Motori di ricerca a confronto per il turismo

Un'indagine esplorativa per alcune aree protette nel Friuli Venezia Giulia

Andrea Favretto e Giovanni Mauro*
Dipartimento di Scienze Geografiche e Storiche
Università degi Studi di Trieste

1 – Introduzione

Il turismo, per sua stessa natura, è un prodotto complesso ed intangibile che, dal lato dell'offerta, risente pesantemente della difficoltà, per l'utente, di valutare concretamente il suo futuro acquisto, al fine di confrontarlo con le altre possibilità offerte dal mercato ed arrivare in tal modo ad una scelta consapevole[1]. Quando si tratta di scegliere una destinazione per una vacanza, spesso infatti il turista si affida prevalentemente alle sue sensazioni, mentre per esempio legge un catalogo o ascolta le soluzioni che la sua agenzia di viaggi di fiducia gli prospetta.

Un altro elemento determinante, nella scelta della destinazione, è indubbiamente il passaparola, ovvero i racconti delle esperienze di viaggio di amici e parenti (Lanza et alii, 2007).

Il progressivo ed inesorabile affermarsi di Internet ed in particolare del *World Wide Web* nei Paesi sviluppati ha determinato, per certi mercati, il superamento dei tradizionali modelli di *marketing*, portando alla ribalta un ulteriore importante canale informativo; in tal modo è nato ed è subito stato apprezzato, per immediatezza ed economicità, il cosiddetto "turismo fai da te", scollegato dalle agenzie di viaggi (almeno per la fase informativa del processo di acquisto del servizio) ed appoggiato alla rete ed al gran numero di portali e siti turistici in essa presenti. Internet ha ormai profondamente rivoluzionato il modo di

fare pubblicità in ambito turistico e sta velocemente cambiando anche i sistemi di vendita dei servizi connessi. Il punto nevralgico della questione è, come sempre, legato ai costi: sono bassi per l'utente che si collega da casa per ricercare informazioni su una meta per le vacanze; sono altrettanto competitivi per l'operatore turistico che vuole attrarre clienti attraverso la pubblicità. È infatti meno dispendioso allestire un sito turistico su un server del *World Wide Web* e pubblicizzarlo in Internet, piuttosto che appoggiarsi ai canali di promozione tradizionali (ad es.: la posta a domicilio, gli spot pubblicitari in televisione o alla radio, gli annunci su quotidiani o riviste). Internet arriva capillarmente ai potenziali utenti, presso le case ed i posti di lavoro; la sempre maggior velocità di accesso, unita alla differenziazione ed alla moltiplicazione dei media attraverso i quali connettersi, favoriscono inoltre una pratica consultazione nei ritagli di tempo anche durante l'orario di lavoro (IPK Int., 2001).

La rete viene sempre più intensamente utilizzata anche a fini turistici, per acquistare (cosiddetto *online booking*), oppure solo per informarsi e contattare poi direttamente l'erogatore del servizio (*online looking*), e ciò si può verificare facilmente sulle statistiche del turismo. Ad esempio, IPK International, nel suo *World Travel Trends Report* (WTTR) di novembre 2007 riporta che, nel periodo gennaio-agosto 2007, in Europa, il 50% dei viaggiatori ha utilizzato Internet per prenotare i servizi turistici (soprattutto biglietti per l'aereo o per gli altri vettori di trasporto), nonché per acquisire informazioni turistiche (rispettivamente, il 36% ed il 14%); la restante metà dei turisti ha utilizzato canali di informazione e vendita tradizionali, quindi non collegati alla rete (IPK Int, 2007).

Il Web, come noto, è un enorme magazzino di informazioni, alcune corrette, molte spazzatura; è inoltre estremamente variegato per ciò che riguarda le tipologie di argomenti che si possono ottenere ed il loro livello di approfondimento; ma, soprattutto, nel Web tutto è drammaticamente dinamico, nel senso che l'aspetto delle pagine, i contenuti delle stesse e soprattutto gli indirizzi telematici presso i quali reperire le informazioni (sono i cosiddetti "url"[2]) variano molto velocemente nel tempo. Da quanto ricordato, è facile arguire che la ricerca di informazioni nel *Web* non può essere un'operazione semplice e routinaria. Eppure le statistiche dimostrerebbero il contrario e ciò è facilmente verificabile non solo in ambito turistico, ma per qualsiasi argomento che possa suscitare un qualche interesse[3]. Ma allora, come si fa una ricerca di informazioni in rete? La risposta al quesito è ovvia ed immediata: ci si collega ad un motore di ricerca e lo si interroga attraverso una serie di parole chiave. Pochi sentono il bisogno di approfondire l'argomento al di là della conoscenza della sintassi, più o meno corretta, che permette di ottenere velocemente dei risultati (ci si accontenta, insomma, di conoscere le sole "istruzioni per l'uso"). Il tutto è molto simile a quello che succede quando si guida una macchina: lo si sa fare, ma non si dispone di una conoscenza approfondita della struttura dell'automobile e del suo funzionamento. Tutto sommato, non è necessario conoscere la meccanica delle automobili: quando si rompono, le si porta dal meccanico; è una delle comodità della divisione del la-

voro, sulla quale è notoriamente basata l'economia dei paesi economicamente sviluppati. Si vive bene lo stesso se non si sa molto anche della ricerca in rete, se si ignora il funzionamento degli algoritmi che la realizzano o quali siano le strutture hardware e software che la gestiscono. È però innegabile che una maggiore conoscenza dell'argomento garantisce dei risultati più efficienti (ovvero risposte più veloci e maggiormente approfondite e dettagliate ai quesiti posti).

In questa nota viene presentata una breve analisi di come le risorse turistiche di tipo naturalistico vengono valorizzate e comunicate nel particolare caso della ricerca in rete mediante un motore. Dopo una schematica introduzione alla ricerca in Internet e una generale spiegazione di come funzionano i motori che la realizzano, sono verificati i risultati di alcune *query* (quesiti), di argomentazione turistica, sui tre attualmente più diffusi motori di ricerca (Google, Yahoo, MSN). I quesiti vertono su informazioni di tipo generale, indicazioni su pernottamenti e ristorazione per alcune aree protette della Regione Friuli Venezia Giulia, allo scopo di focalizzare le diversità delle risposte nei tre casi e soprattutto se e come, attraverso la rete, vengano comunicate le caratteristiche principali degli ambiti tutelati.

2 – La ricerca in rete e i motori che la realizzano

Dal 1993, quando Internet consisteva in poche centinaia di siti, alle decine di miliardi di pagine Web attualmente disponibili (per non parlare dei documenti disponibili in rete in un formato diverso da quello delle pagine html, quali ad esempio i testi, le immagini, i filmati o i file audio) la ricerca ha dovuto raccogliere sfide impegnative, ma nello stesso tempo è riuscita a trasformarsi, per mantenere un alto standard di risultati, tali da giustificarne l'esistenza. Come giustamente fanno osservare alcuni Autori, conservare l'enorme mole di dati presenti in rete ha senso solo se questi ultimi sono accessibili agli utenti (Battelle, 2006).

