

Parte 6

Altri servizi Internet Come funzionano e come usarli

Contenuti

- Architettura di Internet
- Principi di interconnessione e trasmissione
- World Wide Web
- Posta elettronica
- Motori di ricerca
- *Netiquette*
- Antivirus
- Personal firewall

**Tecnologie delle
reti di calcolatori**

**Servizi Internet
(*come funzionano
e come usarli*)**

**Servizi Internet
(*come difendersi*)**

Parte 6

Modulo 1: Posta elettronica



Posta elettronica

- E' stata la **prima "killer" application di Internet** (*ora è il Web*)
- L'idea di comunicare via computer in realtà anticipa Internet...
 - Anni '60: scambio informazioni tra utenti dello stesso computer
 - Poi comunicazione tra computer diversi (Arpanet)
- E' una delle prime applicazioni di Internet
- Lo "stile" moderno vuole **email** invece di **e-mail**

Indirizzo di posta elettronica

rossi.mario@mail.unimo.it

nome_utente / host.domain

nome_utente è un nome scelto dall'utente o dall'amministratore del server, che identifica in maniera univoca un utente (o un gruppo di utenti), mentre *host.domain* è un nome DNS (hostname)

Indirizzi email

nome_utente@host.domain

- **Formato di indirizzi email**
 - Max 64 caratteri nella parte nome_utente
 - La componente locale dovrebbe essere trattata come “case sensitive”, anche se molti mail server, per garantire maggiore compatibilità, scoraggiano o escludono questa possibilità
 - Max 255 caratteri nella parte host.domain
 - Solo caratteri, cifre, -, .
 - Questa parte non è “case sensitive”
- **Gli indirizzi di provenienza delle email non devono essere considerati affidabili perché non il sistema di email non prevede meccanismi di autenticazione!**

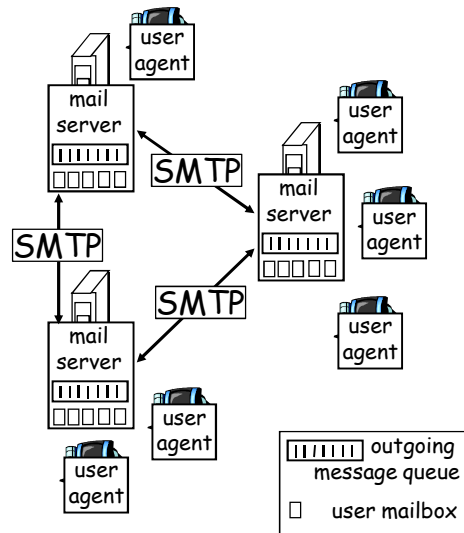
Posta elettronica

- Tre componenti principali:

- **user agent**
- **mail server**
- **Simple Mail Transfer Protocol (SMTP)**

- **User Agent**

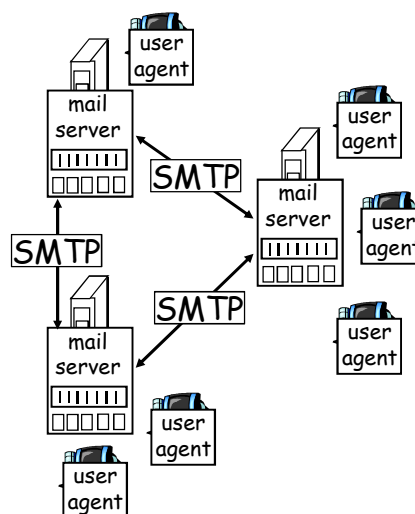
- "lettore della posta"
- composizione, editing, lettura messaggi di posta
- esempi: Eudora, pine, elm, Netscape Messenger, Outlook, Express
- messaggi in ingresso/uscita memorizzati sul mail server



