

Introduzione

1. *Comparazione tra diversi approcci storici all'Information Retrieval System (IRS)*

1.1 Metodi del retrieval

1.2 On line database services

1.3 WWW

1.4 Sviluppo storico della logica di interrogazione

1.4.1 Boolean retrieval

1.4.2 Probabilistic retrieval

1.4.3 Ranking

1.5 Quattro generazioni di OPAC

1.6 System oriented studies/User oriented studies

2. *Schema di information retrieval system (IRS)*

2.1 Schema di information retrieval system (IRS)

2.2 Indicizzazione dei documenti

2.3 Tipologie di documenti digitali

2.3.1 On Line Public Access Catalogue

2.3.2 Collezioni di documenti in formato digitale

2.3.3 Collezioni di pagine web

3. *Chi sono, cosa vogliono: identità, preferenze e modalità di ricerca degli utenti di sistemi IR*

3.1 Chi sono

3.2 cosa vogliono

3.3 Come si esprimono

3.4 Gli elementi della strategia di ricerca

3.5 Il piano specifico delle operazioni

4. *L'approccio user oriented nella relazione uomo-macchina. L'analisi di Nahl D.*

Appendice

Bibliografia

**NOTA REDAZIONALE:**  
Questa tesi si compone di 54 pagine

## Introduzione

Il principale obiettivo del presente studio è quello di indagare le modalità attraverso le quali l'utente reperisce l'informazione nell'ambito dei sistemi informatici. Lo studio prende avvio da una panoramica storica relativa allo sviluppo dei metodi di ricerca (capitolo 1). In particolare, verranno presi in esame l'evoluzione storica della relazione uomo-macchina, i fattori di successo dei software di interrogazione delle banche dati, lo sviluppo tecnologico del linguaggio di formulazione della domanda di interrogazione (query).

Nel secondo capitolo verranno presentati i più recenti modelli di sistema di recupero dell'informazione (IRS, Information Retrieval System), mettendone in rilievo le caratteristiche salienti.

Nel terzo capitolo l'analisi si rivolge alla delineazione dell'*identità* degli utenti di questa tecnologia e all'analisi dei loro bisogni informativi. In particolare verrà preso in esame l'approccio *user oriented* proposto da Nahl D.(data) sulla cui base verrà formulato un questionario da somministrare agli utenti della Biblioteca Centrale d'Ateneo dell'Università degli Studi di Milano Bicocca, con l'intento di fornire un contributo alla valutazione del servizio reso nella ricerca di materiale bibliografico dall'OPAC d'Ateneo.

## Capitolo 1

### Comparazione tra diversi approcci storici all'Information Retrieval System (IRS)

Un metodo per raggiungere l'informazione desiderata è cercare le notizie all'interno di una collezione vasta facendole passare ad una ad una (brute force method) (Meadow 2000).

Questo metodo non richiede alcuna preparazione personale, tanto meno una qualche tecnica nella modalità di interrogazione; è efficace in caso di collezioni di documenti ristrette, mentre in caso di collezioni di documenti numericamente rilevanti, si può ovviare ai costi in termini di tempo della ricerca attraverso la creazione di indici o file separati che funzionano da mappe orientative per addentrarsi nella collezione.

Si creano così indici della più svariata natura e si può ritenere che la storia di IRS sia per la gran parte la storia di come creare e trovare nuovi indici.

## *1.1 Metodi del retrieval*

Il primo computer studiato appositamente per IRS è del 1957 ed è sviluppato dall'US Naval Ordnance Test Station di China Lake in California (Bracken-Tillit): utilizza indici su nastro magnetico. Le domande, inoltrate tramite IBM cards sfruttano la logica booleana. Le risposte vengono inoltrate all'utente tramite la carta forata (Meadow 2000).

## *1.2 On line database services*

Il Chemical Abstract Service nel 1961 produce il Chemical Titles, un indice di circa 600 titoli tra i più importanti tra le riviste del settore. Segue l'AIM, Abridged Index Medicus database.

Il più semplice schema per produrre un database è l'analisi umana che porta alla redazione di un indice e di abstracts di documenti. Da questi elementi si ricava l'informazione bibliografica, utilizzata dal software di ricerca per l'interrogazione del database. Si creano così database che funzionano da archivi di records che rappresentano documenti letti dalle macchine.

I database possono essere interrogati da programmi software i quali estraggono records che contengono dati in campi specifici. Nascono così attorno agli anni '70, i software per il recupero dei records dai database: la National Library of Medicine negli Stati Uniti inizia un percorso che porta alla nascita di Medline. Nello stesso periodo un altro software interattivo per l'indicizzazione degli articoli da riviste scientifiche è allo studio presso il Lockheed Palo Alto Research Laboratory. Il software diverrà operativo nel 1971 col nome di Dialog System. Tutti i software prodotti sono proprietari e funzionano per database specifici poi venduti ai centri di ricerca.

I software raggiungono la commerciabilità nel 1972 ed utilizzano la tecnologia cosiddetta di packet switching che consente l'invio di file di peso notevole nella rete telematica senza provocare intasamenti. In sostanza all'invio i file vengono scomposti, segmentati in piccoli puzzle per essere poi ricomposti presso il computer ricevente. Ogni frammento contiene l'indicazione del file e del calcolatore di invio con la posizione che il frammento ha nel puzzle. Ogni frammento è un sottoinsieme omogeneo con informazioni in grado di ricomporre il file. In questo modo il database che si trova in una certa località del globo, può venire letto da tutti coloro che sono collegati dalla rete telefonica. La prima rete commerciale che funziona con tale tecnologia è quindi del 1972 (Tymshare...) e connette computer locali sparsi in varie regioni, collegati dalla rete telefonica che giungono ai dati in modo ragionevolmente veloce. Il sistema sviluppato nel '72 dal Data Central Corporation e successivamente acquisito da Mead Paper Corporation è l'antesignano di Lexis/Nexis, database giuridico, e IRS relativo ai dati giuridici. Fu il primo sistema commerciale per il reperimento del testo completo dei documenti anziché solo dei surrogati del documento.

La tecnologia packet switching è la stessa utilizzata da World Wide Web (WWW).

### 1.3 WWW

La nascita di Internet risale alla fine degli anni '60 e si collega al proposito delle autorità militari statunitensi (Department of Defence) di prevenire un attacco ai centri nevralgici di calcolo americani: un grosso centro di calcolo in una sola località sarebbe stato un bersaglio troppo facile da individuare da parte della potenza sovietica per cui viene proposto dall'Advanced Research Project Agency (Arpa) il progetto di diffondere sul territorio le macchine ad alta tecnologia. La tecnologia utilizzata è quella della scomposizione dei file (packet switching) (Agosti 2002) per cui nel dicembre del 1969 si forma una prima rete sperimentale detta Arpanet con quattro nodi residenti in Los Angeles (University of California LA), Santa Barbara (UCSB), Stanford Research Institute (SRI) e University of Utah. Nel 1969 vi è il primo esperimento di invio di files tra le quattro sedi. Nel 1973 i nodi diventano 23 mentre nel 1974 viene sviluppato il protocollo TCP/IP (Transmission Control Protocol/Internet Protocol) che condensa l'insieme delle regole che consentono al file trasmesso da un calcolatore, di giungere a destinazione integro. In sostanza il software che effettua l'invio da un nodo all'altro della rete (provider) deve rispettare alcune regole di invio (le regole appunto del protocollo TCP/IP). Nel 1983 il protocollo TCP/IP diviene il protocollo per l'interconnessione internazionale, la rete

viene aperta al pubblico con la diffusione della posta elettronica. Nel 1981 la Apple mette in commercio il Personal Computer. La rete riconosce i diversi computer sparsi per il mondo attraverso gli indirizzi dei destinatari. Ogni pc quindi è dotato di un indirizzo specifico e coerente composto da una sequenza di 32 bit (numero IP).