Da un punto di vista generale, l'accesso all'informazione digitale conservata nei database è divenuto problematico in seguito al progressivo incremento nelle dimensioni degli archivi informatici stessi. Spesso si sente dire che quella attuale è l'era dell'informazione. Attraverso la tecnologia database per la registrazione e la gestione dei dati in formato digitale sono infatti decenni che viene ammassata una miriade di informazioni, delle più svariate tipologie; ad esempio possiamo ricordare le transazioni di tipo economico-commerciali di aziende ed individui, i dati scientifici, quelli medico-sanitari, generali e personali, le immagini da satellite, i testi, i video, ecc. Diventa pertanto di strategica importanza lo sviluppo di metodi efficaci per l'estrazione, l'analisi e l'interpretazione dei dati archiviati. A tal fine operano i processi cosiddetti di *Knowledge Discovery in Databases* (KDD, Figura 1), procedure composite formate da diversi livelli operativi che, partendo dalla massa dei dati residenti nell'archivio, ne elaborano progressivamente le parti di volta in volta utilizzate, in modo da riconoscere in esse

eventuali caratteristiche comuni o modelli secondo i quali si organizza la loro distribuzione di frequenza; il fine ultimo di tutta la procedura è, naturalmente, quello di arricchire la conoscenza finale degli utilizzatori del database.

Lo stadio più importante dell'intero processo è conosciuto come *data mining* ed esso è cruciale nel sistema, visto che prevede l'applicazione di tecniche evolute[4] per identificare connessioni o causalità fra le variabili.

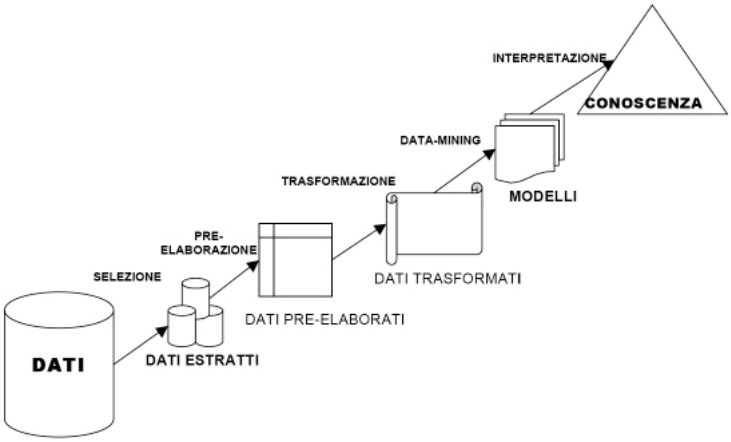

Figura 1
Il processo cosiddetto *Knowledge Discovery in Databases*, del quale il *data mining* è il cuore

Fonte: elaborazione da Zaiane (1999)

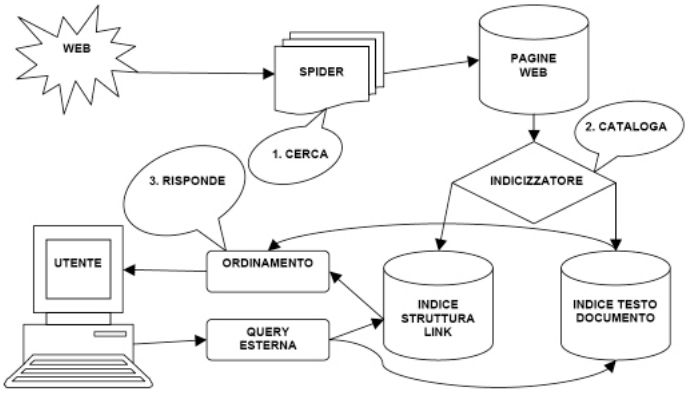

Figura 2
Schema del processo funzionale di un motore di ricerca

Fonte: elaborazione da Zaiane (1999)

Come ricordato, i processi del tipo KDD si applicano ai database di grandi dimensioni, per i quali cioè le *query* in linguaggio SQL (*Structured Query Language*), tradizionalmente utilizzate nei modelli relazionali, non sono soddisfacenti, nel senso che non riescono a circoscrivere i dati che si vogliono estrarre dall'archivio o addirittura producono risultati sbagliati (ad es. vengono estratti dati non richiesti). La tipologia dei dati non è una limitazione per i processi KDD, che funzionano con archivi multimediali, spaziali (Sistemi Informativi Geografici), di serie storiche, di transazioni commerciali e, naturalmente, con il più eterogeneo e dinamico dei contenitori attualmente esistente: il *World Wide Web*.

L'utilizzo delle tecniche proprie di *data mining* per scoprire ed estrarre in modo automatizzato informazioni dal Web è conosciuto come *Web mining* (Kosala e Blockeel, 2000). Come è facilmente intuibile, data la grande dimensione dell'archivio da utilizzare e le caratteristiche di dinamicità ed eterogeneità del suo contenuto, le tecniche di *Web mining* sono particolarmente complesse. Comunemente sono distinguibili tre principali sottocategorie funzionali, che si differenziano in quanto utilizzano diverse componenti dell'informazione registrata nel Web per arrivare al risultato finale (Furnkranz, 2004):

- *Web content mining*, ovverosia l'applicazione della tecnologia in oggetto a testi strutturati o semi-strutturati (ad es.: i documenti in formato html);
- *Web structure mining*, si utilizza la struttura dei collegamenti che caratterizza il Web come fonte di ulteriori informazioni (rispetto a quelle ricavate dai testi);
- *Web usage mining*, un ulteriore fonte informativa è costituita dalle interazioni fra l'utente che richiede le informazioni in rete e il server che viene interpellato ed eventualmente fornisce le informazioni (ad es.: i file di *log*[5] dei server in Internet).

I motori di ricerca in Internet sfruttano le tecniche di *Web mining* per garantire dei risultati ai quesiti dei navigatori della rete. Nella fattispecie, viene sfruttata la struttura ipertestuale dei documenti registrati nel Web attraverso algoritmi appositamente sviluppati (punto 2); i risultati del lavoro degli algoritmi sono inoltre combinati con ciò che fuoriesce dall'analisi dei contenuti dei documenti (punto 1) ed anche con eventuali informazioni ricavate dal traffico registrato dai server.

Un motore di ricerca utilizza le parole chiave digitate dall'utente per cercare, in particolari file chiamati indice, tutte le pagine ed i file collegati alla richiesta dell'utente. I file indice sono stati precedentemente costruiti dal motore di ricerca attraverso l'analisi di documenti scaricati dalla rete e residenti su molti computer collegati in parallelo (Figura 2).

Le funzionalità di un motore di ricerca possono essere scomposte in tre stadi: raccolta, indicizzazione ed ordinamento.