Mail server

Mail Server

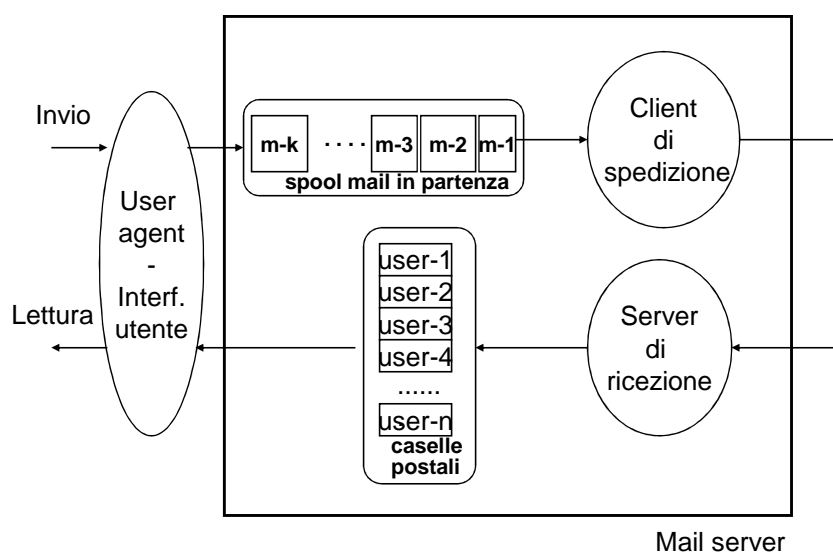
- **mailbox** (casella di posta) contenente i messaggi in ingresso (ancora da leggere) dell'utente
- **coda di messaggi di posta in uscita** (da inviare)
- **protocollo SMTP** tra mail server per inviare messaggi di posta



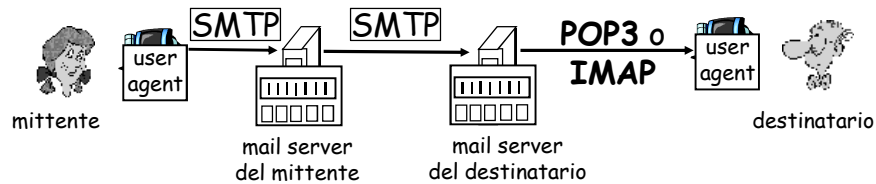
SMTP: Simple Mail Transfer Protocol

- "Protocollo elementare di trasferimento postale"
- Protocollo tra mail server
 - Definisce comandi per la comunicazione
- Paradigma client/server, in cui:
 - **client**: mail server del mittente
 - **server**: mail server del destinatario
 - **client e server in esecuzione su ogni mail server**
- Permette soltanto di inviare messaggi di posta, ma non di richiederli!
 - La fase di richiesta serve solo al client di posta (user agent), che userà altri protocolli (POP3 e IMAP)

Architettura di mail



Protocolli di accesso alla posta



- **SMTP: consegna/memorizzazione al mail server del destinatario. (NON ALLO USER AGENT del destinatario)**
- **Servono protocolli di accesso alla posta arrivata sul mail server. Es.**
 - **POP: Post Office Protocol**
 - **IMAP: Internet Mail Access Protocol**
 - più caratteristiche (maggiore complessità)
 - manipolazione dei messaggi memorizzati sul server
 - **HTTP: accesso alla mail tramite Web (Hotmail, Yahoo! Mail, ecc.)**

Accesso alla posta tramite Web

- Servizi di posta elettronica tramite Web (es., Hotmail e Yahoo! Mail)
- **Il Web browser è lo user agent**
 - Comunicazione con mailbox su mail server tramite HTTP
- **Vantaggio**
 - Permette all'utente di modificare la propria mailbox come se fosse locale
 - es., definizione di cartelle (folder) remote nella mailbox
 - es., ricerca di determinati messaggi nelle cartelle remote
- **Svantaggio**
 - Maggiore lentezza

Parte 6

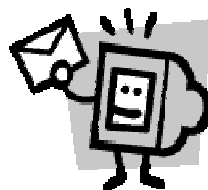
Modulo 2: Netiquette

Netiquette

What is Netiquette? Simply stated, it's network etiquette -- that is, the etiquette of cyberspace.

And "etiquette" means "the forms required by good breeding in social or official life."

In other words, Netiquette is a set of good rules for behaving properly online



V. Shea

La filosofia della *netiquette*

- Il codice della strada si riassume in:
 - non creare pericolo per gli altri
 - non creare intralcio agli altri
- La ***netiquette*** si riassume in:
 - Non sprecare risorse (di tempo, di rete, di calcolo e di spazio di memoria) non proprie
 - Ricordarsi che gli altri interlocutori sono esseri umani anche se la comunicazione è mediata da una tastiera e un video (non li vediamo in faccia...)

