

La ricerca dei documenti giuridici in Rete

di

Michele Iaselli

Gli operatori logici Booleani

Gli operatori logici Booleani sono fondamentali per qualsiasi ricerca documentaria di carattere informatico.

Essi sono:

- AND
- OR
- NOT

Il simbolo con cui viene indicato l'AND è l'asterisco *

L'operatore funziona in base al seguente principio: perché si abbia un certo risultato in uscita, devono essere contemporaneamente presenti all'ingresso tutte le condizioni richieste.

Il simbolo con cui si indica l'OR è quello algebrico della somma +.

L'operatore funziona in base al seguente principio: perché si abbia un certo risultato in uscita, deve essere presente all'ingresso almeno una condizione richiesta.

Il simbolo con cui si indica il NOT è quello algebrico della sottrazione racchiuso tra parentesi (-).

L'operatore funziona in base al seguente principio: c'è sempre un solo ingresso ed una sola uscita. Se alla porta di ingresso c'è una data condizione, il risultato all'uscita è costituito dalla condizione opposta.

Gli *operatori logici booleani* non sono altro che il meccanismo attraverso il quale la nostra mente forma e verifica tutti i concetti. Un concetto, difatti, è un insieme di elementi, alcuni dei quali devono coesistere, altri possono indifferentemente esserci o non esserci, ed altri ancora non devono esserci.

Resta inteso però che non è la macchina che stabilisce se due o più valori debbano essere confrontati in AND ovvero in OR ovvero in NOT; la macchina ha bisogno di dati forniti dall'uomo, su questi applica, attraverso le sue strutture, un qualsiasi operatore logico booleano (o una loro combinazione) indicatogli dall'uomo ed ogni suo ragionamento non può avvenire altro che servendosi di tale strutture. Non si deve, perciò, mai parlare di "logica del computer", ma di "logica nel computer".

Le banche dati

Le tecniche descritte nel paragrafo precedente hanno trovato applicazione principalmente nelle cd. *banche dati* di carattere informatico, concetto relativamente nuovo con cui si designano *sistemi automatici di raccolta, conservazione, elaborazione e ricerca di dati di qualsiasi natura.*

Le banche dati si distinguono dagli archivi cartacei per cinque ragioni, ciascuna delle quali dà luogo ad un *principio fondamentale di libertà per l'individuo*.

In un archivio tradizionale si può effettuare una ricerca *soltanto se si conosce il dato* corrispondente a quello prescelto dall'organizzatore dell'archivio come criterio base per l'ordinamento del medesimo; nella ricerca automatica, invece, *il ricercatore è libero di scegliere il dato su cui imperniare la ricerca*, in quanto il computer riesce a rintracciare i documenti contenenti quel dato ovunque siano stati collocati e in qualsiasi punto del documento quel dato si trovi (principio della libertà di scelta del dato).

In un archivio tradizionale *non è possibile indirizzare la ricerca ad un livello più specifico* di quanto non consenta l'ultima delle sottoripartizioni o classificazioni più capillari predisposte dall'organizzatore dell'archivio stesso; invece, in un archivio elettronico, il ricercatore può sempre, mediante la libera combinazione di più dati posti in *and, or, not, far collimare la selezione automatica a qualsiasi livello* e ottenere dal computer una risposta adeguata alla particolarità o varietà dei suoi interessi (*principio della libera combinazione dei dati di ricerca*).

In un archivio tradizionale *per eseguire una ricerca indirizzata occorre conoscere non soltanto il dato corrispondente al tipo prescelto per l'ordinamento dell'archivio, ma tutti i caratteri che lo costituiscono e in particolare quelli iniziali*; invece, in un archivio elettronico, il ricercatore può utilizzare non soltanto *qualsiasi tipo di dato, ma anche quelli dei quali, per ignoranza o dimenticanza, non sia in grado di indicare una parte dei caratteri che li costituiscono*, ancorché trattasi della parte iniziale o preponderante di essi, essendogli sufficiente sostituire tutti i caratteri del dato non indicati con un segno convenzionale espressivo dell'incognita (*principio del mascheramento dei dati*).

Nell'archivio tradizionale *una ricerca indirizzata di semplice orientamento può eseguirsi solo in ordine all'unico dato prescelto dall'archivista per ordinare l'archivio*; invece, in un archivio elettronico, tale ricerca *può essere consentita per qualsiasi tipo di dato*, nel senso che il ricercatore può sempre limitarsi ad indicare *soltanto il tipo di dato cui esso appartiene* e ottenere, così, dal computer un prospetto di sintesi in cui, per ogni dato concreto corrispondente a quel tipo, sono indicati quanti dei documenti in archivio lo contengono (*principio della estraibilità delle informazioni dai documenti mediante le cd. analisi spettrali*).

Nella ricerca manuale e in quella automatica *si può pervenire alla selezione dei documenti mediante approssimazioni successive*, frutto di un processo interattivo tra la documentazione via via reperita e il ricercatore adattando la strategia di ricerca alla *valutazione dei risultati progressivamente ottenuti (principio del carattere colloquiale della ricerca)*.

