

Terza esercitazione di I.G. Reperimento di informazioni in rete

Esigenza informativa

Ho deciso di ricercare notizie in linea su di una località della Campania, Palinuro(SA), dove mi reco in vacanza da ormai sedici anni. La scelta è caduta su tale argomento in primo luogo per la curiosità di vedere in che modo questa cittadina sia rappresentata sul World Wide Web ed in seconda battuta in quanto la mia lunga frequentazione dei luoghi mi consente di poter valutare più facilmente la veridicità delle informazioni reperite. Inoltre mi interessa verificare quanta parte di questo materiale sia dedicato al timoniere di Enea, Palinuro appunto, che secondo il racconto virgiliano dell'Eneide cadde in mare annegando proprio all'altezza del promontorio che oggi porta il suo nome. In partenza ho stabilito di ricercare solo informazioni in lingua italiana e riservarmi di decidere circa un'ulteriore indagine sulla base dei risultati ottenuti.

Strumenti utilizzati e motivazione

Nel condurre la mia ricerca ho deciso di avvalermi sia di un indice sistematico sia di alcuni motori di ricerca veri e propri. La scelta dell'indice sistematico è caduta su YAHOO(<http://www.yahoo.com>) e sulla sua versione italiana(<http://www.yahoo.it>) in quanto si tratta del più famoso rappresentante della propria categoria. Per quanto riguarda i motori di ricerca per termini sono stati utilizzati:

1. VIRGILIO.COM, scelto per verificare la resa di un motore di ricerca che nella sua collezione includa solo documenti in lingua italiana.
2. ALTAVISTA.COM, che consente una ricerca avanzata attraverso operatori booleani ed altre opzioni avanzate e della sua versione italiana(<http://www.altavista.it>)
3. GOOGLE.COM, caratterizzato dal fatto di fornire per ogni ricerca una graduatoria di rilievo consentita da un complesso algoritmo matematico che prende in considerazione il numero di altri siti che rimandano a quello reperito tramite la ricerca. Inoltre l'utente può navigare sulla copia locale del sito che il motore di ricerca riproduce.

(Fonte: M. Calvi F. Ciotti G. Roncaglia M.A. Zela *Internet 2000 – Manuale per l'uso della rete*, Laterza Roma-Bari 1999)

Fase di ricerca(informazioni aggiornate al 16/01/2001)

Nell'iniziare la mia ricerca ho optato per una query del tipo *Palinuro Cilento*(la zona territoriale in cui si trova la località di interesse)*Salerno*(la provincia), presupponendo così di individuare siti relativi almeno al contesto geografico soggetto alla mia ricerca. L'ordine delle parole inoltre è stato pensato per ottenere risultati significativi sia con i motori e gli indici per cui questa ricerca comporti il reperimento di tutte le parole della frase di richiesta(ovvero che la interpretino come *Palinuro AND Cilento AND Salerno*) sia per quelli che ricercano almeno una delle parole("OR"), di solito stabilendo un ordine di importanza da sinistra verso destra.

Il primo motore interrogato è stato, vista la restrizione del campo ai soli siti in lingua italiana, VIRGILIO che alla richiesta *Palinuro Cilento Salerno* ha fornito 39 pagine: di queste 8 si sono rivelate realmente attinenti trattando in un caso esclusivamente di

Palinuro(<http://www.palinurus.com>) e nei rimanenti del Cilento in genere e del relativo parco nazionale. Sei di questi siti sono stati presentati nella prima delle 4 pagine di risposte fornite, 1 nella seconda e 1 nella terza con una discreta presentazione per importanza. Fra i siti non di interesse molti sono relativi a strutture alberghiere e turistiche in genere nonché a fornitori di servizi di vario genere sia nella realtà del Cilento sia rispetto all'intera provincia di Salerno. Bisogna però rilevare anche l'assenza del sito ufficiale del Parco nazionale del Cilento(<http://www.parco.cilento.it>), URL reperito tramite un link del già citato <http://www.palinurus.com> che riporta anche l'indicazione del sito <http://www.palinuro.net>, associazione umanista reperita anche da Virgilio. Nessun cambiamento di rilievo digitando *Palinuro Cilento* (45 risposte, niente di pertinente in più).

