

SECONDA PARTE DEL CORSO – lezioni da 11 a 22 (test il 2 dicembre)

Fondamenti di Informatica Lez. 11 lun 2 novembre 2015

Revisione argomenti della prima parte del corso.

Capitolo 1 - 4 del libro di testo, Slide del corso 1 - 10.

La lezione è stata dedicata alla preparazione del test del 4 novembre

Fondamenti di Informatica Lez. 12 mar 3 novembre 2015

Internet, protocolli di comunicazione, indirizzamento delle risorse in rete.

Tutto quello che abbiamo visto fino ad oggi riguarda un solo computer. Cosa accade se vogliamo mandare dei dati ad un altro computer?

La situazione cambia perché devo far quadrare due computer.

slide 11.1 internet

Problemi. Uno dei primi approcci fu quello di gestire più calcolatori...

slide 11.2 Mainframe vs Internet... immagine con un computer centrale

è una rete tra le più semplici, c'è un computer più potente al quale si collegano i computer più piccoli...che devono dare al computer principale dati e istruzioni, il quale che poi svolgerà i compiti.

slide 11.3 Mainframe vs Internet... immagine di rete di computer paritetici

Il salto grosso è togliere il calcolatore centrale ed avere in relazione tra computer simili: ciascuno lavora per conto suo e scambia i risultati con gli altri. Ma come faccio a sapere con quale, tra i tanti computer, scambiare ed instradare le informazioni?

La rete internet nasce fine anni sessanta in U.S.A.

slide 11.4 Arpanet

3 località indicate sono in California...Arpanet precursore rete internet

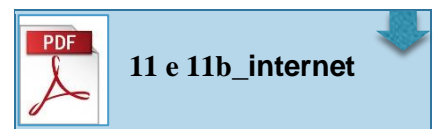
slide 11.5 Arpanet 1971...mappa U.S.A....collegamento costa ovest con costa est

slide 11.6 Internet Backbone

la dorsale internet, un insieme di cavi fisici che vanno da continente a continente.

slide 11.7 Dimensione delle linee internet

LAN sono reti di calcolatori che hanno un diametro massimo tra 10 metri e 1 km...casa...uffici...università...
MAN dimensione metropolitana, WAN ampia, GAN globale...



slide 11.8 Protocolli di comunicazione

bisogna dare delle regole di ore per governare ed indicare le attività di scambio.

Ci sono due standard:

ISO/OSI = standard che descrive le regole di comunicazione

TCP/IP = standard de facto più importante di tutti. Protocollo scritto così bene che ha sorpassato lo standard ufficiale.

slide 11.9 Commutazione di circuito

come fanno a parlare i due calcolatori collegati alle reti?

Posso fare la commutazione di circuito (che non si fa più): cerco la strada libera, più breve che posso, a quel punto la blocco per me.

slide 11.10 Commutazione di pacchetto

L'altra scelta si chiama commutazione di pacchetto: faccio il messaggio in pezzettini e lo mando nella strada che è libera, senza però bloccarla. Non si occupa mai stabilmente la strada, si cerca solo quella libera al momento: è necessario individuare ogni pezzettino di info inviata per consentire di riunire tutto alla fine e comprendere il messaggio.

per fare tutto questo bisogna fare delle regole, un protocollo.

slide 11.11 Parlare con i protocolli...immagine

è un modo per far passare le info anche se ai livelli più alti non si conosce un'altra lingua.

E' una metafora che riporta l'idea dei protocolli di internet

slide 11.12 Standard ISO/OSI

questo standard nel '78 era dettagliato ma...

slide 11.13 Protocollo TCP/IP

lo standard de facto è stato molto più semplice.

Il primo livello sopra è l'applicazione, cioè il programma che stiamo utilizzando. Il programma spezza il contenuto in pacchetti, poi internet li trasferisce.

slide 11.14 Modello di riferimento TCP/IP... dati...due colonne di rettangoli gialli...

slide 11.15 immagini che rappresentano il processo di pacchettamento ed invio dei dati

slide 11.16 immagini che rappresentano il processo di pacchettamento ed invio dei dati

si tratta della scomposizione in pacchetti che vengono spediti e ricostruiti.

Se ci sono pacchetti che mancano, parte un cronometro che dice quanto si aspetta per ricevere o segnalare la mancanza di un pacchetto per completare l'informazione.

