ed un collegamento telefonico mobile

### Sistemi Informativi su Web

- Ripensamento del processo di sviluppo del sistema informativo con riferimento sia alle componenti esplicitamente sviluppate sia a quelle da integrare per far fronte alla maggiore complessità e diversificazione di applicazioni che il web permette e propone

### Sistemi informativi su WEB

- Motivazioni:
  - interazioni di componenti sviluppati in modo indipendente
  - cooperazione di compiti prima attuati separatamente
  - cooperazione fra compagnie diverse
- requisiti richiesti principalmente:
  - autonomia
  - eterogeneità
  - distribuzione dell'informazione
- Ripensamento del processo di sviluppo del sistema informativo con riferimento sia alle componenti esplicitamente sviluppate sia a quelle da integrare per far fronte alla maggiore complessità e diversificazione di applicazioni che il web permette e propone

Pubblicazione e consultazione su WEB (siti catalogo)

### **Comunicazione unilaterale**

- Richiedono un approccio sistematico sempre più importante al crescere delle dimensioni
  - pagine html: presenta i dati, l'organizzazione ipertestuale, la presentazione grafica
  - i dati possono cambiare, come si evolve il sito?
  - I siti presentano ridondanza dell'informazione per leggibilità e navigabilità
  - Spesso le basi di dati per queste informazioni esistono già

## Uso di DBMS

- Un organizzazione del sito che **affidi ad una base di dati** la gestione non ridondante delle informazioni e che permetta di costruire in modo automatico le pagine, permette sia di separare i vari aspetti (grafica e dati, p.e.) , sia di eliminare i problemi connessi con la ridondanza

⇒il browser diventa un'interfaccia standard per sistemi informativi

## Transazioni sul WEB

### **Comunicazione bilaterale**

- Interrogazioni e aggiornamenti della base di dati attraverso un interfaccia web
- ⇒applicazioni che permettano a tutti di eseguire operazioni che non siano di sola lettura
- nascono i problemi e le esigenze tipiche del settore delle basi di dati:
    - identificazione utenti e sicurezza
    - affidabilità e gestione delle transazioni

## Sistemi Informativi su Web: nuove applicazioni

- Commercio elettronico
  - se i cataloghi sono la principale applicazione dei siti di consultazione, è naturale pensare alla “vendita per corrispondenza”, sicurezza e affidabilità diventano essenziali
- gruppi di discussione: ad esempio, recensioni di libri
- sistemi di supporto al lavoro cooperativo: workflow management system (WFMS) (pe. Posta elettronica strumenti di produttività individuale)
- Sistemi “embedded”: telefonini, sistemi di navigazione, Internet TV

## Web come patrimonio informativo: prospettive

- “**L'intero Web è un sistema informativo**”
  - eccessivo: l'intero web non è certo sotto il controllo di un singolo soggetto (e neanche di più soggetti coordinati fra loro)
- “**Ogni pagina è diversa, potenzialmente autonoma, e la considero separatamente**”
  - estremo opposto: impraticabile se le pagine di interesse sono più di qualche unità e vanno mantenute aggiornate e coordinate
- “**Un sito è un sistema informativo e il Web è una federazione di siti fra loro autonomi**”
  - prospettiva equilibrata: ciascun sito è sotto la responsabilità di un soggetto (o gruppi di soggetti)

## Problemi con molti siti web

### **Nella loro struttura**

- L'informazione è spesso male organizzata ed è difficile trovare ciò che interessa
- Non sempre è chiaro quale e che tipo di informazione è disponibile
- La struttura di accesso è casuale
- Lo stile di presentazione è eterogeneo

## Problemi con molti siti web

### **Il mantenimento**

- Difficoltà nel mantenere aggiornato il contenuto
- Difficoltà nel tenere corretti i riferimenti ad altre pagine (sia tra le pagine di uno stesso sito che con il resto del mondo)
- Difficoltà nel cambiare la struttura definita all'inizio
- Difficoltà nel cambiare i dettagli delle presentazioni

## Lo sviluppo di siti web è attività complessa

- E' necessario un approccio metodologico
- analogo a quello seguito per sistemi informativi e basi di dati "tradizionali"
- così come le basi di dati tradizionali hanno una organizzazione che cerca di semplificare l'accesso riunendo i dati omogenei, può avere senso progettare porzioni di web con struttura controllata, con metodologie simili (p.e., l'università)