Raggiunto il mercato la rete si diffonde in maniera esponenziale (nei paesi occidentali a capitalismo 'avanzato') e da un 'isolato esperimento' si trasforma in 'icona culturale': per quanto riguarda le risorse informative di tipo elettronico reperibili in www, uno studio compiuto dall'Online Catalogue della Library of Congress Office of Research (OLC) parla di 3.080.000 risorse per il solo web pubblico (corrispondente a circa il 35% del web totale) (O'Neill 2003).

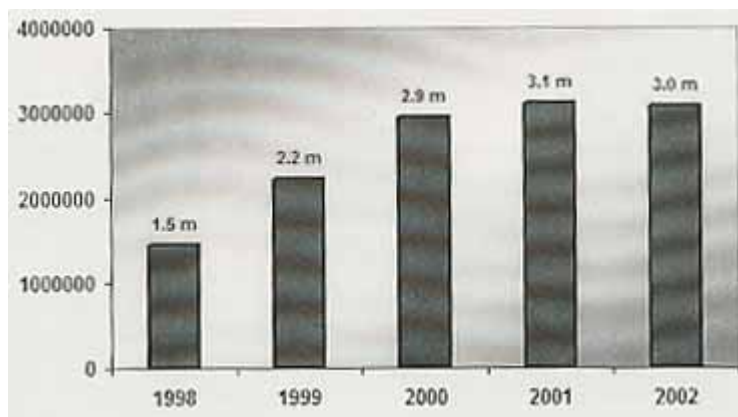


Fig. 1: L'incremento del Public Web dal 1998 al 2002. Tratto da: O'Neill (2003)

Il web pubblico (surface web) viene definito in questo studio una risorsa informativa elettronica che offre agli utenti un accesso senza restrizioni a

significative porzioni del suo contenuto. È la porzione di web più visibile e accessibile dalla maggior parte degli utenti del web.



#### *1.4 Sviluppo storico della logica di interrogazione (Crestani 1998)*

Il problema centrale di IRS è trovare una serie di documenti ritenuti rilevanti da una collezione praticamente infinita e che soddisfano l'esigenza informativa del soggetto espressa attraverso una domanda. Al fine di recuperare un documento l'IRS deve comparare il contenuto semantico della domanda (query) con il contenuto semantico di una serie di documenti. IRS considera rilevante un documento per una query se la query è contenuta nel documento o se la query può essere recuperata in qualche modo nel documento. È opportuno ricordare che il soggetto che pone la domanda ha una valutazione propria di ciò che viene recuperato per cui la valutazione di IRS può essere diversa dalla valutazione fatta dal soggetto che interroga il sistema. Per avvicinare la valutazione del soggetto alla valutazione del sistema sono stati ideati vari linguaggi di interrogazione che vanno dalla logica booleana al linguaggio naturale.

Data la tendenza di IRS di interpretare la query con la logica booleana, con algoritmi probabilistici o con il linguaggio naturale, oltre a verificarsi una discrepanza tra valutazione dei documenti recuperati da parte del soggetto che formula la query e IRS, può accadere anche che la formulazione della query, la sua grammatica, venga ritenuta errata per cui il soggetto deve ristrutturarla.

Il documento è rilevante se il soggetto che formula la query lo ritiene tale. A partire da tale presupposto, si sono sviluppati sistemi di retrieval orientati sempre più verso la valorizzazione della semantica dell'interrogazione.

#### *1.4.1 Boolean retrieval*

Nella logica booleana si assume che i documenti vengano rappresentati da indici o da parole chiave e che la richiesta sia una logica combinazione di questi termini con l'utilizzo di AND (entrambi i termini devono essere presenti), OR (almeno uno dei termini deve essere presente), NOT (il termine non deve essere presente) (Agosti 2002). Il modello booleano ha il vantaggio di essere "efficace in ambienti controllati e con utenti bene addestrati " ma per contro vi è "poco controllo sul numero dei documenti recuperati"; è "impossibile l'ordinamento per una qualche misura di similarità; impossibile la pesatura dei termini" e manifesta "limitazioni dovute alla bassa amichevolezza della logica booleana". Inoltre "richiede agli utenti di sapere precisamente cosa cercano". In tale logica un documento è ritenuto rilevante da IRS e quindi recuperato se i termini indice sono elementi semantici e quindi se il documento ha subito quella mappatura necessaria che porta alla sua evidenza semantica.

Questo discorso si mostra in tutta la sua chiarezza se si tengono in considerazione i dati del Dublin Core Metadata Initiative che fanno riferimento ad uno standard per l'indicizzazione dei siti web (O'Neill 2003).

Nel 2002 solo lo 0,7% dei siti pubblicati facevano riferimento allo standard internazionale; la stragrande maggioranza dei

**Table 1: Metadata Usage on Public Web Sites, 1998 - 2002**

	1998	1999	2000	2001	2002
<b>No. of Public Sites</b>	1,457,000	2,229,000	2,942,000	3,119,000	3,080,000
<b>% of Sites with Metadata on Home Page</b>	70 percent	79 percent	82 percent	85 percent	87 percent
<b>Mean No. of Tags</b>	2.75	2.32	2.72	2.97	3.14
<hr/>					
<b>% of pages with Metadata</b>	45 percent	50 percent	59 percent	62 percent	70 percent
<b>Mean No. of Tags</b>	2.27	2.16	2.46	2.63	2.75

Fig. 2: utilizzo dei metatag nella struttura html di un sito web pubblico. Tratta da: O'Neill (2003)

siti venivano

quindi

pubblicati

senza

indicizzazion

e (metatag),

oppure con

l'indicizzazi

one automatica dell'editor html di pubblicazione, quindi non strutturati in alcun modo, senza regole, specifiche o schemi di qualche valenza.

Si comprende quindi quale sia la difficoltà di reperire informazioni utili in rete in quanto è su tale processo di indicizzazione che i vari motori di ricerca fanno riferimento qualora interrogati dall'utenza.

Il modello di interrogazione booleano non tiene conto della rilevanza.

#### *1.4.2 Probabilistic Retrieval*

Due utenti con diverse idee di rilevanza e valutazione dei documenti, pur immettendo la stessa query nel form di interrogazione, si aspettano diversi risultati.

Come separare il giudizio soggettivo circa la rilevanza dei documenti dalla loro implicazione logica prevista da IRS per cui un documento è ritenuto tanto più pertinente tanto più la query è rinvenuta frequentemente in un documento (meglio: negli indici che rappresentano il documento)?.

Poniamo che  $S$  sia la descrizione parziale di un documento che può essere una serie di frasi o un termine-indice presente nei metatag della risorsa informativa elettronica, mentre  $Q$  la domanda o query formulata dall'utente.

Se  $S$  è vero in un documento  $D$ , allora  $S$  verso  $Q$  è vero. Se  $S$  non è vero in un documento  $D$  allora il sistema passa ad esaminare il documento  $D_1$  e riconsidera la proposizione ( $S$  è vero in  $D_1$ ?). Nel caso  $S$  sia vero, allora  $S$  è vero anche nei documenti dove  $Q$  occorre.

È questo il caso dei descrittori. Se per esempio l'utente immette in un sistema di retrieval la query ( $Q$ ): Programming Language (Crestani 1998), per cui è previsto il descrittore ( $S$ ) Fortran, siamo nel caso in cui  $S$  è diverso da  $Q$  a livello semantico per cui il documento può essere recuperato solo per via indiretta.

La differenza di questo procedimento con la logica booleana è che il documento non verrebbe recuperato nemmeno per via indiretta. Nella logica booleana se un documento contiene determinati termini-indice (S), noi recuperiamo quei documenti che implicano la domanda che noi abbiamo composto. Per cui se S=Fortran e Q=Programming Language, D (contenente S=Programming Language) non viene recuperato. Nella logica probabilistica quindi viene introdotto in IRS un meccanismo che calcola la relazione di vicinanza tra la domanda e il documento e tra i documenti tra di loro. Il risultato di questo calcolo può essere utilizzato per confrontare la rilevanza o pertinenza tra i documenti recuperati e la domanda formulata dall'utente. Il che significa introdurre una discriminante tra i documenti con la speranza che il giudizio che il sistema utilizza per recuperare i documenti sia il più vicino possibile al giudizio che formula l'utente che ha introdotto la query di ricerca.