Il timeout, che appare quando non si scarica una pagina internet, succede proprio quando non si è ricostruito tutto il messaggio e il tempo è scaduto

slide 11.17 immagine grande riquadro giallo...TELNET DNS SMTP...

http è un protocollo, cioè un insieme di regole, utilizzate da un certo gruppo di programmi.

ftp si usa ...

slide 11.18 Alcuni protocolli applicativi...

ftp è diverso leggermente dall'http...quando si vuole trasferire un file dal mio ad un altro computer...
smtp per inviare un messaggio
pop per ricevere un messaggio

slide 11.19 Risorse informative in Internet...immagine di calcolatori, con informazioni o di utenti

ci possono essere diverse richieste (posta elettronica, biblioteca digitale...ecc.) e stanno accadendo tutti i passaggi visti finora (pacchetamento, ecc.)

ma come faccio a scoprire dove sta chi ha l'informazione?

slide 11.20 e 11b.1 indirizzamento delle risorse in rete

problema: quello che si fa è cercare di nascondere alcuni problemi...ad esempio il fatto che esistano dei cronometri, la divisione dell'informazione, o quando copiamo un file dal l'hard disk alla chiavetta...

Problemi suddivisi a strati e ciascuno risolve i propri.

Dare l'indirizzo replica il sistema del file system, indirizzi facili da ricordare e ricostruire. Vedi indirizzamento gerarchico.

.com è il livello gerarchico superiore
all'interno c'è, ad esempio...Google
all'interno ancora c'è l'account...

si legge da destra a sinistra e all'inizio c'è https, vuol dire che mi sono immesso in questa pagina con la versione criptata di http.

ogni livello ha un proprio indirizzamento.

slide 11b.2 Indirizzamento nei vari protocolli (immagine scheda gialla)

partiamo dal più basso, l'indirizzo MAC

slide indirizzo MAC...IP e di porta...identifica il computer

indirizzo IP...rappresenta l'indirizzo di chi naviga in internet...

gli indirizzi IP vengono assegnati da istituti internazionali perché si cerca di suddividere i numeri in maniera razionale per area geografica.

ultima cosa è l'indirizzo di porta, cioè la strada sulla quale si vuole parlare.

Se ci sono due computer che vogliono parlarsi, devono trovare la corsia del cavo sulla quale parlare (è un numero da 16 bit che per alcuni protocolli è fissato) ad esempio l'http parla sempre sulla corsia 80, ecc. altri corrono su corsie libere.

E' importante perché per la sicurezza, li hacker si inseriscono sulle corsie non attribuite altrimenti i computer se ne accorgono...

Fondamenti di Informatica Lez. 13 mer 4 novembre 2015

Prima prova in itinere.

Internet, Indirizzamento delle risorse in rete.

Indirizzo MAC, indirizzo IP, indirizzo di porta, indirizzo DNS.

World Wide Web.

Breve riassunto di quanto trattato nell'ultima lezione.

slide internet...mainframe

il problema è, se abbiamo tanti calcolatori e vogliamo farli parlare tra loro: così fine anni sessanta nasce Arpanet, poi internet...la rete è una rete fisica di cavi, o collegamenti wireless satellitari...

slide acronimi...GAN...WAN...MAN...LAN

slide protocolli di comunicazione...

slide...protocollo TCP/IP

problemi a vari livelli...4

slide modello di riferimento TCP/IP

slide livelli...Host A Host B

ad ogni livello corrispondono varie "lingue"...i due protocolli più importanti sono quelli che riguardano internet...

siamo arrivati al punto chiave! Indirizzamento delle risorse in rete.

slide risorse informative in Internet...immagine

come faccio a dare un nome ad ogni computer che partecipano alla rete e dire a loro il nome di tutti quelli che partecipano alla rete?

slide indirizzamento delle risorse in rete.

riprendiamo con slide non ancora viste prima:

slide 11b.3 Indirizzo MAC, IP, e di porta

Abbiamo iniziato a vedere quali sono questi indirizzi.

Ad esempio questo computer ha MAC Address...60:.....

Ogni scheda di rete ha un codice unico, all'interno di ogni un numero alle proprie schede. A livello di ente locale, quando ci si connette bisogna dire il proprio indirizzo.

Poi a livello più alto c'è l'indirizzo IP a 32 bit

slide 11b.4 Indirizzo IP...esempio

slide 11b.5 Indirizzo IP...spiegazione

una possibile domanda sul l'indirizzo IP potrebbe essere da quanti bit o byte (32 bit o 4 byte)

slide 11b.6 Indirizzo di porta...

ogni calcolatore ha un IP, ma ogni calcolatore può parlare anche su vie diverse, gli indirizzi di porta, che sono fissati: ad esempio per il protocollo FTP la strada è la 21. è solo quella.

Funziona perché dall'altra parte il server sta ascoltando quella strada...

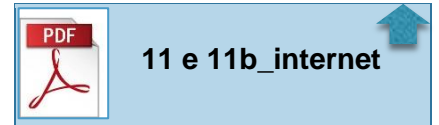
Sono un modo per dire quale strada deve impiegare il computer.

slide 11b.7 Domain Name System (DNS)

rappresentazione dei nomi, visto che tutti hanno un numero identificativo dato da IT, basterebbe quello per comunicare...ma come ce se lo ricorda il numero? Magari tra cento indirizzi?
Chi fa la traduzione da numero a nome? (nr..... = Unipd...o Google...ecc ecc).