## Basi di dati e web: classificazione delle applicazioni

- Applicazioni che prevedono solo consultazione vs quelle che includono aggiornamenti e cooperazione (*siti che offrono informazioni vs siti che offrono servizi*)
- Applicazioni in cui il contesto operativo è controllato (solo i dipendenti) oppure a tutti gli utenti internet (*siti con interfacce più o meno standard e leggere*)
- Applicazioni in cui vi sia o no la presenza di più soggetti, distribuiti e autonomi, che cooperano fornendo informazioni o servizi (*distingue le iniziative che richiedono attività di integrazione*)

## Con riferimento alla basi di dati il terzo punto porta a distinguere

- siti che permettono di interrogare fonti informative diverse (autonome ed eterogenee, oltreché distribuite)
- data warehousing (data farming):
  - stante la grande disponibilità di informazioni sul Web, una azienda o ente può pensare di acquisire informazioni di interesse proprio attraverso il web.

Problematiche di estrazione dell'informazione, di integrazione e qualità dei dati

## Basi di dati e web: similitudine e differenze nel mettere a disposizione l'informazione

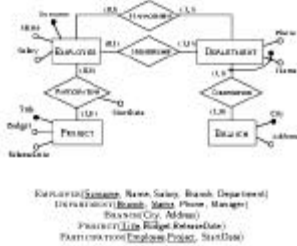
- La struttura degli ipertesti è semplice e irregolare, nati per correlare dei documenti,
- Basi di dati nate per organizzare informazioni in modo articolato e dettagliato permettendo l'utilizzo proficuo dei singoli elementi
- All'aumentare dell'importanza delle attività produttive su web si possono riscontrare delle regolarità delle pagine, che ne favorisce la navigazione da parte dell'utente che può familiarizzarsi con l'organizzazione del sito e cercare in modo analogo informazioni concettualmente diverse

## Progettazione di siti web "centrati sui dati" (data-intensive)

- Fa riferimento prevalentemente a siti che siano assimilabili alla "pubblicazione del contenuto di una base di dati" a fini di consultazione.

## Un modello logico

- Facciamo riferimento ad una base di dati con lo schema concettuale e logico della figura sottostante



## Sito per la consultazione di essa

- Home page:** con link verso le sedi e verso le pagine che contengono link ai progetti e agli impiegati
- una pagina per ogni sede:** con elenco dei dipartimenti, con evidenziato il direttore e i propri impiegati
- una pagina per progetto:** con lista degli impiegati che vi partecipano
- una pagina per impiegato:** con le informazioni personali ed i link verso la sede e i progetti cui partecipa

## Osservazioni

- All'interno di ogni categoria le pagine saranno molto simili
- Esistenza di due livelli simile ai modelli dei dati:
  - uno relativo alla struttura (lo schema della base di dati)
  - uno all'effettivo contenuto (l'istanza della base di dati conforme allo schema)

Si può parlare di uno schema di pagina (in inglese *page-scheme*)

... ha senso quindi definire un modello per ipertesti web, che descriva la struttura delle pagine e le modalità secondo cui queste pagine sono correlate

- Esempio:

```
page-scheme Employee
{
  EmployeeID: string;
  Name: string;
  Salary: integer;
  Dept: string;
  Branch: link(string: #Branch);
  Projects: list-of
    { Project: link(string: #Project);
      StartDate: date;
    };
}
```

N.b. i **link** sono costituiti da un attributo con il ruolo di ancora

## Livelli di rappresentazione negli ipertesti web: premesse

- Il livello logico descritto in precedenza descrive in modo compiuto la struttura ipertestuale, ma non descrive tutti i dettagli dell'ipertesto (l'effettivo lay-out, link di servizio)
- Le stesse informazioni potrebbero essere offerte attraverso un insieme di pagine diverse

## Livelli di rappresentazione negli ipertesti web

- Contenuto informativo:** in un sito centrato sui dati è costituito dalla base di dati. Si può parlare di schema e istanze
- Struttura ipertestuale:** costituita dalla organizzazione delle pagine e dalle modalità secondo cui è possibile navigare fra le pagine.
  - fa riferimento allo schema della base di dati
- Presentazione:** corrisponde agli aspetti grafici e alla disposizione del contenuto e dei link all'interno della pagina
  - fa riferimento allo schema dell'ipertesto

## Livelli di rappresentazione negli ipertesti web: osservazioni

- Esiste una forma di “indipendenza degli ipertesti” analoga all’indipendenza dei dati
- La struttura ipertestuale del sito (se i dati di interesse rimangono inalterati) può essere modificata senza alterare la struttura della base di dati
- Analogamente la presentazione può essere modificata, senza modificare né lo schema della base di dati che la struttura ipertestuale