Nella logica booleana il giudizio di vicinanza tra documenti recuperati e soddisfazione dell'utente (Retrieval Status Value, RSV), è pari a zero o ad uno, a seconda se il documento sia recuperato o meno (S=Fortran, Q=Programming Language: RSV=0).

Nella logica probabilistica RSV oscilla tra zero e uno (S=Fortran, Q=programming Language: RSV= 0, 0,1, 0,2, ... 1).

### *1.4.3 Ranking (Ordinamento)*

Una volta stabilito che RSV valuta la corrispondenza all'interno di IRS, tra documento (S) e domanda (Q) il passo successivo è allineare i documenti in ordine decrescente rispetto alla loro percentuale di pertinenza RSV.

Il sistema di retrieval giunge così alla redazione di una lista di documenti che pone ai primi posti quelli con un maggior RSV e che quindi dovrebbero soddisfare maggiormente le richieste dell'utente.

Ogni documento assume così un peso per cui siamo nella nuova situazione (rispetto alla logica booleana) di avere risposte non solo circa l'esistenza o meno di un determinato documento, ma di valutare quale dei documenti esistenti più risponde maggiormente alle nostre esigenze informative.

### *1.5 Quattro generazioni di OPAC (Fattahi 1997)*

La prima generazione di cataloghi in linea è la traduzione compiuta dalla macchina, delle schede catalografiche cartacee con le loro caratteristiche tradizionali.

In contrasto con le aspettative degli utenti, il loro impiego nelle biblioteche non sfrutta la potenzialità dei database, per cui non apportano sostanziali vantaggi rispetto al catalogo cartaceo. Il recupero delle informazioni da questi OPAC è possibile solo introducendo nel form, l'esatto termine (autore, titolo...) o frase di ricerca nell'esatta forma di linguaggio comprensibile dalla macchina. Ciò a causa delle limitate capacità raggiunte dalla tecnica di indicizzazione che ancora non sono in grado di produrre un vocabolario controllato dei termini.

Il risultato per l'utente è meno importante che non con l'interrogazione del catalogo cartaceo in quanto vi è grossa difficoltà, anche di memoria, ad introdurre la forma esatta dell'interrogazione, e ad ottenere quindi una risposta sensata dalla macchina. La ricerca non è quindi così "magica" come l'utente si aspetta rispetto all'interrogazione del catalogo cartaceo. Non sono disponibili ricerche per 'parola chiave'; non è possibile raffinare la ricerca limitando la 'query' ad un determinato linguaggio, data di pubblicazione dell'opera,... L'interfaccia di questi OPAC ricalca la scheda catalografica tradizionale; le ricerche per soggetto, autore, titolo o altro

non comunicano tra loro e fanno riferimento ad indici separati.

Lo sviluppo dell'informatica si fa sentire con la seconda generazione di OPAC in quanto tale scienza rende possibile la configurazione di un sistema più sofisticato di recupero dell'informazione bibliografica. L'OPAC si caratterizza ora, per la distanza dalla tradizionale scheda catalografica e per l'inclusione di elementi informatici che consentono un effettivo accesso all'informazione come ad esempio la ricerca per 'parola chiave', la ricerca con operatori booleani, la ricerca incrociata con diversi indici (es: autore e titolo). Viene riconosciuta la possibilità di raffinare la ricerca per linguaggio e data di pubblicazione del materiale catalogato. Vi è quindi un effettivo incontro tra le esigenze delle biblioteche e le tecniche di interrogazione utilizzate dai sistemi di information retrieval. Con l'aggiunta dei metodi di ricerca combinata, si offre all'utente della biblioteca una possibilità non disponibile nella prima generazione di cataloghi in linea. L'aggiornamento informatico della gestione delle banche dati attraverso software di ricerca, rende possibile alla seconda generazione di OPAC di arricchire le notizie bibliografiche con sommari, note, abstracts e links alla risorsa elettronica full text. L'interfaccia è resa più maneggevole dalla presenza di strumenti di assistenza in linea come messaggi di errore, help on line.



Gli OPAC di terza generazione aggiungono rispetto ai precedenti, la possibilità di compiere interrogazioni con linguaggio naturale..

A partire dagli anni Ottanta si sviluppano le WIMP (windows, icons, mouse and pointers) interfaces (graphical users interfaces, GUIs) che semplificano e rendono la ricerca molto più veloce (quarta generazione di OPAC).

Con queste interfacce, sviluppate soprattutto dal monopolista Windows, l'utente ha la possibilità di accedere all'informazione bibliografica da punti diversi dell'interfaccia, ognuno dei quali attiva una diversa funzione. Questi sistemi aumentano l'accesso per 'parola chiave'. Viene implementato l'accesso via mouse e tastiera. Le possibilità di ricerca nelle versioni Windows degli OPAC sono di gran lunga superiori rispetto alle altre generazioni di cataloghi in linea.

Con il linguaggio post-booleano, il sistema è in grado di interpretare le richieste dell'utente e allinearle nella pagina a partire da quelle ritenute di maggior interesse (relevance ranking). Viene introdotta la funzione ipertesto in base alla quale l'utente digitando il termine accede ai richiami correlati trovati all'interno del database. In questo modo l'utente può navigare per la base di dati e raggiungere nuove fonti di informazione.

### *1.6 System oriented studies/User oriented studies*

Nahl D. (1996), delinea quali sono i due diversi paradigmi di approccio alla ricerca che pone ad oggetto la relazione tra l'uomo e i sistemi informatici. Secondo un primo approccio denominato *system oriented studies*, il comportamento degli utenti, le loro risposte alle sollecitazioni del sistema informatico, sono considerate come situazioni estranee al processo di miglioramento dei software di interrogazione. Questo approccio non tiene conto delle modificazioni che i sistemi informatici hanno prodotto nell'uomo e viceversa, ed assume che il comportamento umano è lo stesso nel tempo e nello spazio; traslascia variabili come il sesso, l'età, la classe sociale di appartenenza... A questo orientamento di ricerca si associa la formalizzazione di una astratta e generica classe di utenti inesistente cui il sistema di IRS fornisce dati e documenti.

La figura di utente assume in questi casi i tratti paradigmatici di utente statico e acritico. La conclusione cui i ricercatori di questo paradigma giungono è che l'utente, con quanti più documenti gli si mostrano davanti, con più risulta felice e soddisfatto. Con la revisione degli studi di Nilan e Dervin (1986) (Palmquist 1998), si giunge alla formulazione di un nuovo paradigma di ricerca negli studi che hanno ad oggetto la relazione uomo-macchina.

Si affermano gli "user oriented studies" che riconoscono invece il ruolo sociale dell'utente e le sue caratteristiche psicologiche. I valori dell'informazione non sono quelli della macchina, ma quelli che gli utenti percepiscono. L'utente è visto ed analizzato durante il processo di recupero dell'informazione in una situazione di cambiamento dinamico.

La ricerca stessa viene articolata in più fasi, ognuna delle quali con un suo specifico obiettivo e funzione psicologica e motivazionale (sub-goal). Ogni percezione e motivazione dell'utente può dunque assumere valore analitico e venire interpretata come contributo al cambiamento della struttura generale e funzionale di IRS.

## Capitolo 2

### Schema di information retrieval system (IRS)

#### 2.1 Schema di Information Retrieval System (IRS)

Secondo un recente modello di IRS (vedi fig. 1) (Agosti 2002)



Fig.3: Schema di IRS. Tratto da: Agosti (2002)

l'utente recupera un insieme di documenti attraverso la formulazione di una richiesta (query) immessa nel form che dà accesso ad un database (sia esso una collezione di pagine web come google, un

opac, una banca dati, una collezione di documenti digitali). Qualora i documenti recuperati dal sistema di information retrieval e proposti all'utente, vengano considerati rispondenti alle richieste, l'utente abbandona il sistema e considera la sua ricerca conclusa positivamente.