La query iniziale è stata poi sottoposta anche ad ALTAVISTA specificando fra le lingue disponibili quella italiana ed ottenendo 90 risposte. Vagliando le prime 8 pagine di risultati forniti (scelta fatta partendo dal fatto che questo motore di ricerca fornisce le risposte in quello che stima essere l'ordine di pertinenza, peculiarità rivendicata nelle stesse pagine del search engine) sono stati reperiti 6 siti interessanti di cui tre già individuati con Virgilio (palinurus.com in relazione ad un suo file interno sulle case di vacanza e due relativi al Cilento in genere.), è presente invece come prima risposta in assoluto il sito ufficiale del parco anche se in riferimento ad un suo file specifico e non all'home page. Tra il materiale non di interesse ancora moltissime strutture alberghiere spesso riportate più volte e parecchio materiale di vario genere su Salerno. La ricerca diventa invece più fruttuosa con la frase di ricerca *Palinuro Cilento*: i siti diventano 140 ma aumenta notevolmente la loro pertinenza con 13 risposte di interesse nelle prime sette pagine e fra essi il sito ufficiale del parco, <http://www.palinuro.it> (sito specifico su Palinuro), la già citata associazione umanista (pagina dei risultati numero 6) e anche palinurus.com (benché solo nell'ultima schermata consultata). Da segnalare infine che Altavista prevede un conteggio del numero di volte in cui le parole cercate ricorrono nelle risposte fornite e che il motore in fase di ricerca semplice in mancanza di ulteriori istruzioni cerca pagine che contengano almeno uno dei termini ordinandole a partire da quelle che li contengono entrambi.

GOOGLE invece alla richiesta *Palinuro Cilento Salerno* ha risposto con 339 pagine reperite nelle quali, come poi visibile consultando la copia nella cache del motore dove erano evidenziate, risultavano presenti contemporaneamente tutte le parole, almeno nell'ambito delle prime pagine di risposte fornite. Dodici sono risultate le pagine attinenti delle quali 6 presentate nella prime pagine di risposte (fra cui le due esclusivamente dedicate a Palinuro) mentre tra le altre schermate troviamo il sito ufficiale del parco, Cilento nel mondo (<http://www.cilentonelmondo.it>) ma non l'associazione umanista (<http://www.palinuro.net>). Usando l'opzione *I'm feeling lucky* che porta al sito più significativo secondo il sistema si arriva a <http://starnet.it/italy/palinuro/sseverin.htm>, relativo al vicino borgo abbandonato di S. Severino di Centola e già reperito da ALTAVISTA, e tornando indietro fino all'home page si scopre che quest'indirizzo è un alias di <http://www.palinuro.it>!

Bisogna anche dire che GOOGLE ha fornito fra le risposte molte pagine relative alla storia del Cilento, assolutamente assenti negli altri motori, che però non sono state conteggiate fra quelle di interesse. A titolo di esperimento è stata sottoposta anche la query *Palinuro Cilento* che ha fornito 1020 risposte peggiorando inoltre la resa sul piano dell'ordine dei siti forniti (si noti che per ALTAVISTA era successo il contrario).

A questo punto la ricerca è stata affidata a YAHOO a partire dalla versione italiana ed usando le sue categorie e precisamente il percorso è stato *Aree geografiche > Paesi > Italia > Regioni > Province > Salerno > Città* non ottenendo alcun risultato. Una volta chiesto al sistema di cercare all'interno delle sue categorie sulla base della consueta frase di ricerca sono stati reperite 3 pagine relative ad alberghi. Usando la versione internazionale per fare lo stesso percorso all'interno delle directory i risultati sono stati nulli