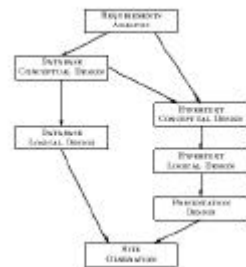
## Percorso per la progettazione

- E’ opportuno definire nell’ordine i tre aspetti dell’ipertesto:
- il contenuto
- la struttura ipertestuale
- la presentazione

## Progettazione: osservazioni

- Anche assumendo la progettazione circoscritta esclusivamente al sito web, è comunque necessario prevedere che l’analisi dei requisiti non riguardi solo gli aspetti specificatamente rivolti ai dati
- Sarà necessario ragionare anche sugli obiettivi del sito, sui potenziali utenti, sulle modalità di interazione desiderate

## Possibile metodologia per la progettazione di siti web centrati sui dati



## Osservazioni

- **Analisi dei requisiti:** progettazione concettuale e logica possono essere svolte come per le basi di dati tradizionali con particolare attenzione alle caratteristiche del Web
- **Progettazione concettuale e logica della base di dati:** definisce l’organizzazione del dominio in modo indipendente dalla realizzazione
  - Lo schema concettuale (ER) è il punto di partenza per la progettazione dell’ipertesto
- **La progettazione concettuale dell’ipertesto:** sulla base dei requisiti e dello schema concettuale dei dati, produce lo schema concettuale dell’ipertesto (nodi e percorsi di navigazione)
- **La progettazione logica dell’ipertesto:** definisce l’organizzazione dell’ipertesto in termine di pagine e link

## Osservazioni

- **La progettazione della presentazione** ha l’obiettivo di definire il layout di ciascuno degli schemi pagina definiti nella progettazione logica
- **Generazione del sito:** ha il compito di individuare una corrispondenza fra i dati nella base di dati e la struttura logica dell’ipertesto, oltretutto l’individuazione delle tecniche per l’effettiva realizzazione
- **L’approccio è modulare:** se la base di dati esiste già si percorre solo il ramo di sinistra.

### Attività principali per la progettazione concettuale di un ipertesto

- Selezione delle entità che corrispondono a nodi dell'ipertesto
  - ogni entità descrive un oggetto del mondo reale e può essere caratterizzata da attributi
- Selezione dei percorsi di navigazione tra entità
  - i percorsi possono essere orientati, perché la navigazione è caratterizzata da punto di partenza e di arrivo
- Definizione della struttura di accesso all'ipertesto
  - aggregazione delle entità in una gerarchia di concetti astratti da presentare all'utente per la navigazione
- Lo schema concettuale dell'ipertesto è derivato dallo schema ER
  - Esempio una entità 'puo' essere derivata da una entità 'ER se rappresenta oggetti rilevanti e indipendenti nell'ipertesto

### Progettazione logica dell'ipertesto

- Obiettivo: progettazione della struttura dell'ipertesto che sarà presentata dal browser
- Questa fase descrive *come* le pagine sono organizzate nel sito web
  - definizione delle informazioni rilevanti in una pagina (testo, immagine, ...)
  - definizione dell'organizzazione delle informazioni (pagina "piatta", lista, lista nidificata)
  - si ignorano gli aspetti di realizzazione fisica delle pagine
- Gli ipertesti sono grafi di pagine
  - occorre esplicitare la struttura logica a partire dal modello concettuale dell'ipertesto

### Progettazione della presentazione

- Un aspetto grafico gradevole è spesso la preoccupazione principale di chi vuole creare un sito Web
- Caratteristiche degli strumenti per la progettazione:
  - possibilità di prototipazione rapida per definire una struttura base per tutte le pagine del sito
  - possibilità di raffinare i dettagli di presentazione di singole pagine
  - possibilità di definire le caratteristiche delle pagine per mezzo di pagine HTML esempio (template)
- Elementi fondamentali
  - Stile dell'attributo: specifica come un attributo (della base di dati) deve essere visualizzato nella pagina.
  - Stile della pagina: descrive le scelte di presentazione per uno schema di pagina specifico.
  - Stile del sito: descrive le scelte di presentazione comuni a tutto il sito.