Nel caso in cui l'utente si consideri insoddisfatto per i risultati ottenuti (ASK, anomal state of knowledge) a causa della non rispondenza dei documenti alle effettive esigenze

informative dell'utente, ha la possibilità di formulare una nuova richiesta o di rivedere la precedente ricerca in base ai documenti recuperati.

La prima parte del processo ad U consente di individuare le fasi di interrogazione, acquisizione e valutazione dell'informazione da parte delle più svariate categorie di utenti.

Alcuni specialisti (Capra 2003) si sono chiesti come gli utenti

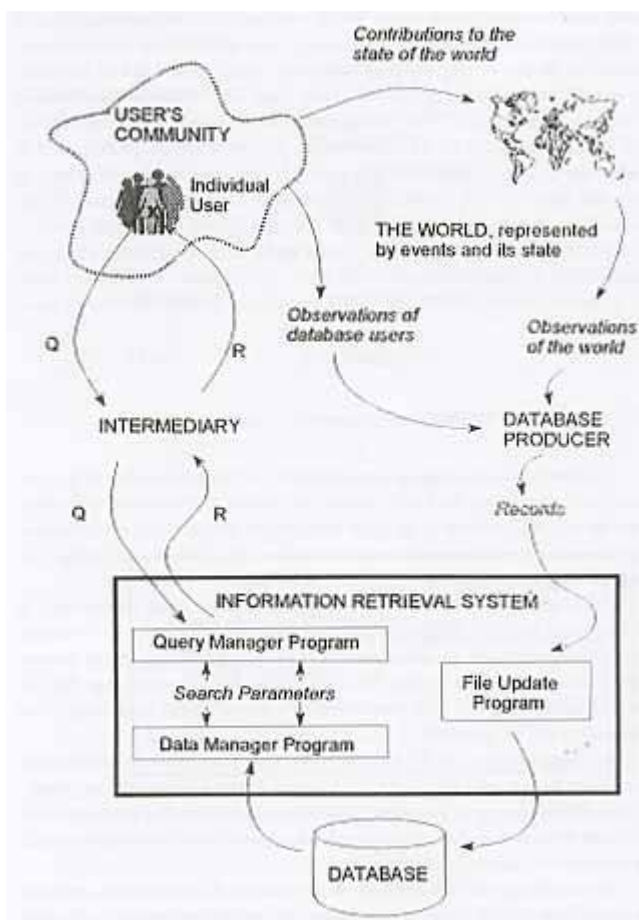


Fig. 4: Schema di IRS. Tratto da: Meadow (2000).

organizzano e integrano l'informazione recuperata. Infatti accade spesso che gli utenti del web (quindi anche degli opac, delle banche dati elettroniche o dei giornali elettronici) recuperino innumerevoli informazioni ma abbiano difficoltà nell'organizzare e integrare l'informazione acquisita. Effettivamente una considerevole mole di studi viene compiuta al fine di aiutare l'utenza a

recuperare l'informazione nel web.

## 2.2 Indicizzazione dei documenti.

A prima vista la formulazione della richiesta di informazione espressa in linguaggio naturale da un utente può apparire in conflitto con quanto l'informatica concepisce come definizione di informazione (insieme di bit).

Per informazione l'informatica intende l'"insieme dei dati forniti dall'ambiente esterno a un essere vivente, o a una macchina, espressi in forma comprensibile e significativa in un particolare contesto" (Agosti 2002).

L'informatica (contrazione di information-automatique) concepisce l'informazione in termini quantitativi, come insieme di bit (contrazione di binary-digit), l'unità minima

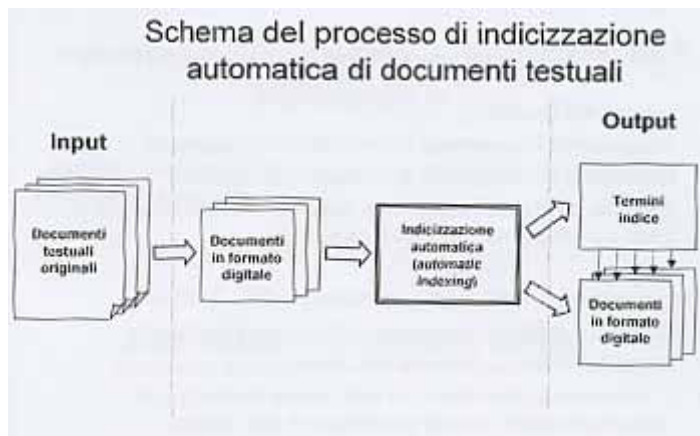


Fig. 5: da documenti a termini indice.  
Tratto da Agosti (2003)

di informazione. Dalla definizione di informazione si comprende quindi come la macchina è concepita in maniera da comprendere ciò che le viene ordinato per il tramite di un software che traduce per essa le

regole di comunicazione del linguaggio umano (naturale). L'informazione riguarda il modo in cui un sistema acquisisce conoscenza dal mondo esterno.

Un qualsiasi dato acquista valore solo in relazione ad altre informazioni che ne definiscono il contesto (20021114 può a seconda significare nulla, una data, la collocazione di un libro in una biblioteca...).

Il conflitto tra le esigenze dell'utente e il linguaggio della macchina è mediato dal processo di indicizzazione automatica dei documenti testuali "che è il processo che esamina automaticamente gli oggetti informativi che compongono il documento e, utilizzando degli algoritmi appositi, produce una lista di indici che vengono utilizzati come surrogati per la rappresentazione del documento originale e che possono quindi essere utilizzati al suo posto durante la fase di recupero" (Agosti 2002). In un contesto di IRS anche la frase immessa dall'utente nel form di richiesta viene indicizzata. Gli indici relativi alla frase dell'utente vengono confrontati con gli indici interni del sistema e quelli simili vengono proposti in risposta all'utente. Il confronto avviene dunque tra la domanda e gli indici, non tra la domanda e i documenti. Questi ultimi sono sottoposti al processo di indicizzazione che ne consente la reperibilità.

Per il tramite di un software si estraggono automaticamente dal documento le parole ritenute più ricche dal punto di vista semantico al fine di rappresentare il documento.

Il processo di indicizzazione è quindi un processo complesso e si compone di diverse fasi: analisi lessicale e selezione delle parole, rimozione delle parole molto comuni (stop

words), riduzione delle parole originali alle rispettive radici semantiche (stemming), creazione dell'indice, eventuale pesatura degli elementi dell'indice.

E' importante sottolineare che l'indicizzazione non avviene mai sul documento originario digitale ma sempre sul surrogato del documento. Colui che esprime l'esigenza informativa in IRS non giunge mai ai documenti digitali originari ma raggiunge per lo più una lista di parole estratte a mò di indice dai testi digitali originari.



### *2.3 Tipologie di documenti digitali.*

#### *2.3.1 Online Public Access Catalogue (OPAC).*

Sono collezioni di documenti digitali che descrivono i contenuti di collezioni fisiche di documenti.

Negli opac si trovano le notizie bibliografiche e le descrizioni catalografiche formulate secondo determinate regole biblioteconomiche, che servono per raggiungere in maniera univoca la risorse bibliografiche presenti in una biblioteca.

In questo caso il documento digitale è un elemento che rinvia a qualche cosa d'altro.

#### *2.3.2 Collezioni di documenti in formato digitale.*

In questo caso (ad esempio <http://gutenberg.net>) il documento oggetto della ricerca viene rinvenuto all'interno del sistema in quanto l'indicizzazione si riferisce ad un documento totalmente recuperato in formato digitale.

### 2.3.3 Collezioni di pagine web

Sono i Search Engine (SE) come google, lycos, Virgilio... In questo caso l'utente che pone una domanda ottiene in risposta un elenco di link a pagine web rinvenute dal SE e ritenute pertinenti in base alla domanda formulata. Gli indici che un SE utilizza per rispondere alle domande dell'utente sono generati e mantenuti in modo autonomo e automatico.