II DNS!

DNS trasformazione degli indirizzi IP in sequenze comprensibili!!!!



slide 11b.8 Livelli di dominio

di primo, secondo, terzo livello

Come funziona la scoperta del nome relativo all'indirizzo IT : è un percorso a salire. Se ci connettiamo via la nostra adsl il nostro computer chiede al server se conosce la traduzione a parole dell'indirizzo IP...se non la trova chiede ad un livello più alto...

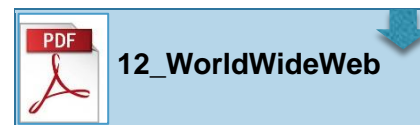
se chiede a alice.it...facebook.com...DNS chiede ad un livello superiore finché lo trova.

Si può avere un indirizzo IP variabile o fisso. Serve fisso se vogliamo dare un servizio di tipo server web o server ftp, cioè nei casi in cui dobbiamo fornire un indirizzo costante.

Le tecnologie del Web: HTML, URL, HTTP, MIME

Le reti Peer-to-peer (P2P)

slide 12.1 WorldWideWeb
slide 12.2 WWW e Tim Berners Lee



Internet è una rete di calcolatori, fisicamente rete di interconnessione dei calcolatori e su questa rete comunico perché esistono protocolli e indirizzi.

il web è una collezione di informazioni collegate tra loro e mantenute su calcolatori diversi presenti sulla rete internet.

internet rete fisica, web rete di informazioni.

Web nasce fine anni ottanta con Tim Berners Lee. C'è una differenza tra internet e web, anche se comunemente si frainendono. Web è una rete di documenti salvati da qualche parte, da un insieme di calcolatori presenti e collegati da internet.

slide 12.3 Tim BL, Answers for Young People
slide 12.4 Il Web

prima del web per trovare info diverse su computer diversi bisognava andare sul computer che le possedeva, autenticarsi, prendere le info.
Invece tramite la rete internet prendo le informazioni senza spostarmi.

Le informazioni che creano le reti web, che sono collezioni di file collegati tra loro, sono gestite in calcolatori che si chiamano **SERVER** (aziende private come Facebook, organizzazioni pubbliche, ma anche dai singoli privati che possono mettere in rete le proprie informazioni).

slide 12.5 Modello Client-Server...immagine

come funziona il web? al livello più alto il linguaggio è HTTP. Il client fa richiesta dell'informazione, il server la riceve e fornisce la risposta.

slide 12.6 Modello Client-Server
slide 12.7 Modello Client-Server
slide 12.8 Il Web

gestione dati: è il server, che deve gestire i dati. Dalla parte del client, utente che riceve le info, non ha l'obbligo di qualificarsi, ma solo di presentare i dati una volta pervenuti.

E' importante sottolineare questo livello di comunicazione: c'è un computer più importante degli altri, il server che ha le informazioni, e poi c'è l'utente, il client, che chiede al server delle pagine: non succede il contrario!

Il processo del server è un programma in esecuzione che attende dalla strada 80 le richieste dei client.

slide 12.9 Terminologia

il sito web ha alcune caratteristiche...la pagina iniziale, poi le altre pagine sono strutturate con testi e immagini o video...poi ci sono le zone cliccabile, ancore o bottoni, cioè punti che permettono di arrivare ad un'altra pagina, il modo per arrivare è chiamato link.

- sito web...è una collezione di info, una serie di pagine. viste come insieme info unitarie...
- la home page... e quella che dovrebbe darmi accesso a tutto il sito web
- data una pagina web, abbiamo un bottone, ancora, che è cliccabile, che dovrebbe portarci su un altro punto del sito, collegamento iper-testuale.
- link...

slide 12.10 L'ipertesto

quando navighiamo non facciamo altro che copiare sull'hard disk le pagine cercate.

slide 12.11 Localizzazione delle pagine

come faccio a localizzare una pagina web, a conoscerne un indirizzo?

E se non abbiamo le due opzioni nella slide?

Magari pubblico una pagina sotto il nome....mio sito/ecc...

Ma non conosco indirizzo o link

Allora col motore di ricerca? ma se non conosce il link il motore di ricerca non lo trova!!!!

C'è un enorme mondo web invisibile!

Come faccio a scoprire dove è una pagina web?

2 modi,

- sapere l'indirizzo esatto di una pagina
- conoscere una pagina vicina e via link successivi si trova

slide 12.12 Tecnologie del web.

