INTRODUZIONE AL WEB DINAMICO

Internet: la rete delle reti

Internet rappresenta oggi l'espressione del forte bisogno dell'uomo di comunicare, al fine, non solo di incrementare le conoscenze, ma anche di sviluppare i rapporti interpersonali che sono di fondamentale importanza nella formazione e nella crescita di ogni individuo.

La storia offre molteplici testimonianze di come l'evoluzione dell'uomo sia stata, ed è tutt'ora, fortemente condizionata dallo sviluppo di sistemi che permettano agli individui di capirsi e di comprendersi l'un l'altro. Di fondamentale importanza, in questo ambito, è la continua ricerca di nuove forme di espressione che denotano la forte esigenza di creare un linguaggio, di definire un insieme di regole per esprimere idee, concetti, per condividere informazioni.

In questo processo di evoluzione l'individuo prende coscienza dei vantaggi che la comunicazione apporta e si riorganizza in gruppi, con la convinzione che collaborare nella risoluzione dei problemi possa migliorare le sue condizioni di vita.

Si costituiscono le prime comunità in cui, come nelle attuali società, le leggi definiscono precise regole di comportamento che gli individui sono tenuti a rispettare; società diverse con lingue, usi, costumi e leggi differenti sono distribuite geograficamente in posti diversi.

Un tale modo di organizzarsi e di funzionare è stato costantemente modificato nel corso dei decenni, fino ad arrivare oggi alla formazione di società strutturate in modo tale da offrire all'individuo la possibilità di usufruire di servizi in qualsiasi settore dell'attività umana (trasporti, scuola, commercio, sanità, etc.).

Internet segue lo stesso modello di una società reale, e come quest'ultima, anch'essa può essere vista come costituita da diverse comunità sparse in tutto il mondo e collegate tra loro al fine di comunicare e di condividere risorse.

Società meglio conosciute come *reti di computer*, costituite non da individui ma da macchine, collegate tra loro non da strade o ponti ma tramite cavi, fibre ottiche ed etere, rispettando non leggi o lingue, ma precise regole imposte da un protocollo.

Internet non è una nuova rete di computer ma è un sistema di interconnessione di reti già esistenti.

Il suo enorme successo è dovuto non solo all'aver definito un linguaggio unico (TCP/IP) che permette di far comunicare tra loro, reti di computer con tipologie e piattaforme differenti, ma anche all'aver creato l'infrastruttura di una società virtuale in cui gli *utenti-clienti* possono accedere ai suoi servizi (posta elettronica, trasferimento file, web, etc.) tramite l'utilizzo di semplici interfacce, senza dover necessariamente conoscere i dettagli tecnici del suo reale funzionamento.

Internet nasce libera per essere libera. ARPANET era il suo nome quando, ai tempi della guerra fredda (metà degli anni sessanta del secolo scorso), il Dipartimento della Difesa Americano decise di realizzare una rete di comando e di controllo che potesse sopravvivere ad una eventuale guerra nucleare.

Le tradizionali reti telefoniche a commutazione di circuito erano considerate troppo vulnerabili, in quanto la distruzione della linea avrebbe interrotto tutte le comunicazioni. La soluzione migliore fu quella di implementare un modello di comunicazione *client/server* e di utilizzare la commutazione di pacchetto per il trasferimento delle informazioni. La commutazione di pacchetto costituiva una valida alternativa a quella di circuito in quanto, a differenza di quest'ultima, il percorso che l'informazione doveva compiere fino a destinazione non veniva stabilito a priori.

Nel 1974 con l'invenzione del modello e dei protocolli TCP/IP, progettati specificatamente per gestire le comunicazioni tra reti differenti, il numero di macchine collegate ad ARPANET crebbe rapidamente, ed il 1° Gennaio del 1983 TCP/IP divenne l'unico protocollo ufficiale.

La crescita continuò in maniera esponenziale e nel 1990 Internet collegava 3000 reti e 200000 calcolatori.

Ma la vera svolta si ebbe, tra il 1990 e il 1991, quando ai tradizionali servizi di posta elettronica e trasferimento file, Tim Berners-Lee, fisico del CERN di Ginevra, affiancò una nuova applicazione destinata a cambiare radicalmente le sorti di Internet: il WWW.

Il World Wide Web (ragnatela attorno al mondo) richiamò alla rete milioni di nuovi utenti attratti dalla possibilità di condividere pagine testuali di informazione distribuite in tutto il mondo.

Pochi anni dopo la prima implementazione del World Wide Web, negli USA la NCSA (*National Center for Supercomputing Applications*) sviluppò MOSAIC, un programma lato client che facilitava la consultazione dei documenti HTML; si trattava di un software non commerciale, disponibile gratuitamente. Mosaic fu realizzato in C e permetteva di vedere documenti con immagini incluse, trasferire in rete suoni o registrazioni video e collegarli a un documento tramite un link. Mosaic venne creato tra la fine del 1992 e l'inizio del 1993, da Marc Andreessen e dai suoi colleghi. Mosaic è in grado di interfacciarsi con vari servizi di rete come e-mail, telnet, ftp, news e Gopher (all'epoca, il sistema di ricerca documenti Gopher rappresentava la caratteristica più importante).