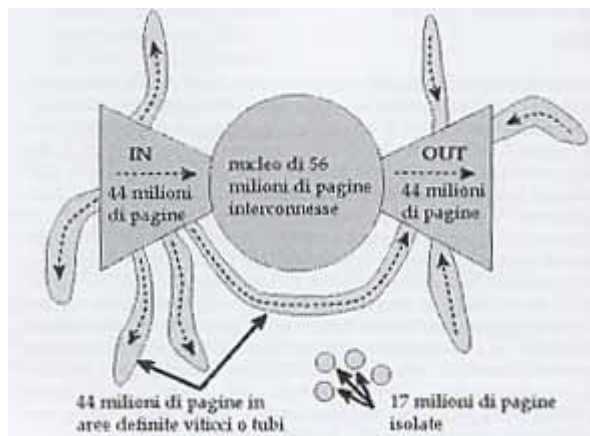


Fig. 6: Schema a farfalla. Tratto da: Gambari (2003)

## Capitolo 3

### **Chi sono, cosa vogliono: identità, preferenze e modalità di ricerca degli utenti di sistemi IR**

Chi si cela sotto la parola utente? E di quale bisogno è portatore?. Una volta delineato per sommi capi il processo che porta al recupero dell'informazione, occorre individuare il soggetto che interroga la rete e cosa vuole in cambio dall'IRS.

### 3.1 Chi sono

I soggetti vengono generalmente suddivisi in tre classi principali (Meadow 2000):

- 1) gli information specialists e i reference librarians,
- 2) i professionisti di una certa disciplina come fisici, chimici, avvocati, docenti universitari...
- 3) gli utenti inesperti (novices end users).

Con le prime due categorie si identificano gli utenti esperti, "in grado di definire esaurientemente le proprie esigenze informative" anche attraverso l'utilizzo di "linguaggi avanzati nelle interrogazioni". Con la terza categoria si identificano gli utenti casuali che non conoscono esattamente cosa stanno cercando, formulano "interrogazioni in maniera generica affidandosi alle potenzialità del sistema di recupero".

Più in dettaglio, gli information specialists (specialisti del settore informativo) sono definiti utenti che assistono altri utenti a trovare le informazioni desiderate in rete

Queste figure non creano informazione per i loro clienti ma sono specializzati nel trovare l'informazione e quindi nell'interrogare e nell'interpretare diversi sistemi di IRS a seconda delle richieste. Sono considerati intermediari tra gli utenti finali e il sistema di IR per cui devono conoscere tutta la vasta gamma di informazione prodotta a diverso livello in una disciplina o specialità presente in una

biblioteca o in altra istituzione nella quale lavorano. Devono conoscere la meccanica di IRS per essere in grado di raggiungere l'informazione contenuta in un database di un CD ROM oppure in una risorsa elettronica remota (RER). Devono inoltre conoscere la terminologia relativa alla disciplina con cui hanno a che fare (medicina, giurisprudenza, chimica...) e il metodo per interrogare l'utente finale in modo da conoscere le esatte informazioni necessarie per iniziare la ricerca (ovvero l'esatta espressione del loro anomal state of knowledge - ASK).

L'information specialist è una figura che si afferma con la diffusione capillare dei personal computer nel mercato statunitense prima, euro-giapponese poi.

A partire dagli anni '80 inoltre, le aziende produttrici di software di Information Retrieval, sono in cerca di nuovi consumatori per i loro prodotti (biblioteche ed agenzie governative erano già state raggiunte) e trovarono negli utenti privati (end users) possessori del personal computer, la possibilità di espandere enormemente le vendite.

Tipicamente un end user è un membro di una professione altamente specializzata (sono medici, avvocati, chimici, fisici,...) e necessita necessariamente di aggiornamento informativo costante. Vi sono figure in ambiente professionale o all'interno di gruppi aziendali appositamente dedite all'acquisizione e all'interpretazione dell'informazione.

A differenza dei references librarians, non è richiesto a costoro la conoscenza del meccanismo di IRS.

Ci sono infine gli utenti inesperti, rappresentati da studenti delle scuole superiori e universitari.

### 3.2 Cosa vogliono

Per molto tempo IRS ha riguardato una cerchia ristretta di ricercatori, l'impatto sul mercato delle loro ricerche si è rivelato poco incisivo sino agli anni '80.

L'interrogazione dei database si basava su standard booleani e vi era scarsa attenzione al risultato della ricerca, al processo di formulazione della domanda, alla pesatura dei termini-indice, al ranking (ordinamento dei risultati) e al feedback formulazione della risposta).

Ad oggi la situazione è del tutto cambiata e molte delle tecniche di interrogazione di software di IRS una volta ritenute inaccessibili o esoteriche, sono divenute di dominio comune.

Le esigenze degli utenti attuali sono le seguenti (Croft 1995):

□ **Relevance feedback:** è il processo che consente agli utenti di identificare documenti rilevanti da una iniziale lista di documenti recuperati. Per fare questo sono stati creati nuovi algoritmi per l'automatica identificazione di documenti rilevanti all'interno dei documenti recuperati.

□ **Information extraction:** implementazione delle tecniche di estrazione dell'informazione da database. Oggi è possibile estrarre da un database semplici categorie di informazioni come nomi di ditte, nomi di persone, autore e titolo di un libro...

□ **Multimedia retrieval:** si vuole avere accesso non solo all'informazione testuale ma anche a video, suoni, immagini... In questo settore investono molte ditte di servizi che hanno notevole interesse a sviluppare queste tecniche.

□ **Effective retrieval:** è considerato il cuore del IRS, è la valutazione della precisione nel recupero dell'informazione. Abbiamo visto i progressi che hanno portato le tecniche di interrogazione dal linguaggio booleano a quello naturale. Le ditte che vendono ed utilizzano sistemi di IRS sono interessate nella precisione e nel recupero dell'informazione: avere un motore di ricerca efficiente è un punto in più rispetto alla concorrenza da un lato, e per gli utenti finali è un notevole risparmio di risorse.

□ **Magic:** una delle cause principali del fallimento di un IRS è un vocabolario ambiguo: l'informazione di cui si ha bisogno viene filtrata come abbiamo visto dai descrittori o termini-indice che rimangono per lo più sconosciuti al vasto pubblico. Tecniche che raggiungono documenti che, a prima vista, non contengono la parola chiave introdotta nel form di interrogazione, appaiono magiche e sono vissute come altamente desiderabili. Oltre al ricorso alla logica probabilistica, risultati del genere si possono ottenere con il Latent Semantic Indexing.

□ **Efficienza:** riguarda sostanzialmente il tempo di risposta del sistema alla domanda introdotta dall'utente.



□ Integrazione tra i diversi sistemi: IRS è un insieme di soluzioni parziali a problemi parziali di organizzazioni specifiche. Lavori per lo standard dell'architettura e di una comune piattaforma sono largamente auspicabili.

### 3.3 Come si esprimono (Hsieh-Yee 1998)

Le modalità con le quali le diverse tipologie di utenti possono interrogare i sistemi di information retrieval (search strategy) si differenziano a seconda degli elementi bibliografici conosciuti dall'utente che interroga il sistema (oltre che al grado di esperienza maturata) "Occorre una effettiva capacità di condurre strategie di ricerca impostate secondo tecniche di volta in volta adattabili e proprie di quel determinato strumento informativo che si va ad utilizzare; è indispensabile saper operare una selezione di strumenti e di metodologie" (De Robbio 1999).

□ Nel caso si conosca chi o cosa si sta cercando (know item search), chi interroga il sistema sa esattamente quali records devono essere trovati e nel caso venga rinvenuto il surrogato del documento (in opac o banche dati) o il documento (digital library), questo è riconosciuto come oggetto specifico della ricerca.

□ Si conosce chi o che cosa si sta cercando solo dopo averlo visto e riconosciuto (specific information search): in questo caso si cercano specifiche informazioni, non necessariamente specifici records.