– *Raccolta*
La raccolta dei documenti, anche detta *crawling*, viene fatta da speciali software chiamati *spider* o *Web robot*, il cui compito è quello di visitare un grande numero di pagine Web e di scaricare quelle che ritiene opportune sui computer del motore di ricerca. Lo *spider* inizia con alcune pagine "radice" e, dai collegamenti presenti su queste pagine, raggiunge nuove pagine, attraverso i cui collegamenti procede nell'ispezione del Web. Le pagine scaricate vengono conservate in un database, che viene periodicamente aggiornato in seguito ad ulteriori processi di raccolta, data la dinamicità del Web. Naturalmente il processo di raccolta è molto importante, perché la bontà dell'informazione fornita dal motore di ricerca dipende chiaramente dall'attendibilità dei documenti presenti nel suo database.

– *Indicizzazione*
L'indicizzazione di un documento avviene in base al testo, che viene rappresentato come una sequenza di termini:

$$d = [(t_1, k_1), (t_2, k_2), \dots (t_n, k_n)]$$

ove d è il documento da indicizzare; $t_1, \dots, t_n$ sono i termini presenti nel documento; $k_1, \dots, k_n$ sono le ricorrenze del rispettivo termine nel documento.
I termini vengono opportunamente pesati, in modo da scremare eventuali termini ripetuti in molti documenti, che quindi non indicano una peculiarità del testo stesso (ad esempio articoli, congiunzioni, parole molto utilizzate quali Internet, Web, libro, ecc.).
Relativamente ai documenti, vengono analizzati e pesati anche i cosiddetti TAG[6] di formattazione presenti sul formato informatico del documento, propri del linguaggio html per la costruzione delle pagine Web.
All'indicizzazione segue la catalogazione del documento, che consiste nell'inversione dell'indice, ovvero la costruzione di un riferimento relativo a documenti presenti nel database del motore di ricerca, in corrispondenza di ciascuna parola possibile oggetto di ricerca.

– *Ordinamento*
In conseguenza di un'interrogazione sono quindi selezionati, mediante il file indice, tutti i documenti in cui sono presenti i termini dell'interrogazione stessa. Generalmente si tratta di migliaia di pagine, che devono quindi essere ordinate per importanza, prima di essere presentate all'utente sotto forma dei collegamenti consigliati nelle schermate. La funzionalità che esegue tale ordinamento viene definita *ranking* e, come si può intuire, è molto importante, visto che, generalmente, l'utente medio visiona solo i collegamenti consigliati nelle prime due pagine dell'elenco a video (in genere ci sono una decina di collegamenti per schermata). I collegamenti presenti nelle prime

posizioni dei risultati sono considerati quindi più importanti dei successivi. L'importanza del collegamento dipende *in primis* dalla rispondenza della pagina Web con i termini di ricerca inseriti dall'utente, poi anche da considerazioni collegate alla struttura Web, ovvero all'importanza ed al "prestigio" che la pagina ha all'interno del *World Wide Web*. Molto spesso l'ordinamento dei collegamenti deriva da un'integrazione fra le caratteristiche del testo, in riferimento alla *query*, e l'importanza del sito che ospita la pagina Web.
Google, ad esempio, ha brevettato un meccanismo automatico che calcola l'importanza della singola pagina all'interno del Web, denominato *page-rank*. L'algoritmo si ispira al mondo delle citazioni accademiche (il cosiddetto *impact factor* degli articoli scientifici) ed è stato sviluppato da Sergey Brin e Lawrence Page, due studenti di dottorato in informatica all'Università di Stanford (California), che, successivamente, hanno fondato un'impresa che ha completamente rivoluzionato Internet. Google, attualmente, vale in borsa più di Ford e General Motors messe assieme (Vise e Malseed, 2005). L'idea chiave dell'algoritmo *page-rank* è che una pagina Web è tanto più importante quanto maggiormente essa è citata nel Web (cioè è un collegamento consigliato su altre pagine); non solo: l'importanza di una pagina dipende anche dall'importanza dei siti sui quali sono presenti i collegamenti alla pagina stessa. In altre parole, se una determinata pagina Web è consigliata ad esempio dal sito della NASA, tale collegamento in entrata le conferisce un'importanza più elevata (cioè un alto valore di *page-rank*) rispetto ad un collegamento in entrata dal sito personale di un privato, o di un'associazione meno prestigiosa (conseguentemente con un *page-rank* meno elevato) della NASA. Tutto ciò è naturalmente espresso in formule ed applicato all'intero Web, scaricato sui computer di Google dai programmi *spider*, di cui si è parlato sopra (per approfondimenti, si veda Brin e Page, 1998). Il valore di *page-rank* viene integrato con altri cento "ingredienti" per l'ordinamento e, naturalmente, la ricetta di Google è segreta, come quella della Coca Cola (Carpineto e Romano,2005).
Yahoo, anch'esso fondato da due dottorandi di Stanford[7], ha un approccio totalmente diverso da Google, in quanto basa le sue ricerche su un catalogo di pagine Web costruito manualmente e costituito da *directory*, strutturate gerarchicamente ad albero in base a categorie quali arte, scienza, business, ecc. La struttura delle categorie viene anche qui integrata con altri elementi per l'ordinamento, ma il cuore del motore è diverso, umano; Google, invece, ne ha uno bionico. Tutto ciò non può non influenzare i risultati delle ricerche che, in questo caso, sono più orientate alla richiesta di ulteriori raffinazioni da parte dell'utente (ne è un esempio l'opzione *also try*, che Yahoo propone con i suoi risultati per consigliare il navigatore sulla base del contenuto delle *directory*).
Non si desidera in tal sede approfondire ulteriormente le differenze fra i vari motori di ricerca, per le quali si rimanda alla bibliografia (Battelle, op. cit. e, per le differenze più squisitamente operative, Lunghezzani, 2004).

3 – Parchi e Riserve Naturali in Friuli Venezia Giulia

Il Friuli Venezia Giulia è una regione caratterizzata da una grande complessità ambientale in un'area piuttosto limitata, ricca di contenuti anche di valore internazionale (dalle lagune al Carso, dalle risorgive alle Dolomiti). Le profonde trasformazioni che hanno interessato il territorio (il riordino agrario, la bonifica di estese aree della Bassa Pianura Friulana, lo sviluppo di nuovi poli industriali a San Giorgio di Nogaro, a Monfalcone e a Osoppo) hanno fatto nascere l'esigenza di una pianificazione territoriale per riconoscere il ruolo e l'importanza degli ambiti naturali e per garantire la loro tutela. Tra alterne vicende[8] nel 1996 sono stati istituiti due parchi naturali regionali, dieci riserve naturali, venti aree di reperimento e venti biotopi[9] per un'estensione totale di circa 89.000 ettari di aree protette[10] (Azienda dei parchi e delle foreste regionali, 1999).