Quattro tecnologie importanti nel web: HTML URL HTTP MIME

l'ultima MIME riguarda se voglio inviare una immagine. come faccio ad inviarla come informazione? c'è un modo per trasformarla in una sequenza di bit come fosse del testo.

slide 12.13 Hypertext Mark-up, Language (= HTML)

come scrivo le informazioni? Ci sono vari modi, ma di base si usa il linguaggio HTML, che non è di programmazione ma **linguaggio di presentazione, cioè un modo per dire al browser come faccio ad evidenziare i contenuti.**

slide 12.14 HTML

la pagina HTML se viene aperta come file di testo, la prima cosa che si vede sono le istruzioni di inizio e fine della pagina web. tutte le pagine web devono essere esplicitate così.

on line ci sono già i programmi per impostare un sito.

slide 12.15 Uniform Resource Locator (=URL)

è il modo di associare un nome ad una pagina: protocollo...http, macchina...www.unipd.it...un file che è /Index.html

slide 12.16 URL

protocollo, dominio, file

.it e il dominio di primo livello...Unipd è dominio di secondo livello...

la home page nel 80 90 % dei casi ha nome Index.html

la terza cosa è come parlo per raggiungere l'informazione, o fare richiesta/risposta

slide 12.17 HTTP

slide 12.18 http (protocollo)

la slide riporta l'elenco della decina di cose che può chiedere il client e che può rispondere il server.

Tutte queste cose le fa il browser.

Tutto questo deve essere fatto in un certo tempo (secondi) altrimenti ci sono degli errori.

slide 12.19 Multipurpose Internet Mail Extensions (= MIME)

mail...non è un caso perché questa parte gestisce info non testuali.

Ma: come faccio a inviare dei caratteri che non fanno parte dell'ASCII?

Audio, video, pdf...sono sequenze di bit che non c'entrano nulla coi testi

E nata la possibilità di inviare info non testuali con questo'approccio:

all'inizio ogni messaggio precisa che da adesso i avanti ti sto inviando non più testo ma una codifica diversa, come jpeg...PDF...file zip...

slide 12.20 MIME...la lista dei formati supportati...

slide 12.21 MIME...è un sistema progettato per supportare...

sequenza di bit che può esse tutto... prendo gruppi di 6 bit e metto due zero in testa, cioè sto facendo una mappatura dal nulla al ASCHII.

tutto quello inviato in allegato viene trasformata in una sequenza di caratteri che hanno 8 bit ma i primi due sono 0 0

slide 12.22 Il browser Web



è un programma che dato un indirizzo url apre tutte le comunicazioni con la macchina con cui vogliamo parlare, con il server, è una volta ottenuto il file da parte del server il browser lo legge.

Fondamenti di Informatica Lez. 16 mer 11 novembre 2015 LABORATORIO

Utilizzo del servizio OneDrive di Microsoft.

Dropbox e modifica dei file con OneDrive.

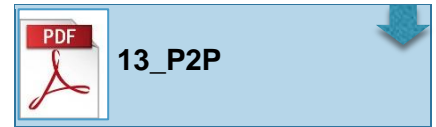
Utilizzo del servizio Weebly.com per la creazione di siti Web.

Reti peer-to-peer.

Reperimento dell'informazione, definizione del problema e definizione di utente esperto

Riprendiamo qualcosa della volta scorsa. quattro slide che concludevano le reti.
Abbiamo visto un modo per comunicare...client-server ma...

slide 13.1 Peer- to-peer
cioè pari a pari.



slide 13.2 Sistemi di comunicazione

si tratta del processo di comunicazione nei 5 elementi presenti: SORGENTE-TRASMETTITORE-CANALE DI TRASMISSIONE-RICEVITORE-DESTINATARIO

slide 13.3 Comunicazione tra due soggetti

immagine...sorgente → trasmettitore → canale di trasmissione → ricevitore → destinatario

slide 13.4 Comunicazione

slide 13.5 Canale di trasmissione

slide 13.6 Protocolli di comunicazione...Perché la comunicazione si realizzi...

slide 13.7 Protocolli di comunicazione...immagine riassuntiva...

non solo impedimento fisico (muro o distanza) ma anche mancanza di comprensione delle info trasmesse! I protocolli applicativi ci consentono sia di oltrepassare le barriere fisiche, sia far parlare "lingue" diverse (due livelli HTTP...poi DSP...poi...ai vari livelli medesimo linguaggio).

Ci sono più livelli perché ciascuno risolve problemi diversi...come inviare...quale strada fare...come trovare il computer destinatario...

C'è un parallelo col sistema operativo, nel senso che i protocolli risolvono i vari problemi

slide 13.8 Reti di calcolatori

come faccio a scoprire dove una risorsa, cioè info o computer? Primo approccio: server centrale. Secondo approccio: distribuisco tutta l'informazione su tanti calcolatori.

slide 13.9 Modello Client-Server...immagine

slide 13.10 Modello Client-Server...svantaggi e vantaggi

prima modalità: client-server, vantaggi di sapere dove sono le info. Il problema è il singolo punto d'ingresso, se capita qualcosa non riceviamo più informazioni. Questo è della famiglia dei modelli gerarchici.

slide 13.11 Modello peer-to-peer (P2P)

slide 13.12 Architettura P2P...immagine

slide 13.13 Modello P2P

seconda modalità: è l'opposto, basta unico calcolatore e info è su tanti calcolatori. Ma a chi chiedo? Vantaggio è che se si rompe un calcolatore ne ho tanti altri. Ma allora come faccio a sapere dove sono le info?