□ L'utente cerca qualcosa qualcosa di generico non di specifico (general information search). Si cercano in rete informazioni su un soggetto in generale come ad esempio: come

comprare un'azienda o un metodo per risolvere un'equazione differenziale. Probabilmente l'informazione si troverà più di un record. È possibile che molte notizie non vengano considerate utili anche se recuperate.

□ Non si sta cercando qualcosa di specifico (exploration the database): oggetto della ricerca in questo caso, è indagare quali notizie possono essere recuperate nel database (browsing), quindi non si interroga il database perché questo risponda ad una specifica domanda.

Un IRS deve considerare che l'utente, nei vari momenti di cui si compone una ricerca, passa dall'una modalità all'altra nel mentre della sessione. Quindi non si deve immaginare che l'utente si approcci al IRS con un percorso prefigurato in mente, come se le varie modalità di ricerca appartenessero a comparti separati della memoria.

### *3.4 Gli elementi della strategia di ricerca*

Una ricerca, in un modo o nell'altro, per raggiungere un obiettivo, deve fare i conti con limitate risorse di tempo, denaro e informazione. Un piano o una strategia di ricerca, deve fare in modo che, date queste limitazioni, l'obiettivo venga comunque acquisito.

La strategia si compone quindi di un oggetto della ricerca, di un piano generale e di un piano specifico delle operazioni.

Occorre chiaramente identificare l'oggetto se si vuole procedere nella ricerca, altrimenti, se non si sa cosa si vuole o non si sa dove si è e dove si vuole arrivare, si rischia di lasciarsi prendere dal fascino delle notizie e perdere il filo della rete.

L'oggetto della ricerca può essere identificato come tipo e quantità di informazione desiderata.

Occorre considerare le risorse disponibili come il tempo della ricerca,..., e formulare una mappa mentale per pianificare le formulazioni delle varie query di interrogazione, anziché estrarre dal database ciò che è disponibile e poi costruirsi un piano mentale a posteriori.

### 3.5 Il piano specifico delle operazioni

Riguarda lo specifico materiale che si intende consultare per estrarre le informazioni desiderate. Interrogazioni presentate a Google non possono raggiungere gli stessi risultati di interrogazioni presentate in un database CSA (Cambridge Scientific Abstract), oppure in un Opac di una biblioteca.

Esistono quindi specifiche modalità per interrogare il web, specifiche modalità per interrogare database e opac...

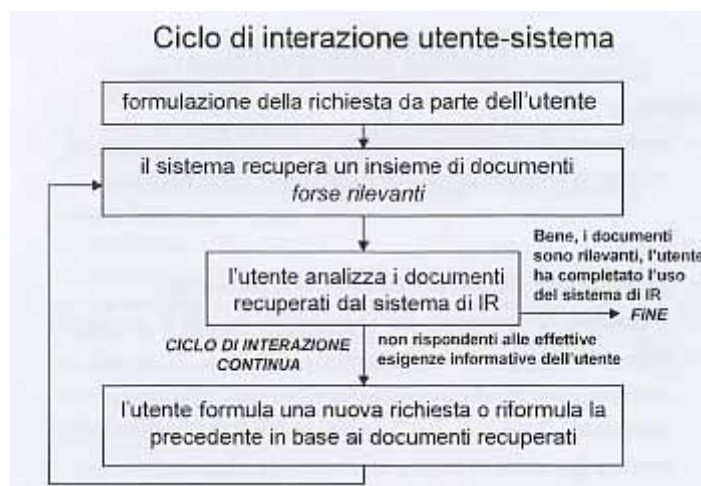


Fig. 7: interazione utente-sistema. Tratta da: Agosti (2003)

Per quanto riguarda la ricerca in web (search engine), si deve tener conto della particolare struttura della rete per cui si avranno buoni risultati sfruttando la gerarchia dei domini (parent directories) e i link (browsing). Quando nulla è trovato circa un

soggetto, si ricorre ad un altro motore di ricerca.

Le tattiche di ricerca nei database oppure negli opac, si avvalgono dei termini indice, delle parole chiave, dei soggetti per cui si ottiene un numero elevato di records su un determinato soggetto e poi si analizzano i risultati, oppure si raffina la ricerca digitando un altro termine o parola

chiave. Quando nulla è trovato circa un soggetto, si ricorre ad altra banca dati, ad altro opac oppure si chiede al bibliotecario.

Introdurre nella query un termine più generico o specifico, cercare in soggetti relazionati e utilizzare sinonimi, sono invece tattiche comuni ad opac, database e web.

Si conclude quindi che il web ha mantenuto le tecniche di ricerca specifiche di IRS quali opac e banche dati (broader, narrower, related, synonymous) (Hsieh-Yee 1998)., aggiungendo i canali interrogativi tipici della sua struttura. Vi sono da ribadire qui, le riserve circa l'indicizzazione del web per cui la stragrande maggioranza delle risorse informative elettroniche remote seguono tecniche di indicizzazione che non vengono riconosciute dai motori di ricerca e rimangono quindi nascoste.

## Capitolo 4

### **L' approccio *user oriented* nella relazione uomo-macchina.**

#### **L'analisi di Nahl D (Nahl 1996).**

Nahl D. propone una ricerca che tende a meglio comprendere l'interazione tra gli utenti e i sistemi di IR, partendo dai bisogni degli utenti espressi sottoforma di domande durante le sessioni di ricerca. Gli utenti, principalmente i nuovi utenti dei sistemi di IR, pongono un'infinità di domande se ne hanno la possibilità. Capire questa complessa interazione e questi bisogni informativi, aiuterà gli operatori della biblioteca a sviluppare migliori sistemi di ricerca e a meglio approntare metodi di ricerca per gli utenti finali. Negli studi di Nahl, viene analizzato l'approccio alla ricerca di un articolo in un full-text database, da parte di studenti universitari di facoltà umanistiche. Durante la sessione di ricerca, agli studenti viene chiesto di pensare ad alta voce in modo da registrare su nastro magnetico domande, considerazioni, perplessità...

La combinazione tra i dati empirici recuperati in questo modo e le strategie di ricerca registrate nei downloads dei risultati delle ricerche, portano l'autrice a formulare l'ipotesi che, di fronte alla macchina si attivano nell'utente

*domande* che chiamano in causa tre abilità psicologiche ben specifiche: sensomotorie, cognitive e affettive.

La domanda è un mezzo per trovare nuove informazioni e si relaziona ad una esistente mappa mentale: può essere considerata un indice spaziale, esterno di un processo interno di tipo cognitivo e affettivo.

L'assunto di base è che il bisogno informativo genera domande che sono atti comportamentali con i quali si intende soddisfare il bisogno.

L'aspetto affettivo riguarda le motivazioni, le intenzioni e gli scopi che spingono l'utente ad entrare in un processo di IR. Oltre ad essere un background psicosociale viene continuamente sollecitato dalla macchina, la quale richiede abilità sensomotorie specifiche (saper utilizzare il mouse, riempire un form, interpretare i messaggi pop-up, interrogare l'help on line...) che innescano una dinamica emozionale la quale si ripercuote sulla qualità e il successo della ricerca. Le risposte emotive (frustrazione, ostinazione, speranza, disappunto, eccitazione, disincanto...) alle sollecitazioni del sistema informativo, coinvolgono l'intero mondo motivazionale ed emozionale dell'utente. Occorre distinguere l'intenzione nella ricerca dai risultati ottenuti; giudicare i risultati della ricerca in base alla soddisfazione espressa dall'utente; assecondare gli sviluppi della ricerca che procede per piccoli passi (sub-goal) e che giunge alla completa o parziale soddisfazione passando continuamente tra ansietà ed euforia.



L'aspetto affettivo dell'utente aiuta a determinare questo movimento dinamico tra risultati parziali.

Le domande che gli utenti pongono spontaneamente, identificano il luogo e il contesto di un'acquisizione o di un cambiamento di strategia durante il processo di ricerca; identificano quindi i luoghi di IRS, dove occorre intervenire tecnicamente per migliorare l'interazione uomo-macchina.