Nell'ambito del presente studio sono state prese in considerazione alcune aree naturali del Friuli Venezia Giulia. In particolare sono state studiate le modalità con cui vengono presentate informazioni a carattere generale e di tipo turistico (pernottamento e ristorazione) delle seguenti aree protette:

- Parco Naturale Regionale delle Dolomiti Friulane
- Parco Naturale Regionale delle Prealpi Giulie
- Riserva Naturale Regionale Foci dell'Isonzo
- Riserva Naturale Regionale Foci dello Stella e Valle Canal Nuovo
- Riserva Marina Naturale di Miramare
- Riserva Naturale Regionale dalla Val Rosandra

La scelta è ricaduta su queste aree perché: a) i due parchi costituiscono le aree protette più importanti per estensione e per peso istituzionale (per ogni parco è stato costituito un ente regionale); b) le riserve naturali della Foce dell'Isonzo e dello Stella sono rappresentative delle zone umide costiere, ossia delle aree lagunari che interessano l'intera regione; c) la riserva naturale della Val Rosandra ricade nell'area del Carso Triestino e Goriziano, in cui insistono ben cinque aree protette[11] (tanto da dare luogo all'ipotesi dell'istituzione del "Parco del Carso"); d) la riserva marina naturale di Miramare, per le sue valenze di natura ambientale e per la contemporanea presenza di elementi architettonici a valenza storico-culturale, costituisce un elemento di forte richiamo turistico per la stessa città di Trieste.

4 – L'informazione su Internet derivante da ricerca semplice o combinata

Per cercare di comprendere come vengono fornite le informazioni su Internet, sono state eseguite alcune ricerche con modalità differenziate utilizzando tre

motori di ricerca tra i più diffusi: Google, Yahoo e MSN. In primo luogo, per acquisire informazioni a carattere generale sull'area naturale prescelta, è stato digitato Parco Naturale X o Riserva Naturale X[12]. È stata analizzata solamente la prima pagina dei risultati, ossia i primi dieci collegamenti suggeriti dal motore di ricerca.

I link derivati sulla prima pagina sono stati classificati come "link al sito ufficiale" quando richiamavano il sito ufficiale dell'area protetta, "link a matrice scientifica" quando richiamavano siti di carattere scientifico come Wikipedia o di associazioni scientifico-naturalistiche, come il sito della Federparchi (www.parks.it), WWF, Lipu, ecc., "link a matrice istituzionale" quando richiamavano siti di enti locali come la regione, la provincia, il comune, associazioni di carattere ufficiale, ecc., "link a associazioni di categoria o associazioni amatoriali" quando richiamavano siti di associazioni amatoriali come canoisti, ciclisti, escursionisti, ecc. Sui primi dieci collegamenti nella prima pagina dei risultati è stata valutata la frequenza assoluta delle diverse pagine richiamate così classificate.

Come si può dedurre dalla Figura 3, nel caso di una ricerca non troppo specifica, rivolta solo ad acquisire informazioni di carattere generale, non sussistono grandi differenze tra i diversi motori di ricerca. L'informazione su Google sembra essere più a matrice istituzionale: infatti questo motore di ricerca richiama il sito ufficiale, siti istituzionali e siti a matrice scientifica. Maggior peso ai siti di associazioni amatoriali (con informazioni di natura tecnica, quali guide alla flora, itinerari ciclistici o per escursionisti) viene invece data da Yahoo; tuttavia anche in questo caso il peso di siti a matrice scientifica (in particolare il sito della Federparchi e Wikipedia) appare molto evidente. Nel caso di MSN, mentre per i parchi il tipo di informazione che se ne deriva non è molto distante da quella proposta da Google, per le riserve (soprattutto Miramare e Val Rosandra) l'informazione appare meno precisa. Essa è talvolta a carattere tecnico, talvolta invece appare poco coerente con la richiesta (link a elementi di altre province italiane, senza connessione con la richiesta).

Per verificare quanto siano accessibili informazioni a carattere turistico relative alla possibilità di pernottare è stato digitato Parco Naturale X + pernottamento o Riserva Naturale X + pernottamento. Anche in questo caso è stata analizzata solamente la prima pagina dei risultati (ossia i primi dieci collegamenti suggeriti), classificati secondo le modalità descritte precedentemente. Tuttavia, essendo oggetto della ricerca informazioni di natura turistica, nella classificazione dei link si sono aggiunti i "link a matrice turistica" ossia i collegamenti a pagine di operatori turistici (agenzie, hotel, alberghi, ecc.).

Con questa modalità di ricerca emergono delle differenze abbastanza significative tra i tre motori di ricerca (Figura 4). Nel caso di Google non sono evidenti differenze sostanziali rispetto alla richiesta precedente, cioè quella relativa a informazioni di carattere generale. Ovviamente, in questo contesto, le pagine dei siti istituzionali, scientifici, ecc., richiamano notizie relative al pernottamento (notizie sulla foresteria dei parchi, ad esempio). La presenza