Ufficialmente

- **Il rispetto della netiquette non è imposto da alcuna legge**, ma si fonda su una convenzione ormai di generale condivisione
- Sotto un aspetto giuridico, **la netiquette è spesso richiamata nei contratti di fornitura di servizi di accesso da parte dei provider**
- Il mancato rispetto della netiquette comporta una **generale disapprovazione da parte degli altri utenti della rete**, solitamente seguita da un isolamento del soggetto "maleducato" e talvolta dalla richiesta di sospensione di alcuni servizi utilizzati per compiere atti contrari ad essa

Premessa

- La comunità degli utenti dei servizi di Internet (email, chat, blog, newsgroup, ...) è ampia, variegata, distribuita in tutto il mondo con enormi differenze culturali
 - ➔ **Poche regole generali che siano chiare e condivise**
- Inoltre, si compone anche di molteplici comunità più piccole
 - ➔ **Ogni piccola comunità potrebbe avere proprie regole più restrittive o più permissive delle regole generali**

Due contesti di comportamento

Originariamente nata per regolare la posta elettronica, la **netiquette** oggi si applica a tutte le interazioni sociali in rete

- **Comunicazioni 1 a 1**
 - Email
 - Chat
- **Comunicazioni 1 a molti**
 - Mailing lists
 - Newsgroups
 - Chat
 - Blog
 - Forum

Comunicazioni 1 a 1

- **L'altro utente è un essere umano che può avere una cultura diversa**
 - Attenzione a unità di misura, fuso orario, acronimi, ecc.
 - Es. “ci risentiamo domani mattina” in una mail che è rivolta ad un destinatario oltre-oceano
 - Si usa generalmente un testo non formattato (non tutti i mail system consentono corsivi, grassetti e colori)
 - Uso appropriato di caratteri per dare enfasi:
 - Parti in grassetto si indicano tra asterischi: ***bold***
 - Tutto maiuscolo → equivale a **SHOUTING** (usare con criterio)
 - Emoticon/smiles → :-) ;-) :-(

Nella posta elettronica in particolare

- Il **subject della mail deve esserci** (è considerato molto maleducato non indicarlo) **e deve essere legato al contenuto**
- Usare bene i “Re:”, “Fwd:” nel subject
- **Ridurre lo spreco di banda**
 - Limitare la dimensione degli allegati
 - Comprimere i file prima di spedirli
 - Chiedere prima di inviare allegati enormi (limiti nelle dimensioni delle mailbox)

Quoting

- **Quoting** → riportare sezioni di un messaggio a cui si risponde
- Il quoting aumenta la leggibilità dei messaggi (fornisce contesto)
- Non si quota tutto, ma solo le parti interessanti
- Evitare quoting multipli e insensati
- Usare [...] per indicare parti di testo originale mancante (non riportato con quoting)

Esempio di quoting

MIGLIORE

-- CC wrote:
> Resto in attesa
> della tua opinione
Concordo al 100%
> e ti ricordo
> l'appuntamento
> di domattina
Arrivo alle 7.30

PEGGIORE

Concordo
sull'opinione e ci
vediamo alle 7.30
-- CC wrote:
> Resto in attesa
> della tua opinione
> e ti ricordo
> l'appuntamento
> di domattina

Email con destinatari multipli

- Potenziali problemi di privacy e di sicurezza
 - Alcuni mailer inseriscono tutti i destinatari del messaggio ricevuto nell'addressbook
 - Gli addressbook sono usati anche dai virus nella loro propagazione
- Attenzione al REPLY ALL

**Valutare l'uso di Bcc (Ccn)
Blind Carbon Copy (Copia Carbone Nascosta):
i destinatari non vedono gli altri destinatari**

Comunicazioni 1 a molti

- **Molti principi della comunicazione 1 a 1 si applicano anche alla comunicazione 1 a molti**
- *In più, con molti ascoltatori il rischio di offendere/annoiare qualcuno è maggiore*
 - Serve molta più attenzione
- Newsgroup, mailing list, blog sono piccole società con loro costumi, tradizioni e regole
 - Leggere le linee guida della comunità
 - “Lurk before talk”
 - Guardare archivi, best-of, FAQ

In caso di errore

- Sbagliare è umano
- La netiquette (e l'educazione in genere) raccomanda di chiedere scusa non appena ci si accorge dell'errore

Netiquette 2.0

- **Nuove regole di netiquette nate per il Web 2.0**
- All'interno dei Social Network si instaurano tante relazioni tra singoli utenti
 - E' importante fare attenzione a quali informazioni vengono fornite in questo contesto, evitando di condividere **dati personali e di contatto**, come numeri di telefono o indirizzi, che nella vita reale non si darebbero a persone che non sono degne di fiducia
 - E' bene evitare di **scambiare file con utenti di cui non ci si può fidare** e in ogni caso, anche quando si conosce l'interlocutore, è necessario verificare sempre l'origine dei file ed effettuare un controllo con un antivirus aggiornato