Vi sono domande poste al supervisore durante l'esperimento che indicano lo stato di smarrimento dell'utente giunto ad un punto della ricerca e che intende essere assicurato circa la corretta direzione da lui intrapresa nel percorso. Bisogna immaginare cosa succede in situazioni normali, dove l'utente non ha alcun supporto tecnico o emotivo cui poter rivolgere domande (oltre all'help on line).

L'aspetto affettivo quindi è essenziale per giudicare il successo della ricerca.

L'abilità cognitiva interviene nella comprensione dei concetti mentre la sensomotoria nella pratica col sistema.

Le domande denotano non solo uno sviluppo razionale della ricerca ma anche la dinamica dell'ambientamento che è indice di esitazioni, desideri di conferma, di fuga, speranza... tutte attività che interagiscono con il patrimonio cognitivo dell'utente e determinano il dominio sensomotorio della ricerca.

Può accadere che motivazioni emotive spingano ad abbandonare la ricerca prematuramente senza ottenere i risultati sperati

quando si è ad un passo del successo semplicemente modificando l'impostazione della domanda o della strategia. Il che significa che la sessione di ricerca non procede per concettualizzazioni ma per micro informazioni e che esistono una serie di micro informazioni organizzate gerarchicamente che portano a recuperare l'informazione di cui si ha bisogno. Le domande del dominio affettivo interagiscono col bisogno dell'utente e con la sua situazione cognitiva per progredire nelle decisioni da prendere durante la ricerca. Se queste motivazioni vengono a mancare, la ricerca viene abbandonata. Durante l'esperimento, il supervisore interagisce con gli studenti al fine di evitare di giungere a questo punto. È da rimarcare il fatto che durante le ultime sessioni di ricerca, le domande di dominio affettivo subiscono un drastico decremento.

Le domande di dominio cognitivo caratterizzano l'abilità dell'utente di comprendere i risultati e selezionare i vari passi della ricerca.

Le domande di dominio sensomotorio caratterizzano la necessità di tradurre le intenzioni di procedere in un senso o in un altro nella sessione di ricerca, nella corretta grammatica e sintassi utilizzata dalla macchina.

Lo studio evidenzia che il 46% delle domande nelle prime sessioni di ricerca ineriscono al campo affettivo (interazione sociale, sorpresa per insperati risultati raggiunti, richiesta di conferma circa la procedura di ricerca

pianificata...)). Tra le domande di dominio affettivo, la preminenza spetta a richieste di conferma circa la procedura di ricerca pianificata che rinviano al bisogno dell'utente di essere assistito da continue motivazioni per procedere nella ricerca. Le richieste assumono varie forme come il desiderio di vedere riconosciuto il piano di ricerca impostato, l'assicurazione che ciò che si sta facendo è la cosa giusta, la richiesta di supporto morale circa una decisione.

La caratteristica specifica delle domande di dominio affettivo è che la risposta che dovrebbe essere fornita dal supervisore, è in realtà già contenuta nella domanda posta. Il supervisore quindi non può che rispondere affermativamente.

Per questo motivo le domande di dominio affettivo si differenziano dalle domande di dominio sensomotorio e cognitivo per le quali le risposte non sono comprese nelle formulazioni delle domande.

Le domande cognitive riflettono ciò che l'utente non sa, riflettono il bisogno di informazione sintattica e la scarsa conoscenza circa la logica del motore di ricerca. Gli utenti inesperti hanno bisogno di aiuto in queste particolari circostanze.

Riguardano:

a) l'interazione col sistema (system image questions come ad esempio: il sistema ti permette di mettere una x per selezionare il record? Oppure: è possibile cercare l'anno specifico di un determinato e-journal?...);

b) la strategia di ricerca (search strategy question) ovvero la comprensione del linguaggio booleano, i troncamenti....

c) la comprensione del contenuto dell'informazione che si è recuperata.

Tra i domini vi è una gerarchia: se la domanda di dominio affettivo riceve una risposta positiva, allora si procede nella ricerca e si accede eventualmente alle altre domande degli altri due domini, cognitivo e sensomotorio.

Le domande affettive assumono rilievo negli approcci user oriented, a differenza che negli approcci system oriented, dove non si danno risposte a domande che hanno già in sé stesse la risposta.

Le domande sensomotorie concernono

a) l'utilizzo degli strumenti di cui si compone la macchina come: mouse, tastiera, stampante, schermo... Esempi di tali domande (formatting input) sono: cosa devo premere ora?; come posso continuare a visualizzare le pagine?; devo lasciare uno spazio?...

b) (screen format questions) la localizzazione e l'identificazione delle informazioni che compaiono sullo schermo. Generalmente non sono domande che rinviano a tare cognitive ma ad incomprensioni circa l'utilizzo degli strumenti che consentono il riconoscimento e la localizzazione delle informazioni sullo schermo. Gli utenti non avvezzi alla consultazione delle basi di dati mostrano difficoltà nel seguire le informazioni in modo sistematico sullo schermo, a

riconoscere la struttura di un record e a distinguere tra testo e campo-testo.

A causa della differenza tra i vari sistemi operativi e tra i database stessi, queste domande risultano essere croniche. Durante la ricerca compiuta dall'autrice, per questo tipo di domande gli utenti preferiscono chiedere al supervisore anziché cercare da soli la risposta (vi è quindi un minor costo nel chiedere che nel cercare da soli, magari anche attraverso l'uso degli help on line).

## **Appendice**

*Proposta di questionario per la valutazione del servizio reso alla ricerca da parte dell'OPAC nel contesto della Biblioteca d'Ateneo dell'Università degli Studi di Milano Bicocca*

Sulla base delle indicazioni emerse dagli studi riportati nei capitoli precedenti, si è cercato di formulare un questionario a scelta multipla, con l'obiettivo di valutare la qualità del servizio reso nella ricerca bibliografica dall'OPAC, nel contesto della Biblioteca d'Ateneo dell'Università degli Studi di Milano Bicocca.

E' stata utilizzata una scala Likert a 5 punti. Di conseguenza ad ogni risposta è possibile attribuire un punteggio che varia da 1 a 5 (da una risposta per niente pertinente ad una risposta del tutto pertinente).

## Questionario

a) Quali sono gli elementi della notizia bibliografica con i quali consultati il catalogo in linea (OPAC)? (c)

- Autore-titolo (5)
- Isbn (1)
- Soggetto (2)
- Parola del titolo (4)
- Collocazione (3)

b) Ritieni la consultazione del catalogo in linea (OPAC) (c)

- Utile prima della recupero del materiale (4)
- Indifferente al processo di recupero del materiale (1)
- Utile per sapere se il libro è presente in biblioteca (5)
- Utile per sapere dove è il libro in biblioteca (3)
- Utile alla fine della ricerca (2)

c) ritieni l'OPAC (c)

- Uno strumento per il recupero del materiale in biblioteca
- Uno strumento utile al solo personale della biblioteca
- Uno strumento di informazione generale sui servizi della biblioteca
- Una complicazione in più rispetto all'iter della ricerca
- 

d) ritieni le notizie fornite dall'OPAC di Ateneo (s)

- Organizzate in maniera funzionale alla ricerca (4)
- Di difficile consultazione (2)
- Localizzate sullo schermo in modo ben organizzato (5)
- Localizzate sullo schermo in modo incomprensibile (1)

- Di facile consultazione (3)
- e) Un libro è in prestito (s)
- Quando lo trovi nell'OPAC ma non a scaffale (3)
- Quando lo trovi nell'OPAC ma il personale bibliotecario ti fornisce la notizia che il libro è al momento in prestito (2)
- Quando l'OPAC ti fornisce la notizia che il libro è al momento in prestito (5)
- Quando l'OPAC ti fornisce la notizia che il libro è stato ordinato (4)
- Quando non trovo la notizia bibliografica nell'OPAC (1)
- f) Hai mai fatto la prenotazione di un libro? (s)
- No perché il personale bibliotecario non ti ha fornito la notizia di questa possibilità (1)
- No perché l'OPAC non ti ha fornito la notizia di questa possibilità (2)
- Ci hai provato ma ti mancavano le chiavi d'accesso alla pagina web della prenotazione (3)
- Ci hai provato ma hai dovuto chiedere conferma dell'operazione al personale bibliotecario (4)
- Sì con successo (5)
- g) Ritieni che per la consultazione dell'OPAC (a)
- Sia necessario un corso propedeutico (5)
- Siano necessari opuscoli informativi predisposti dalla biblioteca d'Ateneo (4)
- Ci debba essere personale bibliotecario a supporto della consultazione (3)
- Sia sufficiente in caso di difficoltà, l'help on line (2)