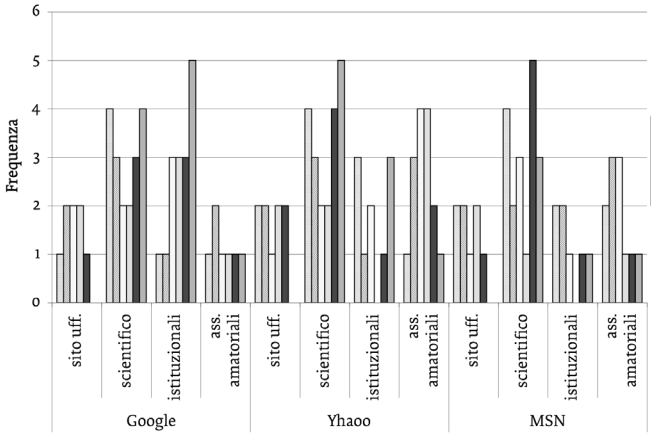

Figura 3
Ricerca di informazioni generali

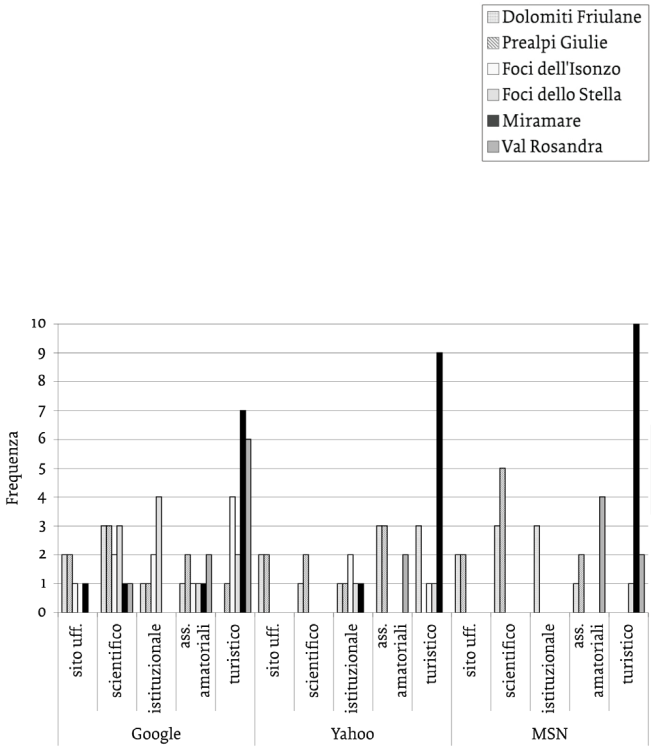

Figura 4
Ricerca di informazioni generali associate alla parola pernottamento

del sito ufficiale dei parchi italiani (il sito della Federparchi) diventa alquanto significativa per tutte la aree protette indagate. Forte è anche la segnalazione di link a matrice turistica, soprattutto relativamente alle riserve: in questo caso la presenza di siti istituzionali o scientifici viene alquanto ridotta a vantaggio dei siti degli operatori turistici (in particolare, per la riserva di Miramare e della Val Rosandra, probabilmente per visite collegabili a quella di Trieste).

Nel caso di Yahoo si rilevano più marcatamente le differenze tra i siti relativi ai parchi e quelli delle riserve. Mentre il sito della Federparchi è particolarmente efficiente nel segnalare le possibilità di pernottamento nei parchi, nel caso delle riserve si viene ad evidenziare una generale carenza numerica di siti individuati (1-2 pagine dei risultati). Inoltre, frequentemente i siti consigliati sono poco inerenti o riportano descrizioni di esperienze personali sulle visite fatte da altre persone alle aree protette, rendendo nel complesso difficoltosa la ricerca del potenziale turista. In particolare, si segnala che nel caso della riserva naturale delle Foci dell'Isonzo vengono individuate in totale solo sette collegamenti e, anche cambiando la ricerca con l'altra denominazione con cui la riserva è conosciuta ("Isola della Cona"), i risultati sono alquanto scarsi. In generale si può affermare che l'utilizzo di questo motore di ricerca per pernottare nelle riserve analizzate porta a risultati non soddisfacenti individuando molti siti non inerenti (Figura 6). I risultati relativi alla riserva di Miramare confermano quanto osservato per quelli della medesima ricerca effettuata con Google (ossia i siti rilevati sono numerosi solo perché il pernottamento è associato ad una visita turistica alla città di Trieste). Ciò viene ribadito anche dal terzo motore di ricerca, MSN, che accentua la profonda differenza di risultati esistente tra parchi e riserve, nell'individuazione di possibilità di pernottamento. La presenza di link a siti non coerenti con la richiesta diventa prevalente (Figura 6), rendendo perciò la ricerca poco efficiente.

Parallelamente a quanto fatto per il pernottamento, è stata realizzata una terza richiesta per verificare l'accessibilità a informazioni relative alla possibilità di ristorarsi presso ristoranti, trattorie, ecc. In questo caso è stato digitato Parco Naturale X + ristorazione o Riserva Naturale X + ristorazione.

Google evidenzia una buona capacità di rilevare l'informazione richiesta (Figura 5). Malgrado una parziale differenza tra l'informazione derivata per i parchi e per le riserve, le soluzioni proposte sono coerenti e potenzialmente utili all'utente di Internet. Per quanto riguarda i parchi viene dato peso ai siti ufficiali-istituzionali, ma soprattutto a quelli scientifici (sito di Federparchi). Relativamente alle riserve, i siti istituzionali acquisiscono meno peso rispetto a quelli di carattere turistico-ricreativo (talvolta andando fuori tema, come nel caso di Miramare, che diventa il riferimento per hotel con lo stesso nome in altre zone d'Italia).

Come nel caso del pernottamento, l'utilizzo del motore di ricerca Yahoo non fornisce risultati soddisfacenti. La presenza di siti poco congruenti riguarda sia le ricerche fatte sui parchi, che quelle sulle riserve. In generale viene dato poco peso ai siti delle associazioni amatoriali, mentre viene dato rilievo a siti con carattere istituzionale, che però si rivelano in realtà poco utili all'utente perché si

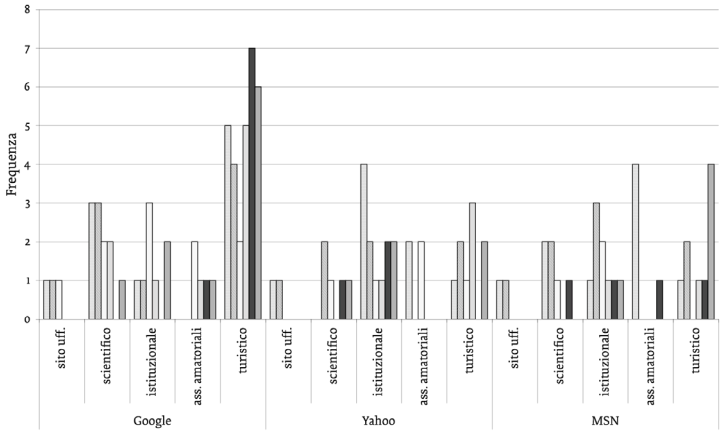

FIGURA 5
Ricerca di informazioni generali associate alla parola ristorazione

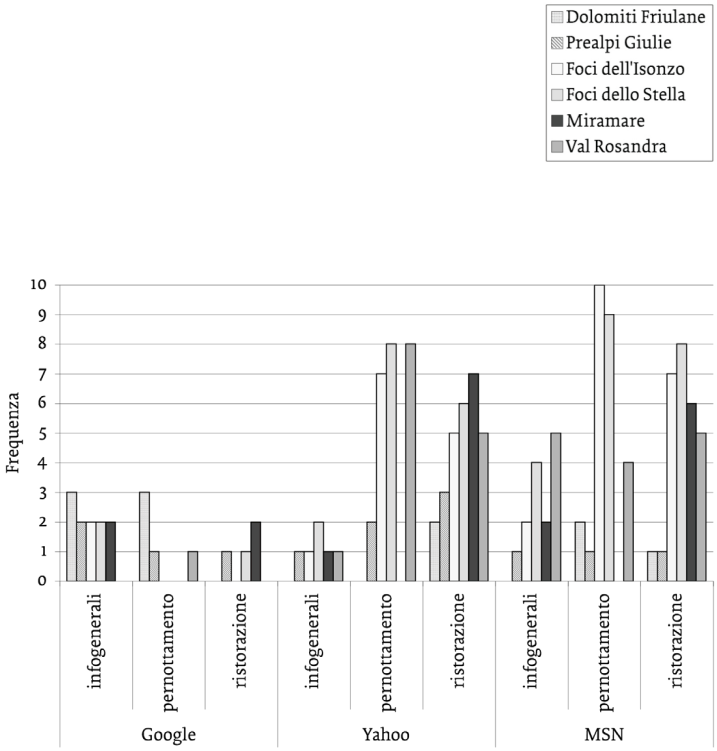

FIGURA 6
Frequenza di collegamenti a siti non coerenti con la richiesta effettuata