ed alla richiesta di cercare nelle directory *Palinuro Cilento Salerno* non è stata ugualmente reperita alcuna risposta. A questo punto vista la scarsità dei risultati si è fatto ricorso al motore di ricerca per termini di Yahoo che, a livello di ricerca semplice, non consente la scelta della lingua, reperendo 63 pagine, in gran parte in Inglese, di cui solo 3, tutte in lingua anglosassone appunto, parzialmente di interesse. Tra esse il sito ufficiale del parco in versione inglese, un file piuttosto scarno, sempre in inglese, già reperito precedentemente in lingua italiana ed un altro file con poche informazioni per altro inesatte (il povero Palinuro diventa timoniere di Ulisse!). Un tentativo con *Palinuro Cilento* non porta a cambiamenti di rilievo anche se aumentano le pagine (130). A questo punto viene tentata la ricerca con la versione italiana che dichiara di appoggiarsi a GOOGLE ottenendo con la prima interrogazione 254 pagine con 11 siti attinenti fra i primi 60. Primo fra questi viene fornito il file sul borgo di S. Severino all'interno dell'alias di palinuro.it (<http://www.stanet.it/italy/palinuro/ssverin.htm>), come in GOOGLE vero e proprio, poi varie risposte spesso già riscontrate fra le quali lo stesso <http://www.palinuro.it> ed il sito ufficiale del parco, manca invece, almeno nelle prime pagine, <http://www.palinurus.com>.

L'interrogazione *Palinuro Cilento* infine provoca un aumento del numero delle risposte ma i documenti significativi restano sostanzialmente gli stessi. Arrivati a questo punto della ricerca è stato fatto un tentativo di eliminare i siti commerciali, gli alberghi ed altro materiale relativo alla zona di Palinuro ma non alla località. I primi tentativi di ricerca avanzata sono stati compiuti con ALTAVISTA a partire dalla query (*Palinuro AND Cilento*) AND NOT albergo* (la wide card in questo motore è ammessa anche nella ricerca avanzata) ottenendo 154 risposte e nella prima pagina di risultati (5 utili) sono presenti vari fra i siti più pertinenti (palinuro.it, il suo alias, palinurus.com, il sito ufficiale del parco nazionale oltre a <http://www.cilento.net> già varie volte reperito). A questo punto però la ricerca si complica in quanto all'esclusione di "hotel", "camping" e via via di altri termini legati a servizi, ristoranti e simili la riduzione del numero di risposte si accompagna alla perdita di pagine significative tra cui anche le due dedicate solo a Palinuro che contengono parole come "hotel", "servizi" e simili. In risposta alla query "*Capo Palinuro*" AND *Cilento* invece sono state associate 1280 pagine in gran parte legate alla meteorologia (a Capo Palinuro è presente una stazione meteorologica dell'Aeronautica), la richiesta di eliminarle ("*Capo Palinuro*" AND *Cilento*) AND NOT (meteo* OR weather) però non ha dato i frutti sperati portando le pagine a circa 500 ma senza guadagnare in pertinenza visto che le prime risposte riguardavano alberghi, agenzie di servizi o di viaggi e altre pagine non pertinenti). Un ulteriore tentativo di raffinare la ricerca ha visto l'uso del qualificatore *title: Palinuro* che ha restituito 35 risposte fra le quali molte attinenti ed in particolare <http://www.palinuro.it> (1° risposta), <http://www.palinurus.com> (3°), <http://www.palinuronet.com> (28°) e <http://www.cilento.it> (31°). Il qualificatore *link* invece ha dato due risposte per palinuro.it (dalla società creatrice del sito che offre inserzioni sulla pagina e da <http://www.novamedia.it/Sienall/salepaes.htm>, pagina che elenca link a tutti i paesi del salernitano, per Palinuro ci sono anche palinurus.com e cilentonet.it) e tre per palinurus.com (<http://www.casevacanze.it>, a sua volta segnalato dal sito richiamato, <http://www.ziezi.tripod.com> ed il già citato elenco dei paesi del Salernitano). Il ricorso alla versione italiana di ALTAVISTA non è stato soddisfacente: la query *Palinuro Cilento Salerno* ha dato infatti 5832 risposte e fra le prime 80 solo la numero 23 (palinurus.com) è risultata attinente. Uniche note positive sono state i ricorsi al qualificatore *title* ed al qualificatore *link*. In questo caso il primo ha fornito 28 risposte pertinenti come nella versione internazionale (l'alias di palinuro.it, palinurus.com, <http://www.palinuronet.com>) anche se peggio disposte come ordine di pertinenza (18°, 23°, 28°), il secondo 3 link a palinuro.it (gli stessi dell'altra versione più <http://www.hotelgraal.it>) e quattro a palinurus.com (casevacanze.it, l'elenco dei paesi del salernitano, una sottopagina