Deve iniziare un sistema di scambi di messaggi...

Distribuire le info su vari calcolatori, anche presenti non interamente su un unico calcolatore, vantaggio perché se voglio nascondere chi ha condiviso, posso...

slide 13.14 Vantaggi e svantaggi

lo svantaggio del peer to peer è quando mi manca un pezzettino di informazione. mando un messaggio a tutti? C'è una moltiplicazione delle richieste!

Come si risolve? Si nominano alcuni responsabili che osservano un gruppo più limitato di computer.

Questo riduce di molto. Ora il responsabile ha l'informazione...ma se esce dalla rete anche lui...si mette in moto un meccanismo di voto dove coloro che sono rimasti senza responsabile ne nominano un altro.

Questo succede con Skype che è una rete peer to peer e alcuni calcolatori vengono eletti responsabili della rete. Ma succedeva che il tuo calcolatore aveva il processore 100% su Skype, perché eri stato eletto responsabile!!!! E quindi impiegavano altri il mio processore!!!!

Ma l'utente non sa quando diventa responsabile!

slide 13.16 Funzioni

slide 13.17 Topologia

questione velocità: se in tanti chiedono ad un server web può essere facile un attacco ad un server web. Invece sulla rete P2P chiedo porzioni di informazione e la velocità è maggiore in quanto se connessi a tanti computer di buona velocità la somma dà un totale maggiore di un singolo server.

slide 13.17 Connessioni...un aspetto fondamentale...

slide 13.18 Storia swl P2P (File Sharing)

peer to peer: due elementi

connessioni strutturali= sono connesso

connessione diretta = parlo solo con chi mi interessa...stessa cosa che capita con Skype...ricorda molto client-server

slide 13.19 Napster

Napster funzionava in quanto c'erano più server ai quali un utente si poteva connettere per mettere a disposizione i suoi file. Così io li potevo far scaricare a chiunque: milioni di utenti in pochissimo tempo. Problema: chiuso il computer centrale...fine degli scambi...

slide 13.20 Gnutella...(GNU approccio opensource)

rete peer to peer pura, cioè davvero tutti i computer alla pari.

la rete è ancora in piedi...

slide 13.21 KaZaa

prima serie di miglioramenti peer to peer

raggruppa la rete in gruppi, è solo i capigruppo si parlano...

slide 13.22 BitTorrent

è lo standard, rete peer to peer più utilizzata al mondo, è un protocollo che parte da un file, file .torrent, che contiene alcune info, indica un punto di partenza, di ingresso, per iniziare a scaricare, da quanti pezzi è formato il file, quanto grandi sono...

il punto debole del Torrent è il punto d'ingresso...il file ha il punto d'ingresso che si trova sul server

slide 13.23 A livello legale?

slide 13.24 Pirate bay

è un motore di ricerca che dà l'accesso ai file d'inizio...

cercare spoty file. !!!!!

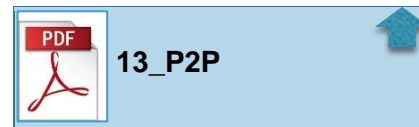
slide 13.25 VOIP

lo stesso principio di condivisione di file lo posso applicare a qualsiasi informazione: cioè telefonate via internet.

slide 13.26 Skype

se vale per l'audio ...

slide 13.27 IPTV



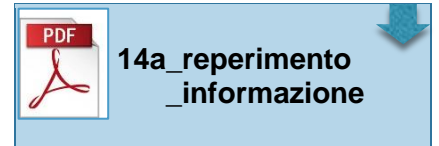
allora posso mandare anche il video...

ho un flusso video che voglio trasmetterlo in line condividendolo con tante persone...la peer to peer television. ... potrebbero esserci trasmissioni *no limits* in tutto il mondo!

vedere net Flix...server a bassissimo costo.

Abbiamo visto che esistono info su calcolatori...e le possiamo vedere se abbiamo l'indirizzo...se non abbiamo queste due cose c'è una terza informazione che può venire dai motori di ricerca.

i motori di ricerca.



slide 14a.1 Reperimento della informazione...quadro generale

il problema del reperimento è quello di pescare nella massa i documenti rilevanti: realizzazione difficile!

slide 14a.2 Reperimento dell'informazione...immagine

slide 14a.3 Reperimento dell'informazione...definizione

slide 14a.4 Disponibilità delle informazioni...problema della disponibilità delle info...

dove si trova la collezione dei documenti? In biblioteca è semplice...e dovrei sapere le info della collezione dei documenti...altrimenti devo cercare in un'altra biblioteca... ovviamente teoricamente si potrebbe leggere tutti i documenti. Allora si costruisce un catalogo preparato dal bibliotecario...

preparazione e la catalogazione o indicizzazione....cioè sintetizzare l'informazione...

slide 14a.5 Disponibilità delle informazioni...il problema di come reperire le info...

c'è la soluzione: se non esistessero i motori di ricerca dei siti web dovrei leggere tutto. Allora si costruisce la struttura dati sintetica, che si chiama indice, che è quella che si utilizza quando si fa la ricerca.

slide 14a.6 Indicizzazione delle informazioni

è l'automatizzazione...

ad esempio un libro ha l'indice...si costruisce un altro indice automaticamente! è più facile e rapido cercare nell'indice...