riferiscono a progetti di vario tipo (progetti sostenibili, censimenti vari, progetti internazionali) sviluppati dagli enti. Anche in questo caso i risultati per la riserva delle Foci dell'Isonzo sono pochi (nove collegamenti), anche se migliorano inserendo la denominazione "Isola della Cona" (tre pagine di risultati). La ricerca per la ristorazione associata alla riserva marina di Miramare è totalmente insoddisfacente (Figura 6), accentuando il fenomeno rilevato in Google (ossia l'associazione a strutture alberghiere con lo stesso nome, ma localizzate in altre zone in Italia). Risultati ancora peggiori si ottengono utilizzando MSN: in questo caso la ricerca va fuori tema anche per la riserva delle Foci dello Stella, oltre che per Miramare.

In sintesi si può affermare che, mentre il motore di ricerca Google si rivela efficiente sia nel caso di una ricerca a carattere generale, che relative a temi turistici, Yahoo e MSN evidenziano i propri limiti soprattutto nella ricerca associata. In particolare MSN rivela limiti anche nella ricerca di informazioni generali. I siti con *page-rank* più elevato (Tabella 1) sembrano venire richiamati con maggiore insistenza.

5 – I siti ufficiali

Per l'analisi dei siti ufficiali delle realtà ambientali considerate sono stati presi in esame la posizione che essi assumevano nella prima pagina dei motori di ricerca utilizzati (Google, Yahoo, MSN), il *page-rank* di Google e la tipologia di informazione offerta. Per poter fare questo è stata attentamente analizzata la struttura del sito (la cosiddetta "mappa del sito"), nella quale sono stati contati il numero di contatti per tipologia di informazione offerta.

I siti ufficiali sono i seguenti[13]:

- Parco Naturale Regionale "Dolomiti Friulane": www.parcodolomitifriulane.it
- Parco Naturale Regionale "Prealpi Giulie": www.parcoprealpigiulie.org
- Riserva Naturale Regionale "Foci dell'Isonzo": www.isoladellacona.it (www.riservanaturalefoceisonzo.it: dominio scaduto)
- Riserva Naturale Regionale "Foci dello Stella": www.riservenaturali.maranolagunare.com
- Riserva Marina "Miramare": www.riservamarinamiramare.it

Come si osserva dalla Tabella 1, non sempre i siti ufficiali ricoprono la prima posizione nella pagina dei risultati. Quando questo non accade, essi vengono preceduti dal sito Internet che la Federparchi ha sviluppato, il quale contiene comunque dei rimandi facilmente accessibili ai siti ufficiali dei parchi o delle riserve. Anche l'utente inesperto riesce facilmente ad accedere a tutte le informazioni che gli possono risultare utili. Per quanto concerne il *page-rank*, l'unico sito con punteggio elevato è quello della riserva marina di Miramare (0,5), ricco di informazioni, in particolare a carattere scientifico (Tabella 2).

Come anticipato, per poter operare un confronto tra i siti ufficiali dei parchi e/o delle riserve, è stata analizzata la "mappa del sito". Sono state classificate informazioni relative al parco, ai servizi offerti e alle cartografie.

Per quanto concerne le informazioni relative all'area protetta (Tabella 2), colpisce l'elevato numero di link a carattere scientifico (in cui vengono presentate informazioni, ad esempio, sulla flora, sulla fauna, sulla geomorfologia, ecc.), del sito della riserva di Miramare, nonché quelli relativi ai progetti e alle ricerche attualmente in corso presso l'ente. Una differenza sostanziale si rileva anche tra i siti dei due parchi: nel caso del Parco delle Dolomiti Friulane, viene dato maggior peso agli elementi istituzionali dell'ente stesso (con numerosi siti dedicati all'ente, ai centri visita, ai comuni interessati, ecc.).

Parco	Google	Page Rank	Yahoo	MSN
Dolomiti Friulane	3	0.4	1	1
Prealpi Giulie	1	0.4	1	1
Foci dell'Isonzo	3	0.3	2	3
Foci dello Stella	3	0.3	3	2
Miramare	1	0.5	1	1

Tabella 1
Posizione sito ufficiale nella ricerca di informazioni generali

Area Protetta						
Nome	Ente	Scientifici	Studio-Ricerca	Centri visite	Comuni	Contatti
Dolomiti Friulane	7	2	9	7	8	2
Prealpi Giulie	1	5	4	1	1	2
Foci Isonzo	-	-	-	-	-	-
Foci Stella	3	1	1	3	3	1
Miramare	6	45	13	1	3	1

Tabella 2
Numero dei link per tipologia di informazione offerta sul parco
nel sito ufficiale dell'ente considerato

	Servizi				
Nome	Pernottamento	Ristorazione	Escursioni/ Trasferimento	Lingue	Varie
Dolomiti Friulane	7	8	6	4	12
Prealpi Giulie	5	5	8	4	7
Foci Isonzo	-	-	3	-	-
Foci Stella	-	-	3	2	3
Miramare	4	4	9	2	7

Tabella 3
Numero dei *link* per tipologia di informazione offerta
sui servizi nel sito ufficiale dell'ente considerato

	Mappe		
Nome	Generale	Dettaglio	Interattive
Dolomiti Friulane	-	si	si
Prealpi Giulie	si	si	-
Foci Isonzo	si	-	-
Foci Stella	si	-	-
Miramare	si	si	si

Tabella 4
Tipologie di cartografia messa a disposizione sui siti ufficiali

Relativamente ai link segnalati sulla "mappa del sito" classificabili come inerenti ai servizi, si nota come non sempre il grado di informazione sia soddisfacente. Tuttavia, per quanto riguarda i due parchi, questa tipologia di informazione è sicuramente da ritenersi esauriente. Le indicazioni di questo tipo risultano invece non molto approfondite nel caso della riserva delle Foci dell'Isonzo (spiegabile con il fatto che il sito è attualmente in fase di manutenzione, cosa che spiega anche la ridotta informazione presente sul parco) e della riserva delle Foci dello Stella. Comunque bisogna evidenziare come anche in questo sito vengono ripe-

tutamente segnalate possibilità di escursioni con ristorazione associata. Altro limite da segnalare, che riguarda tutti i siti, è dato dal fatto che, mentre è possibile talvolta prenotare qualche attività collegabile al parco (escursioni in genere), non è consentito pagare le stesse utilizzando pagamenti automatici su Internet.

Le lingue in cui vengono presentate le informazioni sono generalmente l'italiano e l'inglese, cui si associano talvolta anche il tedesco e lo sloveno. Numerose altre informazioni sono messe a disposizione sui siti, come la situazione meteorologica, le immagini con webcam, i calendari, le foto, ecc. (sotto la voce "varie").