### Generazione del sito WEB

- Opera a partire da:
  - schema logico dell'ipertesto
  - schema logico della base di dati
  - stili di presentazione delle pagine
- Occorre definire la corrispondenza tra attributi dello schema ipertestuale e attributi dello schema della base di dati

### Generazione del sito WEB

- Informazioni per la generazione del sito
  - URL di partenza del sito ("radice")
  - Stile di presentazione del sito
  - Base di dati dove sono memorizzati i dati da pubblicare e modalità di connessione (normalmente JDBC/ODBC)
- Informazioni per la generazione delle pagine
  - Definizione della corrispondenza tra DB e pagine Web elenco delle tabelle (o viste) necessarie corrispondenza tra attributi di pagina e attributi dello schema della base di dati
  - Meccanismo di generazione di una URL locale per ogni pagina generata
    - valore costante (esempio: result.html)
    - valore parametrico, basato sul valore di un attributo della base di dati (esempio: URL(<Matricola>))

### Interconnessione fra basi di dati e siti web

- L'approccio standard è basato sul protocollo CGI
- Il protocollo CGI (Common Gateway Interface) permette di eseguire programmi tramite un server Web
- I programmi possono essere scritti in
  - qualsiasi linguaggio di alto livello (C, C++, ...)
  - linguaggi di scripting (Perl, Tcl, ...) o linguaggi di comandi (shell) del sistema operativo
- E' possibile specificare parametri di esecuzione per il programma

## Interconnessione fra basi di dati e siti web

### Funzionamento

- A. Il server riceve una richiesta, riconosce che la risorsa indicata dalla URL è un programma con cui deve comunicare tramite CGI e lancia il programma
- B. Il programma CGI è eseguito, accede alla base di dati (in modo tradizionale) e restituisce i dati al server
  - A. Apre la connessione con la base di dati.
  - B. Esegue l'interrogazione (o modifica).
  - C. Memorizza il risultato in variabili locali (una riga) o leggerlo mediante cursore (N righe).
  - D. Genera la pagina Web contenente il risultato e scriverla sullo standard output.
- C. Il server inoltra i dati al browser

## Interconnessione fra basi di dati e siti web

### Limiti del protocollo CGI

- Il protocollo HTTP è senza memoria
  - è difficile gestire applicazioni che richiedono operazioni multiple sulla base di dati
- Il programma CGI è un processo, per cui devono essere allocate risorse di calcolo
  - se occorrono risorse notevoli, l'allocazione può essere lenta
  - richieste diverse allo stesso programma causano allocazioni diverse
- Ogni esecuzione del programma CGI, richiede una nuova connessione con il DBMS
  - sovraccarico del DBMS
- I file HTML sono generati dinamicamente
  - è difficile realizzare un servizio di ricerca di informazioni sul sito, perché i dati devono essere estratti dalla base di dati

## Interconnessione fra basi di dati e siti web

### Alternative al CGI

- Soluzioni basate sul server
  - il client è un browser standard
  - il server gestisce l'interazione con la base di dati
- Soluzioni basate sul client
  - estendono il client, per incorporare alcune delle funzionalità del server

## Interconnessione fra basi di dati e siti web

### Soluzioni basate sul server

- Materializzazione delle pagine HTML (approccio *push*)
  - le pagine HTML del sito sono generate periodicamente a partire dalla base di dati
  - adeguata se i dati hanno una dinamicità limitata
- Vantaggi
  - eliminazione dei problemi di prestazioni
  - portabilità del sito in ambienti diversi
- Svantaggi
  - obsolescenza dei dati del sito

## Interconnessione fra basi di dati e siti web

### Soluzioni basate sul server

- Estensione delle funzionalità del server HTTP
  - server API (application programming interface), che permettono al server di interrogare direttamente la base di dati
  - il server HTTP diviene il client del DBMS
- Vantaggi
  - riduzione delle operazioni di inizializzazione e miglioramento delle prestazioni
- Svantaggi
  - mancanza di standardizzazione e scarsa portabilità

## Interconnessione fra basi di dati e siti web

### Soluzioni basate sul server

- CGI leggere (*thin CGI*)
  - programmi CGI molto semplici, che si limitano a passare la richiesta ad un server CGI, processo sempre attivo che si interfaccia con il DBMS
- Vantaggi
  - i processi thin CGI sono piccoli, il server è sempre attivo
  - miglioramento delle prestazioni

## Interconnessione fra basi di dati e siti web

### Soluzioni basate sul client

- Eliminazione della mediazione del server nell'interazione tra client e DBMS
  - estensioni del browser: moduli software eseguiti dal browser come librerie dinamiche (linguaggi Java, Active-X, ...)
  - helper esterni: programmi esterni attivati dal browser, che cede loro il controllo
  - browser proprietari: browser specializzati nell'accesso alle basi di dati
- Soluzioni adatte a contesti noti e predefiniti, quali le intranet