del sito stesso(?) e <http://www.rastir.com>, società che ha realizzato il motore di ricerca interno al sito). La ricerca avanzata di GOOGLE invece non prevede l'uso di operatori booleani ma permette di selezionare tramite dei menù a tendina le parole da includere e da escludere più altri indicatori, quali la data di creazione ed il dominio (presenti anche in ALTAVISTA), non utilizzati per la query. Una volta selezionate le parole da includere o meno la query si presenta con gli stessi segni + e - della ricerca semplice di ALTAVISTA. Per prima cosa è stata tentata la query "Capo Palinuro"+ Cilento che ha restituito 473 pagine relative in gran parte a siti sulla meteorologia ma con presenti fra le prime 60 palinuro.com(2 volte), cilentonet.com e palinuro.it mentre l'opzione *Mi sento fortunato* ha fornito un file interno ad un sito di meteorologia con le previsioni del tempo attorno a Capo Palinuro. Visti questi risultati è stata formulata la query "Capo palinuro"+ Cilento - weather -meteo -vacanze -viaggi (queste ultime due chiavi per limitare la presenza di agenzie turistiche) ottenendo 403 siti con presenti fra i primi 70 palinurus.com, palinuro.it, cilentonet.com, il sito ufficiale del parco(<http://www.parco.cilento.it>) e palinuronet.com, con una buona resa qualitativa nonostante il numero ancora alto di siti reperiti. GOOGLE fornisce anche un unico qualificatore, *link*, opzione presente in un menù a tendina nella sua ricerca avanzata che dà la possibilità di cercare link ad una pagina specifica. Tale strumento è stato usato in relazione ai due siti dedicati esclusivamente a Palinuro ovvero <http://www.palinuro.it> e <http://www.palinurus.com> ottenendo per il primo 10 link (di cui 9 a partire dal suo alias ed uno dal sito di un hotel) e nessuno per il secondo (un link era presente in <http://www.palinuronet.com>, pagina non reperita da questo motore).

Per quanto riguarda VIRGILIO la ricerca avanzata non consente l'uso di operatori booleani (anche se ciò è impossibile da capire a partire dai vari menù *help* presenti) permette solo di scegliere fra *cerca tutte le parole* (in pratica AND) ed *almeno una parola* (OR), con "Capo Palinuro" Cilento opzione *tutte le parole* dunque si ottengono 5 risposte di cui l'unica coerente è palinurus.com. YAHOO infine si appoggia a GOOGLE anche per la ricerca avanzata e consente solo di scegliere fra un'opzione *chiave AND chiave* ed una *chiave OR chiave* mentre tutte le parole scritte in altri modi vengono codificate con dei segni + qualunque sia la loro sintassi. In questo modo "Capo Palinuro" Cilento viene espressa come "Capo+Palinuro"+Cilento ed addirittura "Capo Palinuro" Cilento AND NOT meteo diventa, nella tendina del browser, "Capo+ Palinuro"+ Cilento+ AND+ NOT+ meteo con risultati di ricerca imprevedibili. In ogni caso con la query "Capo Palinuro" AND Cilento si ottengono 102 pagine e fra le prime 40 figurano palinuro.it, palinuronet.com, l'alias di palinuro.it, <http://www.cilento.it> ed il sito ufficiale del parco. A questo punto vista la difficoltà di raffinare la ricerca senza perdere informazioni significative, considerato il fatto che i documenti significativi presentati ormai erano stati individuati e tendevano a ripetersi ed infine che nemmeno i link interni ai siti d'interesse segnalavano altre pagine utili non recuperate ho deciso di terminare qui la ricerca.

Risultati ottenuti e valutazione

In definitiva la ricerca ha consentito di individuare un numero abbastanza ampio di pagine che presentassero una qualche attinenza con l'argomento di interesse ma al contempo solo alcuni di esse si sono rivelate, ad un'analisi più accurata, realmente utili per l'esigenza informativa espressa.