Infine, per quanto riguarda le carte disponibili presso i siti ufficiali delle aree protette studiate, si rileva anche in questo caso una certa eterogeneità. In generale, i siti con *page-rank* più elevato mettono a disposizione cartografia maggiormente curata e ricca di dettagli (in particolare il sito della riserva marina di Miramare ha una sezione dedicata, con informazioni su GIS, telerilevamento e con cartografia interattiva), mentre quelli con *page-rank* più basso forniscono solo mappe generiche. Nessun sito mette comunque a disposizione dell'utente una struttura *WebGIS* interattiva, in grado cioè di fornire informazioni su richiesta in tempo reale.

6 – Conclusioni

Pur nella sua inevitabile, parziale incompletezza, il lavoro svolto ha evidenziato alcuni risultati soddisfacenti, che saranno oggetto di verifica ed integrazione attraverso successivi approfondimenti. L'incompletezza a cui si fa riferimento è collegata ai tre siti per la ricerca interrogati (complessivamente, i motori di ricerca disponibili in rete sono ovviamente più numerosi), nonché al fatto che si è operata una scelta fra le risorse turistico-ambientali del Friuli Venezia Giulia. Si pensa tuttavia che i tre motori esaminati rappresentino un campione rappresentativo della categoria[14], al pari delle aree protette, selezionate in modo da riprendere sia le diverse tipologie di ente di tutela esistenti in regione (parco, riserva), sia le diverse caratteristiche morfologiche del territorio regionale (ambiti montani, aree umide, mare).

Volendo schematizzare i risultati ottenuti, si possono proporre le seguenti riflessioni.

- Per quanto riguarda le differenze riscontrate nelle risposte dei diversi motori di ricerca, si è visto che Google riporta più informazioni su siti di carattere ufficiale o istituzionale di Yahoo, che invece ha più associazioni amatoriali o di categoria fra i collegamenti consigliati; MSN, invece, si pone a metà strada, visto che per i parchi si comporta più similmente a Google, per le riserve naturali a Yahoo.
  Per ciò che riguarda Google e Yahoo, tale differenza è probabilmente spiegabile dal diverso metodo di lavoro dei due. Come visto, Google calcola il valore di *pa-*

ge-rank di una pagina Web (che serve per ordinare i risultati presentati all'utente) in base al numero ed all'importanza dei collegamenti in entrata di quella pagina; tipicamente questo è il caso delle pagine Web dei siti ufficiali (in altre parole: il sito ufficiale di un parco sarà maggiormente consigliato come collegamento da altre pagine "istituzionali", di prestigio, quali ad esempio enti pubblici o ministeri, piuttosto che, ad esempio, il sito di una associazione di ciclisti montani delle Dolomiti Friulane o delle Prealpi Carniche). Yahoo invece lavora utilizzando come base per la ricerca delle *directory* compilate a mano, sulla quale struttura sono innestati i meccanismi automatici che valutano anche i contenuti dei testi delle pagine o quelli dei TAG html. Tale caratteristica può aver determinato un comportamento del motore meno omogeneo e conseguentemente, nei risultati dello stesso, può essere stata attribuita uguale importanza alle pagine dei siti ufficiali e a quelle dei siti di categoria; a questo punto la scelta dell'ordinamento sarebbe stata fatta sulle caratteristiche del solo testo e del suo formato html (ad es. il numero di volte che nel testo compare il nome ricercato dall'utente navigatore, oppure, nel formato html, le parole chiave, la descrizione breve, il titolo, l'autore nei TAG). Le differenze rilevate sono più marcate nel caso di interrogazioni per informazioni di carattere generale alla risorsa turistica, se viene aggiunta alla *query* la richiesta per indicazioni su pernottamento o ristorazione, tutti e tre i motori "diluiscono" i risultati pertinenti con alcuni collegamenti chiaramente fuori tema; ancora una volta Google si rivela migliore rispetto agli altri due motori; inoltre, nuovamente i parchi sembrano essere meglio referenziati rispetto alle riserve naturali.

– Attraverso l'analisi delle varie "mappe dei siti" informativi, si può infine ricavare qualche indicazione utile al "turista internauta". Si è rilevato, infatti, che le informazioni più dettagliate e precise (sia di tipo generale che pratico) si trovano generalmente sui siti ufficiali delle aree protette, che hanno spesso indicati anche dei buoni collegamenti per ristoranti ed hotel nelle vicinanze. Va inoltre segnalato il sito trasversale www.parks.it, un'ottima fonte informativa anche a livello regionale, pur coprendo esso l'intero territorio nazionale.

Si è convinti che Internet giocherà in futuro un ruolo fondamentale per quanto riguarda la commercializzazione del turismo. Come accennato nell'introduzione di questa nota, il peso della rete a tal riguardo sta aumentando considerevolmente, sia da un punto di vista informativo (*online looking*) che commerciale (*online booking*). In tale ottica si pensa che il contributo dei motori di ricerca sia strategico per il futuro: quanto più essi saranno in grado di rispondere alle esigenze del pubblico, tanto maggiormente favoriranno l'affermarsi di un mercato turistico in Internet. Sono sempre molto importanti gli studi di settore, che però non possono essere limitati all'analisi e all'elaborazione dei dati del turismo in rete. È necessario, infatti, un corretto inquadramento della materia nel suo contesto teorico interdisciplinare ed è in tale direzione che il presente contributo ha inteso muoversi.

NOTE

* Il presente lavoro è stato realizzato in piena collaborazione fra i due Autori. Tuttavia si desidera precisare che i paragrafi 1, 2 e 6 sono da attribuire ad Andrea Favretto; i paragrafi 3, 4, 5 a Giovanni Mauro.

1 Per un approfondimento in generale sulla Geografia del Turismo, si rimanda, fra gli altri, a Lozato Jotart, 1999; Innocenti, 2003.

2 URL sta per *Uniform Resource Locator*, indirizzo standard che indica la collocazione di una risorsa informativa presente in rete. In genere si indica con l'indirizzo web di un dominio (ad es. www.units.it è l'indirizzo url del dominio proprio dell'Università degli Studi di Trieste).

3 Viene stimato che quotidianamente vengono fatte ben 500 milioni di interrogazioni al web, attraverso le richieste registrate dalla totalità dei motori attivi in rete (Carpineto e Romano, 2005).

4 Si tratta di tecniche che spaziano dalla classificazione e l'identificazione di gruppi omogenei nei dati, all'identificazione di eventuali modelli sequenziali o regressivi nel loro comportamento, fino ad arrivare a scoprire eventuali periodicità o salti nel caso di serie storiche (cfr., fra gli altri, Seifert, 2004; Zaiane, 1999), applicando algoritmi matematico-statistici e procedure reiterative informatiche di tipo *machine-learning*.