Nel corso del 1995, Mosaic venne praticamente abbandonato (l'ultima versione risale al 1997), per lasciare spazio a un nuovo programma browser: Netscape. Intanto il browser IE 1.0 (*Internet Explorer 1.0*) era stato acquisito da Spyglass, struttura commerciale del progetto NCSA Mosaic, e su di esso Microsoft iniziò lo sviluppo in competizione con Netscape; il rilascio avvenne nel Luglio del 1995 con Windows 95, come package a parte, mentre in Windows 95 venivano implementate le

funzioni TCP/IP. Nel frattempo Netscape lanciava la versione 2.0, introducendo la programmazione lato client con JavaScript, i frame e i plug-in.
Nel modello client/server, il client web ed il server, comunicano tra loro utilizzando un protocollo chiamato HTTP. Tramite l'HyperText Transfer Protocol (protocollo per il trasferimento di ipertesti) il client ed il server negoziano la richiesta di una risorsa informativa. Tipicamente questo genere di informazioni sono organizzate in pagine, realizzate utilizzando dei marcatori speciali (chiamati tag) che ne definiscono la struttura e la formattazione. L'insieme dei marcatori costituisce un linguaggio meglio conosciuto come HTML (HyperText Markup Language) adoperato per la creazione di ipertesti anche distribuiti.
Agli albori del web, tutti i contenuti informativi presenti in rete erano di tipo statico: le informazioni non cambiavano da una richiesta all'altra, o da un utente all'altro, fino a che il webmaster non aggiornava il sito sostituendo i documenti HTML che lo costituivano.
Attualmente il Web è molto più dinamico, non solo a livello client, dove Applet Java e contenuti Flash animano le pagine, ma soprattutto sul lato server, con la possibilità di generare del contenuto dinamico in base alle specifiche richieste degli utenti od in funzione al contenuto di un database.
L'idea base per ottenere contenuti dinamici fu quella di consentire all'utente di inviare al server la richiesta di esecuzione di un preciso programma, corredata da dati inseriti in un apposito modulo web. A seguito di tale interrogazione il server avrebbe dovuto lanciare l'esecuzione del programma, usando come parametri di input i dati inviati dall'utente. Tutti gli output generati, sarebbero stati a loro volta rispediti in tempo reale al client, che li avrebbe visualizzati all'interno di una pagina web.
Per realizzare tutto ciò non bisognava modificare niente della struttura client/server della rete, ma sarebbe bastato aggiornare: i client, i server, aggiungere nuove regole all'HTTP per la spedizione dei dati ed aggiornare l'HTML con i tag necessari per la creazione dei moduli (form).

Le Tecnologie del Web Dinamico

Per supportare siti web dinamici, sono disponibili un certo numero di tecnologie. In origine furono inventati i moduli CGI (Common Gateway Interface), programmi eseguibili dipendenti dalla piattaforma, in grado di generare contenuti testuali dinamici, tramite l'esecuzione di un programma specifico. I moduli CGI vengono chiamati dal web server all'arrivo di una richiesta web, eseguendo un nuovo processo esterno al server. I moduli CGI possono essere scritti in C, oppure in vari linguaggi di scripting, come Perl o Python. Ad ogni modo, questi componenti hanno grossi problemi di performance e scalabilità: il fatto di generare ad ogni richiesta un nuovo processo, comporta un intenso utilizzo delle risorse, soprattutto quando il server deve supportare un certo numero di chiamate concorrenti.

Per cercare di risolvere questi problemi, sono nate nel tempo soluzioni alternative, come le estensioni del server. Queste componenti, implementate come librerie dinamiche, girano all'interno dello spazio di indirizzamento del server Web, aspetto che riduce notevolmente le risorse necessarie all'elaborazione. In questo caso non è infatti necessario creare nuovi processi di elaborazione, ma possono essere direttamente i thread dei web server i deputati ad eseguire il codice di generazione presente nelle librerie di estensioni. Una delle prime società ad implementare questa tecnologia nei propri server web fu Netscape, che sviluppò NSAPI (*Netscape Server API*), interfacce di programmazione per la realizzazione di questo tipo di estensioni. In seguito anche Microsoft adottò questa strategia, quando decise di entrare anche lei nel mondo di Internet. Il suo web server, IIS (*Internet Information Server*), dispone infatti delle ISAPI (*Internet Server API*) alternativa, anch'essa proprietaria, a NSAPI.

Purtroppo anche le estensioni server presentano dei problemi, in quanto le diverse API disponibili non sono compatibili tra di loro e quindi obbligano a riscrivere parte dell'applicazione al cambiare della marca del web server. Questo aspetto costituisce di fatto una limitazione della flessibilità dell'applicazione che può aumentare i costi di manutenzione.