Infatti sono da considerarsi soddisfacenti:

1. <http://www.palinurus.com>, sito non a scopo di lucro creato dalla classe 5°B dell'Istituto per geometri "E.Cenni", il migliore quanto a completezza e correttezza delle informazioni, tra gli altri link quelli a <http://www.palinuronet.com> (vedi) ed a <http://www.parco.cilento.it> (vedi)

2. <http://www.palinuro.it>, l'altro sito dedicato solo a Palinuro, prodotto da un'azienda specializzata in presentazioni Web, è presente sul Web anche con un alias, <http://www.stanet.it/italy/palinuro>.
3. <http://www.palinuronet.com>, sito dell'associazione umanista che porta il nome di Palinuro, ricco di informazioni e con link, fra gli altri, a <http://www.palinurus.com> ed a <http://www.parco.cilento.it> (sito ufficiale del Parco Nazionale del Cilento, di cui Palinuro fa parte, vedi sotto)
4. <http://www.parco.cilento.it>, sito del Parco Nazionale del Cilento che presenta informazioni aggiornate ed attendibili sul parco ma anche sul circondario.
5. <http://www.cilentonelmondo.it>, sito relativo al Cilento in genere ma con un'ampia e dettagliata pagina dedicata a Palinuro
6. <http://www.cilentonet.it>, ancora un sito sul Cilento in genere che oltre a pagine sui servizi offre un'ampia panoramica di tipo turistico sulle varie località, compresa Palinuro.

Sono stati poi reperiti altri siti validi sul Cilento in genere spesso presenti anche nei link dei siti citati ma dal carattere più spiccatamente commerciale.

Inutile ribadire che tutti i motori hanno poi restituito molto più materiale di quello qui riportato, si tratta però nella stragrande maggioranza dei casi di singoli file, spesso all'interno di varie sottodirectory, interni a siti che parlano di zone territoriali più ampie, a volte dell'intera penisola, e che le informazioni fornite spesso non sono corrette ed in ogni caso sono estremamente scarse. La stessa leggenda di Palinuro, che spesso diventa come lamentato anche dagli autori di palinurus.com il nocchiero di Ulisse e non di Enea, prende in molti siti le sfumature più varie. Inoltre scarsissima è la presenza di materiale storico-letterario sul Web, tanto che, ad esempio, la mia preoccupazione di reperire molte pagine riguardanti l'Eneide si è rivelata piuttosto priva di fondamento. La valutazione del materiale reperito è stata fatta dunque, oltre che sulla base delle mie conoscenze in materia, in riferimento ad alcuni parametri: in primo luogo è stato prestato particolare interesse a siti che trattassero in modo specifico di Palinuro o del circondario in modo da limitare la frammentarietà riscontrata in varie presentazioni. Inoltre, nell'ambito dei siti citati, sono stati tendenzialmente esclusi quelli a forte connotazione commerciale (es. fornitori di servizi sul territorio) presupponendo che le eventuali notizie turistico-culturali non fossero state al centro della progettazione del sito; questo criterio è stato tuttavia applicato in modo critico, tanto che nell'elenco dei risultati utili figura <http://www.cilentonet.it>, legato ad un'attività commerciale ma ricco di notizie valide. Ultimo filtro, ma non certo per importanza, è stata la valutazione dell'attendibilità attribuibile agli ideatori/autori dei siti: la scelta è caduta dunque su di un sito non a scopo di lucro progettato da una scuola e che ha ricevuto i complimenti del locale assessore al turismo (presenti in forma di e-mail nel sito stesso), sulla presentazione Web di un'associazione culturale locale, sulle pagine ufficiali di un ente, quello gestore del Parco Nazionale ed infine su tre presentazioni create nell'ambito di un progetto più vasto (l'azienda di palinuro.it per esempio ha creato analoghi siti per molte località della Campania) o segnalate dalle pagine ritenute sicuramente attendibili.