Netiquette 2.0

- Quando si scambiano contenuti multimediali bisogna essere sicuri di averne il diritto e di non utilizzare alcun file coperto da **copyright**
- Quando si riscontra un **comportamento riconducibile ad un illecito** durante una conversazione privata (es. tentativo di approccio sessuale verso minori, stalking o cyberbullismo) si devono usare gli appositi sistemi di reportistica degli abusi offerti dal servizio, segnalando tempestivamente il nickname che ha perpetrato l'abuso
- Quando si contribuisce con del materiale in un ambiente condiviso, esso deve essere **coerente con il contesto e le regole della community**
 - Es. No materiale pubblicitario!

I doveri dei due protagonisti

- Utenti
 - rispettare il buon gusto e avere buon senso
 - rispettare la netiquette
 - rispettare le norme della comunità a cui si accede
- Amministratori di rete, di blog, di newsgroup
 - stabilire linee guida e regole chiare
 - rispondere tempestivamente alle segnalazioni di infrazione alla netiquette e alle regole
 - non abusare del proprio potere nella censura

Parte 6

Modulo 3: Cercare informazioni nel Web

Barzelletta

Un sera un passante vedendo un uomo ubriaco carponi sotto un lampione, si avvicina e gli chiede:

P: Avete perso qualcosa?

U: Sì, le chiavi di casa.

P: Ma dove le avete perse esattamente?

U: Laggiu'. Vicino alla porta di casa.

P: E allora perché la cercate qui?

U: Perché qui c'è più luce!

E' importante cercare nel posto giusto!

Cercare nel posto giusto

Nell'era pre-digitale...

- “Andare in posti noti”
 - Per reperire informazioni sui contributi pensionistici chiediamo all'INPS
 - Per trovare la via più breve tra Lucca e Pistoia consultiamo una cartina della Toscana
 - Per trovare informazioni sui personaggi storici del Risorgimento, andiamo in biblioteca

Come si organizza l'informazione?

→ **Classificazione gerarchica (tipica biblioteca)**

- L'informazione solitamente è raggruppata in un piccolo numero di categorie, ognuna descritta semplicemente (classificazione di primo livello)
- L'informazione in ogni categoria è suddivisa in sotto-categorie (classificazione di secondo livello), e così via
- Alla fine le classificazioni saranno così piccole che sarà possibile scorrere un'intera categoria per trovare l'informazione desiderata

Ma quando l'informazione non è organizzata (come la maggior parte dell'informazione Web)?

Usare un “motore di ricerca”

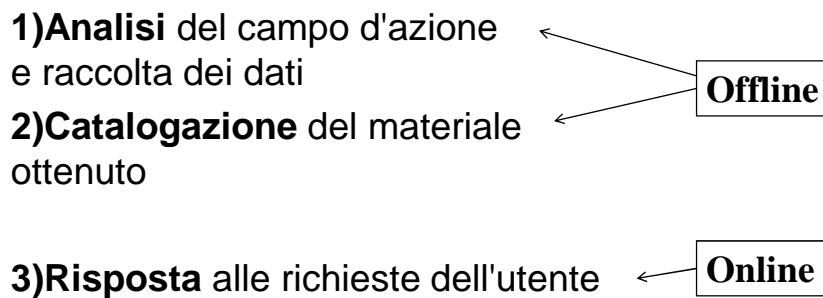
- Un **servizio** che aiuta nella ricerca di documenti su Web
- La ricerca è basata su una richiesta (*query*) basata sull'uso di parole chiave (*keyword*)
- Il risultato è una **lista di risorse Web** che contengono la/e parola/e chiave richiesta/e
- **Funzionamento base:**
 - **analisi** di un insieme di dati (spesso raccolti dal motore stesso)
 - restituzione di un indice dei contenuti disponibili **classificati in base a formule statistico-matematiche** che ne indicano il grado di rilevanza rispetto ad una determinata chiave di ricerca

Interfaccia dei motori di ricerca

- Ogni motore di ricerca contiene una zona scrivibile nella quale digitare la query
 - Anche integrato nel browser Web
- Premendo il tasto SEARCH (o tasti analoghi) il motore avvia la ricerca dei documenti presenti sul Web che si adeguano alla richiesta
 - *NOTA: La ricerca nel Web non è in tempo reale!*
- Viene ritornata una lista di risultati in **ordine di rilevanza**
 - In base a qualche misterioso algoritmo
- Il più famoso è **www.google.it**