Bisogna definire il contenuto che interessa, cioè la base per costruire gli indici che costituiscono i libri. Il testo è fatto di parole...quindi analizzo le parole...

per i libri è facile...

difficile invece per formati non testuali...come si fa cercare l'immagine o il suono di un certo tipo? Come si fa ad automatizzare la procedura?

Parliamo solo dei testi...

costruiamo degli indici di struttura dati.

slide 14a.7 Utenti e reperimento...

importanza dell'utente!

slide 14a.8 Tipologie di utenti...

come ragiona il motore di ricerca...

Reperimento dell'informazione.

Fasi del reperimento: Acquisizione (o localizzazione), indicizzazione, recupero.

Localizzazione delle pagine Web da parte dei motori di ricerca.

Indicizzazione automatica.

https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles?language=it

Oggi vedremo un video sull'utilizzo dei motori di ricerca.

slide 14a.9 Reperimento dell'informazione...finalità...tipologie...modalità...l'utente

E' una slide di riassunto

ora vediamo i vari passi da fare per mettere insieme la ricerca automatica...

slide 14a.10 Fasi del reperimento...immagine

i tre passi sono: acquisizione, indicizzazione, recupero.

Nella fase acquisizione ci sono due alternative, o i doc sono in formato testuale o dobbiamo digitalizzare.

Fase indicizzazione: dal doc in formato digitale devo raccogliere tutti i documenti e tutte le parole (BD nella slide sta per basi di dati, cioè la struttura dati che permette di ricavare i dati)

Poi c'è il recupero, che confronta

questa fino ad ora è la parte del computer.

Ora inizia la parte dell'utente

slide 14a.11 Fasi del reperimento...immagine

indicizzazione e che il computer fa la stessa cosa sulla nostra micro richiesta.

Poi il sistema fa il confronto tra le due indicizzazioni...e ci offre dei documenti. Posso essere soddisfatto oppure leggendo affino l'idea e le richieste, o riformulo completamente la ricerca

slide 14a.12 Fasi reperimento...immagine

è la messa insieme delle due azioni motore di ricerca e utente sopra viste

slide 14a.13 Reperimento delle informazioni sul web

quanto visto finora funziona bene se la collezione di documenti è statica (biblioteca).

Invece nel web c'è un problema...non c'è una autorità che regola l'apertura di siti web.

E' difficilissimo controllare il contenuto di tutte le pagine web.

Questione attendibilità! Uno può scrivere quello che gli pare (mentre in biblioteca vi è l'inserimento di libri scelti appositamente) e quindi come posso valutare l'attendibilità del web?

Altro problema: nascono e muoiono una infinità di pagine al giorno.

Come si fa a trovare l'informazione che ci interessa????

slide 14a.14 Motori di ricerca

slide 14a.15 Motori di ricerca

delle tre fasi quella che cambia radicalmente è l'acquisizione, che cambia nome e diventa "localizzazione" cioè andare sul web e trovare le pagine.

Si tratta di programmi automatici o quasi.

Devo partire dall'indirizzo! Carico la pagina unipd.it, percorro tutti i link e salvo la pagina. Poi esploro quello che trovo nei link...così ho l'informazione e il MdR sa quello che c'è nei link.

Ma così ho esplorato solo un sito!!!

Quindi qualcuno deve fornire dati al Web Search Agent che è un programma che fa queste cose.

slide 14a.16 Localizzazione delle pagine Web

mette da parte i link delle varie pagine

slide 14a.17 Localizzazione pagine web... prima immagine

quello che è automatico e quello di percorre tutti i link...di pagina in pagina...

slide 14a.18 Localizzazione delle pagine web...seconda immagine

e se un MdR non passa mai sopra ad un insieme di siti? Non otterremo mai una risposta su quelle pagine. Ma questa parte è la stragrande maggioranza del Web!!! Forse dieci volte tanto il web scoperto!

slide 14a.19 Localizzazione... immagine

slide 14a.20 Localizzazione... immagine

slide 14a.21 Indicizzazione automatica dei documenti testuali

fino ad ora il MdR ha messo da parte le pagine. Ora deve analizzare il contenuto delle pagine web...

slide 14a.22 Analisi lessicale

è la prima fase.

slide 14a.23 Rimozione stop word

slide 14a.24 Estrazione delle radici

La seconda e terza hanno valore storico ma che oggi non ci sono quasi più!!!! Fino a quindici anni fa sì perché si risparmiava spazio (rimozione e trazione radici).