5 Il file di *log* contiene un elenco di tutte le operazioni effettuate da un computer. Tale file è tipicamente mantenuto dai server web che ospitano i siti informativi in rete e di conseguenza vengono interpellati dagli utenti navigatori.

6 All'interno di ogni pagina in formato htlm sono presenti dei marcatori (i cosiddetti TAG), che hanno differenti nomi a seconda della loro funzione. I TAG importanti per caratterizzare le pagine web ai fini della ricerca sono soprattutto quelli che indicano il titolo del documento, l'autore, la descrizione del contenuto, eventuali parole chiave, ecc.

7 Jerry Yang e David Filo hanno fondato Yahoo nel 1995.

8 In estrema sintesi tali vicende possono essere così indicate: il Piano Urbanistico Regionale Generale del 1978 con l'ipotesi di porre a vincolo circa il 30% del territorio regionale per costituire il "sistema ambiente"; la ridotta applicabilità (a causa della complessità della materia) della conseguente L.R. 11/83; la legge quadro in materia di aree naturali protette L.N. n.394/91 e l'obbligo per le regioni di istituire aree protette con la partecipazione degli enti locali; la conseguente istituzione di aree protette con la Legge Regionale L.R. 42/96 (Azienda dei parchi e delle foreste regionali, 1999).

9 Ai sensi della L.R. 42/96 per Parco si intende «un sistema territoriale che, per valori naturali, scientifici, storico-culturali e paesaggistici di particolare interesse, è organizzato in modo unitario con le finalità di conservare, tutelare, restaurare, ripristinare e migliorare l'ambiente naturale e le sue risorse, nonché perseguire uno sviluppo sociale, economico e culturale, promuovendo la qualificazione delle condizioni di vita e di lavoro delle comunità residenti. Ulteriore scopo è inoltre, promuovere l'incremento della cultura naturalistica». Per Riserva si intende «un territorio caratterizzato da elevati contenuti naturali ed in cui le finalità di conservazioni dei predetti contenuti sono prevalenti rispetto alle altre finalità indicate per il Parco». Per Area di Reperimento si intende «un territorio che, scelto in base a criteri che hanno come fondamento l'esistenza di contenuti naturalistico-

ambientali indiscutibili, costituisce un bacino rispetto al quale può essere prodotta dal legislatore una singola legge per l'istituzione quale nuova area protetta». Infine per Biotopo si intende «un'area di limitata estensione caratterizzata da una grande concentrazione di habitat ancora integri e che è necessario porre in condizione di tutela per evitare che corrano il rischio di essere danneggiati o distrutti» (Azienda dei parchi e delle foreste regionali, 1999).

10 In realtà il Piano Urbanistico Regionale Generale prevedeva circa 243.000 ettari di aree protette perché ambiva ad un progetto "sistema natura" che con la L.R. 42/96 è stato profondamente ridotto (53.000 ettari circa di aree protette, parchi e riserve, 35.000 ettari circa di aree di reperimento e 900 di biotopi). Proprio in quest'ottica deve essere vista la proposta di una rete ambientale per la regione Friuli Venezia Giulia che ha l'ambizione di costituire una rete infrastrutturale «per creare o ripristinare una connessione tra gli ambienti a maggiore valenza ecologica per mitigare gli effetti della frammentazione attraverso l'individuazione di direttrici ambientali primarie e secondarie» (Bonfanti e Sigura, 2004).

11 Riserve Naturali del Carso: "Laghi di Doberdò e Pietrarossa"; "Falesie di Duino"; "Monte Lanaro"; "Monte Orsario"; "Val Rosandra".

12 Nel caso di Miramare è stato digitato Riserva Marina Miramare.

13 La riserva della Val Rosandra non ha ancora un sito ufficiale.

14 Google è il motore maggiormente "algoritmico", Yahoo quello più "umanizzato", MSN può contare sulle innegabili sinergie legate alla tremenda diffusione del sistema operativo più diffuso al mondo, che notoriamente viene distribuito con integrato un programma di tipo *browser*, per la visualizzazione delle pagine web in rete (si tratta del ben conosciuto Internet Explorer).

BIBLIOGRAFIA

A.A.V.V., *Aree Naturali Protette,* Treviso, Le Tre Venezie, 7, 2003.

AZIENDA DEI PARCHI E DELLE FORESTE REGIONALI, *Aree naturali protette – Parchi, Riserve e Biotopi nel Friuli Venezia Giulia,* Udine, Regione Autonoma Friuli Venezia Giulia, 1999.

BATTELLE J., *Google e gli altri. Come hanno trasformato la nostra cultura e riscritto le regole del business,* Milano, Raffaello Cortina Editore, 2006.

BONFANTI P. e SIGURA M., *Una rete ambientale per la Regione Friuli Venezia Giulia*, in P. Pedrocco e S. Fabbro (a cura di), *Le trasformazioni del territorio in Friuli Venezia Giulia: nuove ipotesi di Analisi e pianificazione,* Udine, 2004, pp. 65-74.

BRIN S. e PAGE L., *The anatomy of a large-scale hypertextual web search engine*, in "Computer Network – Proceedings of the 7th International world wide web Conference", Brisbane, Australia, 1998, 30, pp. 1-7.

CARPINETO C. e ROMANO G., *Verso i motori di ricerca di prossima generazione*, in "Mondo digitale", giugno 2005, 2, pp. 19-31.

FÜRNKRANZ J.,*Web Mining*, Draft book chapter, Technische Universität Darmstadt, Germany, 2004.

INNOCENTI P., *Geografia del turismo,* Roma, Carocci, 2003.

IPK Int., *A World Travel Trends Report (WTTR),* November, 2007.

IPK Int., *Use of the Internet for Travel and Tourism,* European Internet Travel Monitor 2000 (Extract), 2001.

KOSALA R. e BLOCKEEL H., *Web Mining Research: a Survey,* in "ACM SIGKDD", 2000, Vol. 2, Issue 1, pp. 1-15.

LANZA E., MALIZIA W. e DEL CORPO B., *Collaborative on-line network: promuovere il turismo con il web 2.0,* Rapporto sullo Sviluppo sostenibile 6.2007, Fondazione ENI Enrico Mattei, www.feem.it,2007.

LOZATO JOTART J.P., *Geografia del turismo,* Milano, Franco Angeli, 1999.

LUNGHEZZANI G., *Risorse bibliografiche in formato elettronico per le scienze sociali,* Università degli Studi di Trento, www.biblio.unitn.it, 2004.

SEIFERT J. W., CRS REPORT FOR CONGRESS, *Data Mining: an Overview,* The Libray of Congress, US Department, fpc.state.gov, 2004.

VISE D. e MALSEED M., *Google story,* Milano, EGEA, 2005.

ZAIANE O. R., *Principles of Knowledge Discovery in Databases. Introduction to Data Mining,* University of Alberta, Dep. of Computing Science, 1999.