Altre tecnologie per la generazione dinamica di contenuti Web sono quelle che consentono l'implementazione delle pagine dinamiche: all'interno della pagina Web vengono inseriti particolari tag, elaborati sul lato server, che permettono di eseguire del codice sviluppato dal programmatore.

In quest'ambito è molto affermata la tecnologia PHP (*Acronimo ricorsivo di PHP Hypertext Preprocessor*), che si basa su un linguaggio di scripting lato server simile al Perl e dotata di un'ampia libreria di base con le più disparate funzionalità. PHP è il successore di un vecchio prodotto chiamato PHP/FI (*Personal Home Page/Form Interpreter*), creato da Rasmus Lerdorf nel 1995 ed inizialmente costituito da un insieme di script Perl, la cui funzione era quella di monitorare gli accessi alla sua home page. Lerdorf chiamò questo set di script "Personal Home Page Tools". Successivamente, quando servì implementare più funzionalità, Lerdorf ne scrisse in C una versione ampliata, dotata degli strumenti necessari per interagire con database e per sviluppare applicazioni web dinamiche. Variabili in stile Perl, interpretazione automatica dei parametri di un form e la possibilità di incorporare codice HTML all'interno di uno script sono i principali fattori che hanno determinato il successo del PHP.

Anche sul versante Microsoft, sono presenti tecnologie per le pagine dinamiche. Inizialmente le pagine ASP, documenti HTML in cui è possibile inserire codice VBA (*Visual Basic for Applications*) e compatibili solo con il server web IIS ed in seguito ASP.NET, la nuova versione implementata e migliorata delle pagine ASP.

Le tecnologie della piattaforma Java per la creazione di siti web dinamici sono le Servlet e le JSP (*Java Server Pages*).

Le Servlet sono estensioni del web server scritte in Java, mentre le JSP sono la tecnologia per creazione di pagine web dinamiche. I punti di forza di queste tecnologie sono le prestazioni, la portabilità, la robustezza, la scalabilità, l'integrazione con i dati, la libreria di base, la gestione delle sessioni e la gestione delle risorse condivise.

La portabilità è un elemento fondamentale che contraddistingue la piattaforma Java: un'applicazione web sviluppata in Java è indipendente dal web server e dal sistema operativo e quindi fornisce una flessibilità maggiore. Non è un problema sviluppare su Windows o Linux e poi passare l'applicazione su server Unix.

Un aspetto importante delle Servlet è la robustezza: il codice Java gira all'interno della Virtual Machine che è un elemento isolante rispetto al sistema operativo. Eventuali errori di programmazione non possono creare malfunzionamenti tali da danneggiare il corretto funzionamento del web server.

La tecnologia delle Servlet è molto scalabile: è possibile cominciare lo sviluppo con un semplice PC ed installare poi il risultato del proprio lavoro su server Unix ben più potenti.

Basandosi sulla piattaforma Java, le servlet possono accedere ad un ampio ventaglio di API per l'accesso ai dati. Le principali sono JDBC (*Java DataBase Connectivity*), che consente l'accesso ai database relazionali, JAXP (*Java API for XML Processing*) che permette di trattare con documenti XML.

HTTP è per sua natura un protocollo *stateless*: per il server, ciascuna richiesta è svincolata dalla precedente. Con il solo HTTP non si è quindi in grado di determinare se più richieste sono legate ad uno stesso utente che sta, ad esempio, lavorando con un sito di gestione della posta elettronica. Questo rende difficoltosa la realizzazione di applicazioni web, in quanto il programma deve gestire questo aspetto. Le Servlet offrono una gestione integrata delle sessioni utente, in grado di indicare se una richiesta è stata inoltrata da un utente già attivo.

Per ottimizzare le performance, le applicazioni basate su Servlet possono condividere informazioni (come tabelle caricate dal database) tra tutti gli utenti connessi.

Tutti questi vantaggi fanno della tecnologia delle Servlet uno dei maggiori successi della piattaforma Java. L'unica tecnologia che può competere con le Servlet a parità di funzionalità è ASP.NET che però è più giovane rispetto alla più matura e consolidata tecnologia di SUN e che ha il problema di tutte le tecnologie Microsoft: è fortemente vincolata a Windows.

Se le Servlet sono le estensioni del server della piattaforma Java, le Java Server Pages sono la tecnologia che consente di implementare pagine dinamiche su questa piattaforma. Il vantaggio di

JSP è che queste vengono compilate in Servlet, di fatto usufruendo di tutte le caratteristiche di quest'ultime.
Il componente che si occupa di questa operazione, e di gestire il funzionamento delle Servlet, è il Web Container (implementato da prodotti come Tomcat, Resin, IBM WebSphere).
Le Servlet sono più indicate dove il contenuto da ritornare non siano semplici pagine HTML, ma contenuti più complessi, come immagini generate dinamicamente, documenti XML o quando sia necessario operare programmaticamente sulle informazioni legate all'intestazione HTTP di richiesta. Le pagine JSP, invece, sono più indicate per presentare informazioni HTML.