Altro modo di risparmiare spazio era ridurre le parole alle loro radici.

slide 14a.25 Creazione indice e pesatura

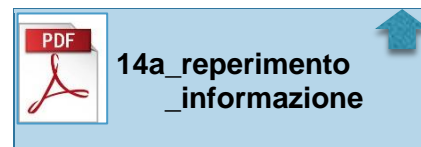
è la penultima fase: creazione dell'indice, ordino tutto quello che ho visto e ad ogni parola devo ricordarmi la pagina a cui appartiene la parole stessa. Si crea una parola e il MdR cerca la parola e crea una lista che dice quante volte è presente e in proporzione ordina e presenta i risultati. Ma il MdR cerca sinonimi e argomenti correlati...

slide 14a.26 Modelli di recupero

ultima fase: la pesatura...importanza dei termini

vedere il servizio internet archive che tiene in memoria le pagine esposte dai siti nel passato...

proiezione del video di Eli Pariser: attenti alle "gabbie di filtri" in rete.



Fondamenti di Informatica Lez. 19 mer 18 novembre 2015 LABORATORIO

Link per inserire le query (clicca qui)

https://docs.google.com/document/d/1RPgog_eVqEaFJPza8HYxW20615LE3c4oyiivvSH8Ucok/edit?pli=1

Fondamenti di Informatica Lez. 20 lun 23 novembre 2015

Motori di ricerca, importanza dei termini e dei link per ordinare le pagine.

Sicurezza: Password e Virus informatici

a laboratorio avevamo saltato qualcosa...cercare Googleworks

si va sul link indicato, sia con accedi che con inizia...

bisogna saltare la configurazione guidata...

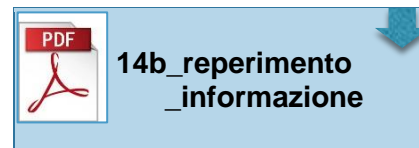
appare la pagina dove è possibile fare strumento di pianificazione delle parole chiave (dal menù strumenti)

.....

lo possiamo usare per capire quali possono essere termini chiave da mettere sul sito web!

.....

Riprendiamo le slide.



slide 14b.1 Modelli di recupero

la localizzazione l'abbiamo vista, l'indicizzazione pure, ora passiamo al recupero: in due modi.

1) modo esatto (exact match)...che contengono la parola

2) best macht: più utilizzato dai MdR, è il confronto migliore, non è detto che l'assenza dei termini escluda del tutto i documenti.

slide 14b.2 Modello di recupero booleano

i MdR attuali permettono di attuare questo metodo...ma bisogna andare in affina ricerca!

slide 14b.3 Modello booleano

non usata spesso dai MdR perché non riesce a controllare il volume di documenti da restituire, non c'è modo di riordinarli! Problema!!! Trovo migliaia di documenti, quali presento prima?

Ecco perché funziona bene per biblioteche ma non per il web

slide pesatura dei termini indice...

slide 14b.4 Pesatura dei termini indice

slide 14b.5 Pesatura tf x idf

importanza della frequenza del termine...per prima in ordine è messa la pagina che ha più frequenza!

ma anche: parola frequente ma molto comune...cosa si fa? cerco "appartamenti in vendita" e io cerco la parola "casa" è frequentissima, ovunque, ma non si riesce a discriminare!

questo problema è risolto da inverse document frequency...è una formula! questo, con la frequenza del termine, dà il peso del termine!!!!

per far salire il documenti bisogna descrivere bene il documento, non con parole troppo comuni!

slide 14b.6 Rilevanza di un documento

rispetto all'utente. cioè alcuni doc sono rilevanti per una persona e non per altri.

slide 14b.7 Valutazione

slide 14b.8 Misure della valutazione

slide 14b.9 dati rilevanti/dati non rilevanti

slide 14b.10 dati rilevanti/dati non rilevanti

aspetti di errore...effetto rumore = il MdR ci propone documenti interessanti ma assieme ad altri che ci rallentano il processo di cernita e valutazione. Effetto silenzio = il MdR nel dubbio non ci fa vedere certe cose... sono i due possibili casi di errore per un motore di ricerca: se siamo troppo generici entra in gioco l'effetto rumore.