Per quello che riguarda la valutazione del lavoro svolto dai motori di ricerca è necessario fare delle distinzioni: infatti senza dubbio valida si è rivelata la ricerca svolta da ALTAVISTA e da GOOGLE i quali attraverso le varie query presentate hanno sostanzialmente restituito tutti i documenti di principale interesse anche se bisogna segnalare che, rispetto all'ordinamento per rilevanza che dovrebbe essere prerogativa di entrambi, i risultati migliori li ha dati GOOGLE. Si tratta di un aspetto estremamente importante in quanto l'utente non può certo vagliare tutto il materiale proposto e dunque, specie in un motore con un numero di pagine indicizzate molto grande come

ALTAVISTA, il rischio di perdere informazioni importanti recuperate ma non riportate fra le prime risulta alto. Per quanto riguarda invece VIRGILIO e YAHOO l'indagine, pur permettendo il recupero di materiale d'interesse, ha risentito delle scarse possibilità di selezione avanzata offerte dai due strumenti con l'aggravante di istruzioni in linea di qualità inferiore a quelle degli altri due motori di ricerca. Resta da dire in ogni caso che tutti i motori in ultima analisi hanno restituito delle informazioni in grado di rispondere alla richiesta informativa e che dunque un utente esperto o almeno consapevole sarebbe stato in grado di reperire del materiale utile attraverso questi strumenti;alcuni problemi riscontrati esulano infatti dalle caratteristiche delle varie search engine e verranno di seguito brevemente trattati

Conclusioni

A questo punto del lavoro è dunque possibile trarre delle conclusioni circa questa ricerca di informazioni online ed anche rispetto alla risorsa informativa del Web in genere. Le risposte dei motori di ricerca riguardo alla mia esigenza informativa mi hanno portato infatti a reperire una grande quantità di materiale che però, purtroppo, si presenta come approssimativo e frammentario; molto difficile inoltre risulta valutare l'attendibilità di ciò che si è reperito se non si possiede una conoscenza minima dell'argomento dato che molti di questi documenti, vista anche la facilità di pubblicazione(assolutamente positiva dal punto di vista del principio) di informazioni sul Web, non fanno capo ad organizzazioni precise o ad enti riconoscibili come fonti credibili. A ciò si aggiunge il problema dell'ancora scarsa standardizzazione nella scrittura delle pagine che nel nostro caso penalizza molto tre dei siti maggiormente significativi, <http://www.palinurus.com>, <http://www.palinuronet.com> e <http://www.parco.cilento.it>, il cui ritrovamento meno frequente rispetto ad altre pagine potrebbe dipendere anche dall'assenza di tag meta nei loro sorgenti, tag che sono invece presenti, con riferimento allo schema del Dublin Core, nel più reperito <http://www.palinuero.it>. Questo problema è particolarmente significativo in quanto i tre siti sprovvisti di tag meta sono anche gli unici tre la cui credibilità sia più semplicemente accertabile(fanno capo ad una scuola locale, ad un'associazione culturale ed all'ente gestore del Parco) nonché i soli a permettere una valida navigazione orizzontale attraverso i propri link(si segnalano tutti a vicenda). Questi problemi dunque, uniti ai limiti strutturali dei motori di ricerca(nel mio caso l'esclusione di parole quali ad esempio "albergo" portava all'automatica eliminazione di pagine importanti nelle quali questo termine era incidentalmente presente, impedendo di rendere più precisa la ricerca), fanno sì che la ricerca sul World Wide Web sia una risorsa preziosa ma che al contempo richieda all'utente un notevole senso critico e non sia oggi ancora effettuabile disgiuntamente dall'uso di altre fonti informative di natura diversa. Ciò non toglie importanza alla funzione dei motori di ricerca che rappresentano invece un'ottima porta d'accesso alle risorse presenti nella rete a patto che poi l'utente ne sappia fare un uso ragionato. C'è da sperare che l'evoluzione del World Wide Web porti con sé una maggiore regolamentazione degli standard in modo da superare le difficoltà a tutt'oggi presenti nell'indicizzazione e permettendo così l'accesso a quel cospicuo numero di pagine che al momento rimangono nascoste all'utente ed insieme renda possibile a quest'ultimo valutare con più facilità e sicurezza l'attendibilità di ciò che egli reperisce, magari a partire proprio dall'ordine di presentazione dei documenti fornito dai motori di ricerca, strada che, per fortuna, si comincia già oggi a percorrere.