- Sia sufficiente migliorare l'interfaccia grafica e renderla più comprensibile (1)
- h) spesso le notizie fornite dall'OPAC sono (s)
- Non esaustive (1)
- Abbastanza esaustive (4)
- Sufficientemente esaustive (3)
- Poco esaustive (2)
- Del tutto esaustive (5)
- i) Trovi che la ricerca abbia più efficacia quando (a)
- Cerchi i libri nell'OPAC da solo (2)
- Cerchi i libri al catalogo con un compagno di studi (3)
- Chiedi all'ufficio prestiti di compiere per te la ricerca (4)
- Chiedi al personale bibliotecario di mostrarti come si imposta la ricerca (5)
- Ritieni essere tuo diritto che la biblioteca possieda il libro che cerchi (1)
- l) tra i motivi per cui non ti soddisfa l'OPAC ci sono (s)
- Una tua scarsa attitudine all'uso del computer (1)
- Incapacità di proseguire nella ricerca di fronte a messaggi del software che non capisco (2)
- Strutturazione inadeguata delle informazioni fornite (3)
- Insufficienza delle informazioni rispetto al recupero del materiale (4)
- Non ci sono motivi di insoddisfazione rispetto all'OPAC (5)

m) nel caso in cui non trovi ciò che cerchi (a)

Abbandoni la ricerca (1)

Riformuli la domanda inserendo altre notizie bibliografiche in tuo possesso (5)

Chiedi al vicino se ti può aiutare (3)

Consulti l'OPAC di un'altra biblioteca che già conosci e frequenti (4)

Ti accerti dell'esito negativo della ricerca rivolgendoti al personale bibliotecario (2)

n) quando trovi discordanza tra la notizia dell'OPAC e la situazione allo scaffale (tipo il catalogo dice che il libro c'è ma il libro a scaffale manca) ritieni (a)

Di aver letto male le notizie del catalogo (1)

Di dover fare presente la cosa al personale della biblioteca (5)

Di aver interpretato male le notizie del catalogo (4)

Che la notizia sia stata data in maniera poco chiara (2)

Che il personale bibliotecario non attui i necessari controlli (3)

o) sei a conoscenza delle diverse modalità di ricerca proposte dall'OPAC (ricerca combinata, catalogo dei periodici, salvataggio della ricerca...)? (s)

Sì le ho sperimentate quando avevo difficoltà a recuperare il materiale dopo la ricerca base (4)

Sì ma non le so usare perché non conosco la procedura per giungervi (2)

Sì ma non le so usare perché non conosco la procedura di ricerca combinata (3)

Sì, le uso abitualmente (5)

- No, l'home page dell'OPAC non ne fa cenno (1)
- p) Ti capita a volte di (a)
- Trarre le notizie essenziali dal catalogo ma di aver bisogno comunque di chiedere chiarimenti al personale della biblioteca (3)
- Trarre le notizie essenziali dal catalogo ma di aver problemi nell'interpretazione del codice di collocazione del materiale (2)
- Giungere all'Ufficio Prestito con le notizie bibliografiche sbagliate e dover ripetere la ricerca (2)
- Sentirmi inadeguato a cercare il materiale da solo a causa della dispersione del materiale nella biblioteca (5)
- Entrare in biblioteca e sentire il bisogno di chiedere aiuto perché non mi so orientare (4)
- q) che idea ti sei fatto del funzionamento dell'OPAC? (c)
- Non hai minimamente idea di come il terminale giunga a recuperare i dati (1)
- Ritieni che qualcosa di magico succeda per cui ciò che cerchi viene recuperato (2)
- Credi che vi sia un linguaggio informatico che traduce la tua domanda in termini comprensibili alla macchina e che ti interesserebbe studiare (4)
- Credi che vi sia un linguaggio informatico che traduce la tua domanda in termini comprensibili alla macchina e che però non è tuo interesse approfondire (3)
- Ritieni che la il funzionamento dell'OPAC dipenda dalla corretta formulazione della tua richiesta (5)

## Bibliografia

Agosti (2002). Dispense di elementi di informatica per studenti di materie umanistiche / Maristella Agosti, Nicola Orio. - Padova: Progetto, 2002.

Capra (2003). Re-Finding Found Things: An Exploratory Study of How Users Re-Find Information [Risorsa elettronica] / Robert G. Capra, Manuel A. Perez-Quinones. - Documento. - S. l.: S. e., 2003. - Modalità di accesso: World Wide Web. URL:<http://it.arxiv.org/ftp/cs/papers/0310/0310011.pdf>. - Descrizione della risorsa al 30 novembre 2003.

Crestani (1998). Information retrieval : uncertainty and logics : advanced models for the representation and retrieval of information / edited by Fabio Crestani, Mounia Lalmas, Cornelis Joost van Rijsbergen. - Boston [etc.] : Kluwer Academic Publishers, c1998.

Croft (1995). What do people want from Information Retrieval? [Risorsa elettronica] / W. Bruce Croft. - In: D-Lib Magazine. - (1995), Novembre. - Modalità di accesso: World Wide Web. URL: <http://www.dlib.org/dlib/november95/11croft.html>. Descrizione della risorsa al 30 novembre 2003.

De Robbio (1999). La biblioteca nel Web, il Web nella biblioteca / Antonella De Robbio. - In: Bibliotime, anno II (1999), 2. - Disponibile anche su World Wide Web. URL: <http://www.spbo.unibo.it/bibliotime/num-ii-2/derobbio.htm>.

Fattahi (1997). The relevance of cataloguing principles to the online environment [Risorsa elettronica] : an historical and analytical study / Rahmatollah Fattahi. - Sidney : New South Wales University, 1997. - Modalità di accesso: World Wide Web.

URL:

<http://web.archive.org/web/20011008015702/wilma.silas.unsw.edu.au/students/RFATTAHI/ch3.htm>. - Descrizione della risorsa elettronica al 30 novembre 2003.

Hsieh-Yee (1998). Search Tactics of Web Users in Searching for Texts, Graphics, Known Items and Subjects: A Search Simulation Study / Ingrid Hsieh-Yee. - In Electronic Resources : use and users behavior / Hemalata Iyer, editor: The Haworth Press, Inc, 1998, pp. 61-85.

Meadow (2000). Text information retrieval systems / Charles T. Meadow, Bert R. Boyce, Donald H. Kraft. - 2. ed.. - San Diego [etc.] : Academic press, c2000.

Nahl (1996). Affective and Cognitive Searching Behavior of Novice End-Users of a Full-Text Database / Diane Nahl, Carol Tenopir. - In: Journal of the American Society for Information Science. - 47(1996), 4: 276-286.

O'Neill (2003). Trends in the evolution of the Public Web: 1998-2002 [Risorsa elettronica] / Edward T. O'Neill, Brian F. Lavoie, Rick Bennett. - In: D-Lib magazine. - 9(2003), 4. - Modalità di accesso: World Wide Web. URL: <http://www.dlib.org/dlib/april03/lavoie/04lavoie.html>. - Descrizione della risorsa al 30 novembre 2003.

Palmquist (1998). Modeling the Users of Information Systems: Some Theories and Methods / Ruth A. Palmquist, Kyung-Sun Kim. - In Electronic Resources : use and users behavior / Hemalata Iyer, editor: The Haworth Press, Inc, 1998, pp. 3-25.

*In ricordo della compagna Mery*