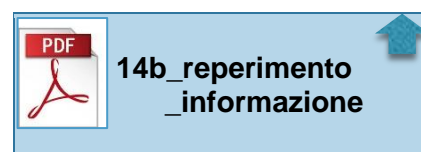
slide 14b.11 Funzioni avanzate

slide 14b.12 Altri approcci alla ricerca sul Web

slide 14b.13 Meta-informazioni nel Web

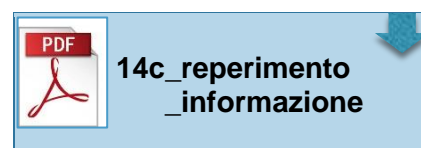
slide 14b.14 Biblioteche digitali

slide 14b.15 Accesso Web ad una biblioteca digitale



slide 14c.1 Come funziona Google (più o meno)

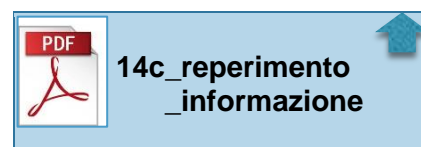
la cosa importante è quanti mi linkano...la pagina diventa più autorevole
il MdR media il fatto che un sito sia autorevole con le cose che ci sono dentro.



slide 14c.2 Una volta che Google ha "trovato"...

slide 14c.3 Google stima l'importanza...

slide 14c.4 L'utente inoltra una query a Google...



slide 15.1 Sicurezza...

c'è internet...c'è il web....modi per accedere e per essere monitorati!
Entra in gioco la sicurezza.
Passi utili per difendersi.

slide 15.2 Autenticazione

fondamentale e avere nome utente e password o almeno una password.

slide 15.3 Password

slide 15.4 Scoprire una pwd

slide 15.5 Autenticazione forte

slide 15.6 Virus informatici

qui ci vuole il secondo livello di sicurezza...

I virus sono programmi che hanno obiettivo distruggere info o rendere inaccessibile il dispositivo. Essendo un programma, il virus deve arrivare alla memoria centrale e ci arriva grazie alla memoria secondaria.

slide 15.7 Virus informatici... Nel 1949...

sono programmi, nascono in assenza della rete internet... questo è alla base dei virus

Fondamenti di Informatica Lez. 21 mar 24 novembre 2015

Sicurezza: Virus, Worm, Trojan horse.

Antivirus, Firewall, Backup.

Crittografia: cenni su chiave pubblica e privata.

vediamo il sito www.zoo.com

è un meta motore di ricerca...

Come funzionano? Inviando l'interrogazione a più motori di ricerca, li raccoglie e li presenta miscelati.

Torniamo alle slide relativi ai problemi di sicurezza...

slide 15.8 Antivirus...

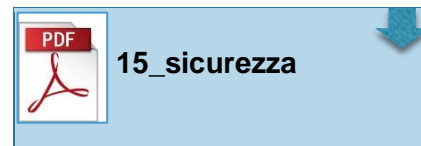
il virus è un programma ridotto per le operazioni che deve fare perché se fosse complesso l'utente se ne accorgerebbe, e anche il computer. Il virus mette poche istruzioni all'interno di un file tra quelli che si aprono con molta probabilità. Ad es. i .doc, che sono diffusissimi.

L'obiettivo del virus è replicare se stesso, e per evitare di replicarsi sullo stesso computer, ha una sequenza di bit che gli permettono di sapersi riconoscere. Se vede che è già presente non si riproduce.

Gli antivirus vanno a cercare proprio queste sequenze.

Bisogna caricare e aggiornare gli antivirus.

Quante risorse occupa un antivirus? è un programma che deve essere in esecuzione sempre! quindi il processore è sempre in attività...e può rallentare.



slide 15.9 Worm

altra famiglia di virus...i "vermi".

slide 15.10 Trojan Horse

slide 15.11 Firewall

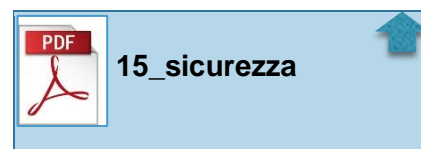
è un programma che controlla le connessioni in entrata e uscita. Il Prof. consiglia di installare anche il Firewall

slide 15.12 Back-up

sono delle copie.

slide 15.13 Back up... copie di back up

oggi le soluzioni migliori sono tipo Dropbox, insomma i servizi iCloud.



slide 16.1 Crittografia...la storia

è lo studio di come nascondere testi cifrando i messaggi.

slide 16.2 Steganografia

slide 16.3 Crittografia... non mira a nascondere

slide 16.4 Crittografia...obiettivi della crittografia

slide 16.5 Crittografia -trasposizione

trasforma, non nasconde, il messaggio. viene compreso solo da chi scrive e da chi legge.

come si può fare? in due modi, per anagrammi (trasposizione) o per sostituzione

slide 16.6 Crittografia- sostituzione

slide 16.7 Classificazione

slide 16.8 Terminologia

slide 16.9 Il modello di cifratura

slide 16.10 Assunzioni

slide 16.11 Comunicazioni radio

slide 16.12 Fine della Grande Guerra

slide 16.13 Sviluppo delle macchine per cifrare

slide 16.14 Enigma

slide 16.15 Crittografia moderna



-----fine delle lezioni-----

Fondamenti di Informatica Lez. 22 mer 25 novembre 2015 LABORATORIO

Aggiungere un URL a Googlebot (link).