Le fasi di funzionamento

Il lavoro dei motori di ricerca si divide principalmente in **tre fasi**



Analisi

- Per analizzare il Web i motori di ricerca utilizzano dei programmi detti **crawler** (o **spider** o **robot**)
- I crawler si occupano di *navigare continuamente ed indipendentemente* tra i siti Web e seguire tutti gli URL che trovano all'interno dei documenti analizzati
- Di volta in volta *inseriscono in un database* tutte le informazioni "importanti" della pagina (il contenuto testuale, più varie informazioni su di essa, come la data di ultimo aggiornamento, e altro)

Catalogazione

- Dopo la fase di analisi, le pagine vengono *inserite nell'indice del motore di ricerca*
- Un programma, chiamato ***indexer***, legge i documenti raccolti e crea un indice basato sulle parole contenute in ogni documento
- NOTA: Ogni motore di ricerca utilizza un **programma proprietario** per indicizzare i risultati in modo opportuno
- L'indice creato durante la fase di catalogazione verrà in seguito analizzato per fornire le risposte alle ricerche degli utenti

Risposta

- **Rispondere alle richieste degli utenti implica la necessità di elencare i siti in ordine di rilevanza rispetto alla richiesta ricevuta**
- **Per stabilire la rilevanza vengono cercati nel database quei documenti che contengono la/e parola/e chiave cercata/e, che poi sono *classificati attraverso algoritmi proprietari***
 - Es. Si controlla quante volte le parole chiave vengono ripetute, quanti link riceve quel documento, in quali punti della pagina sono poste le parole chiave, quanti altri documenti contengono link verso quella pagina, o quante volte un utente ha visitato quel sito dopo una ricerca

Aspettative utente

- *Ci si aspetta che i documenti più interessanti o maggiormente collegati alla lista di parole chiave fornita si trovino in cima alla lista*
- Questo può non succedere se:
 - Il programma di indicizzazione non è efficiente
 - La selezione di parole chiave fornita dall'utente è ambigua o troppo generica

Colpa dell'utente!



Come scrivere una query

- Solitamente, le parole di una query vengono cercate all'interno dei documenti in **qualunque ordine e preferibilmente (ma non necessariamente) tutte insieme**
- Inoltre esse vengono **filtrate** per eliminare parole poco significative (articoli), punteggiatura, ... Questa operazione è detta **stoplist removal**
 - Es. sto cercando: *La divina commedia; opera di Dante*
 - Dopo l'operazione di stoplist removal ottengo: *divina commedia opera Dante*

Come scrivere una query

- Le parole chiave vengono anche filtrate per unificare le parole con radice comune (e.g. Canto cantare cantante cantano canti...)
 - Questa operazione è detta **stemming – espansione della ricerca**
 - Ad esempio se un utente cercherà la parola "Danza" il motore ricercherà e proporrà fra i risultati anche le parole "Danzatrice" , "Danzatore" etc...

Riassumendo

- Gli indirizzi che vengono riportati corrispondono a pagine che contengono:
 - *Tutte o alcune delle parole della query*
 - *Parole simili a quelle della query (p.es. un'altra voce dello stesso verbo)*
 - *A volte, parole con significato correlato a quello delle parole della query*
- Gli indirizzi vengono poi ordinati per **rilevanza** (con criteri che variano da motore di ricerca a motore di ricerca)

Problemi dei motori di ricerca

- Se la parola chiave è **troppo generica o ambigua** (è associata a più significati diversi) si può avere un **sovraccarico di informazioni**:
 - Il motore di ricerca ritorna una lista troppo lunga di risultati
 - Es. parola “calcio”... elemento chimico o sport??

Troppa informazione = nessuna informazione

- Soluzioni
 - **Raffinare** la ricerca, per esempio aggiungendo altre parole chiave (Es. calcio elemento chimica)
 - Utilizzare gli **operatori booleani** forniti dalle opzioni di ricerca avanzata del motore

Raffinare la ricerca

- **Meglio una query con molte parole che una query meno specifica**
- **Strumenti di ricerca avanzata**
- **Ricerca di frasi o di parole adiacenti:**
 - Se si racchiude la frase (o le parole) tra virgolette, sono restituiti solo i documenti in cui la frase compare in modo esatto (con le parole nell'ordine)
- **Termini obbligatori:**
 - In alcuni casi io voglio essere sicura che le parole che scelgo siano comprese nei testi trovati (tipo l'AND logico)
 - Precedo ogni parola obbligatoria da un segno +

Operatori booleani

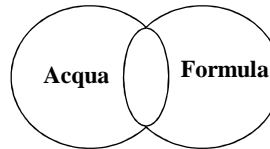
- Il *matematico inglese George Boole* (1815-1864) fondò un campo della matematica e della filosofia chiamato **logica simbolica**
- Il suo nome è rimasto legato ad un insieme di operatori che sono molto utili e molto presenti nel campo dell'informatica: gli **operatori booleani**
- Nell'ambito dei motori di ricerca gli operatori booleani sono utili per definire **operazioni di ricerca avanzate**

Operatori booleani di base

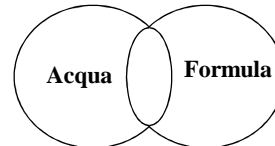
- **AND**
 - Saranno restituiti documenti in cui compaiono **TUTTE** le parole specificate con AND, e non soltanto alcune
 - In alternativa si può usare il segno +
- **OR**
 - Saranno restituiti documenti in cui compare anche solo una delle parole specificate, **NON NECESSARIAMENTE TUTTE**
- **NOT**
 - Saranno restituiti documenti in cui **NON** compaiono le parole precedute da NOT
 - In alternativa si può usare il segno -

Esempi

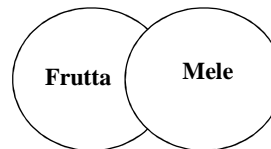
- Acqua + formula



- Paperino OR Topolino



- Frutta -MELE



Combinazione di operatori booleani

- Gli operatori booleani possono essere combinati e raggruppati usando le parentesi per cambiarne la priorità, come gli operatori matematici
- *Es.*
(meringa OR fragola OR cioccolato) AND semifreddo
- Senza parentesi sarebbe ambigua
meringa OR fragola OR cioccolato AND semifreddo
- Poteva essere interpretata come
meringa OR fragola OR (cioccolato AND semifreddo)

Come escludere dei termini

- **Esclusione di termini (operatore NOT):**
 - Talvolta una stessa parola ha due significati diversi. Se vogliamo essere sicuri di ottenere solo risposte relative ad un significato **possiamo escludere esplicitamente parole** che ci ricondurrebbero all'altro
 - *Esempio:* se cerco informazioni relative al calcio (nel senso di materiale chimico) posso provare ad escludere i risultati relativi allo sport nel seguente modo:
 - Calcio -sport -pallone (- equivale a NOT)

Ma quanto è affidabile il Web?

- **Come valutare l'affidabilità delle informazioni?**
 - Non dare nulla per scontato!
- **Prima regola: verificare l'identità della persona o organizzazione che pubblica il sito Web**
 - Può non essere banale: caso famoso di www.gatt.org
 - Il nome del dominio ricorda il General Agreement on Tariffs and Trade (GATT) , accordo per il libero commercio creato negli anni '90
 - Il sito si presenta come pubblicazione della World Trade Organization (WTO) di cui riporta logo e attività
 - In realtà gestito dal gruppo Yes Men, contro il libero mercato
 - Nel 2001 truffa a danno degli organizzatori di un evento per il libero commercio

Caratteristiche dei siti affidabili

- Esistenza nel mondo reale
- Certificazioni e citazioni
- Chiarezza
- Aggiornamento
- Professionalità

In ogni caso la via migliore è:

- Verificare chi ha registrato il nome del sito Web
 - InterNIC (Internet Network Information Center)
www.internic.net/whois.html

La pagina del cane burmese

- Una indicazione dell'esistenza di una espressione è data dal numero di occorrenze rilevate da un motore di ricerca
- Ma bisogna fare attenzione...
- La ricerca su Google di
Burmese AND mountain AND dog
restituisce più di 100.000 risultati, di cui uno è il sito del "Burmese Mountain Dog Club of America" e sembra affidabile...
- Però tra le razze canine il cane burmese **NON ESISTE**

La pagina del cane burmese

- Esiste però un cane di montagna bernese
- La ricerca con
Bernese AND Mountain AND Dog
restituisce 709.000 risultati
- Il cane bernese è identico a quello spacciato per burmese
- In compenso esiste un gatto burmese (molti dei risultati precedenti riportano sia razze di cani che di gatti)
→ Una pagina che sembrava affidabile non lo è!