Le banche dati in Italia

In Italia le prime banche dati sono sorte negli anni settanta; da allora l'informatica giuridica documentaria si è molto sviluppata in Italia ed offre ampie possibilità di acquisire informazioni relative alla legislazione, alla giurisprudenza e alla letteratura di rilevanza giuridica.

Il collegamento alle banche dati telematiche può avvenire: direttamente agli *host computer* (calcolatori-distributori che hanno in memoria una o più banche dati, accessibili *on line* via terminale attraverso una rete di trasmissione dati), su *linea commutata* (cioè la rete telefonica generale appositamente instradata, v. ad esempio Internet) oppure *dedicata* (connessione su circuito diretto per la specifica esigenza).

In Italia esistono diversi sistemi operativi di informatica delle fonti, che con l'avvento di Internet hanno assunto una notevole varietà nonché facilità di consultazione, tra i principali e per così dire *Istituzionali* si ricordano:

- il Centro di documentazione automatica (CDA) della Camera dei Deputati e del Senato;
- il GIURITEL;
- "Norme in Rete" che è un portale di accesso alle norme pubblicate sui siti delle P.A. partecipanti curato dall'Istituto di Teoria e Tecniche dell'Informazione Giuridica (ITTIG);
- la Gazzetta Ufficiale Telematica;
- Il Centro Elettronico di Documentazione (CED) della Corte Suprema di Cassazione.

Di particolare importanza è quest'ultimo per almeno tre ragioni:

- 1) la vastità degli archivi;
- 2) la diffusione del servizio di informazione, aperto nel 1973 a tutti gli uffici giudiziari ed esteso, a seguito del D.P.R. 21 maggio 1981 n. 322, a tutti i cittadini;
- 3) il sistema di ricerca (*Italgiure Find*) appositamente studiato per la documentazione giuridica da un gruppo di magistrati della Cassazione, in collaborazione con tecnici informatici e dotato di una serie imponente di strumenti linguistici e di classificazione.

Oltre a queste banche dati giuridiche *on-line* di carattere istituzionale esistono ulteriori banche dati di carattere privato che sono disponibili sia on line che off line su DVD e specialmente le prime hanno l'indubbio vantaggio dell'aggiornamento in tempo reale.

Queste banche dati possono distinguersi in *archivi di dottrina e giurisprudenza* ed in *archivi di legislazione* forniti da diverse società come ad esempio l'Utet, la Giuffré, la De Agostini, la Tribuna, l'Ipsoa. Tali archivi sono ormai tutti disponibili on line, con la possibilità di avere periodicamente un DVD per una consultazione off line.

Gli archivi legislativi contengono tutta la legislazione vigente in Italia, a livello sia legislativo che regolamentare, con esclusione delle disposizioni contenute nei codici e quasi tutti ormai contengono le principali massime giurisprudenziali collegate ad ogni norma (per esempio De Agostini, Ipsoa).

Gli archivi di dottrina e giurisprudenza partono, in genere, dal 1981 e riportano tutte le massime pubblicate sui repertori; le chiavi di ricerca sono molteplici ed in parte coincidenti con quelle di *Italggiure Find*. La ricerca per unità lessicale è però limitata, in quanto l'elaboratore mostrerà soltanto i documenti che presentano *esattamente la chiave richiesta*, così come è stata formulata, e non i documenti che presentano la stessa parola al plurale piuttosto che al singolare, o lo stesso verbo coniugato piuttosto che all'infinito. Questo limite (come si è visto in precedenza) viene però facilmente superato con l'uso della tecnica del *mascheramento dati*.

La ricerca di giurisprudenza può venire impostata per *full-text* (il metodo più efficace), per riferimento legislativo, per ente giudicante, per voci, per anni di repertorio. La ricerca di dottrina può venire impostata per *full-text*, per autore, per anni di repertorio.

L'uso del *full-text* consente l'uso degli operatori logici di tipo algebrico "+", "-", "and", "or" o "not", oppure "e" ed "o". L'elaboratore fornisce una risposta che consente all'operatore di raggiungere sia un risultato immediato (può leggere subito la massima delle sentenze cui è interessato), sia un risultato mediato (ottiene gli estremi dell'eventuale pubblicazione della sentenza stessa).

Ricerca dei documenti giuridici su Internet

La diffusione delle reti telematiche, in particolare di Internet, ha introdotto un elemento nuovo e per certi versi rivoluzionario. L'accesso all'informazione avviene, in questa prospettiva reticolare, in modo diffuso; l'utente è per così dire immerso in una notevole quantità di dati disordinati che spaziano da un polo all'altro del globo terrestre. L'influenza di questo approccio produce i suoi effetti anche sul modo di ragionare e quindi di fare ricerche; prevale infatti un tipo di ragionamento associativo rispetto a quello deduttivo, al quale siamo culturalmente più abituati.

Alla visione di banche dati centralizzate, che pure hanno indiscussi vantaggi, si va affiancando la visione più ampia di una informazione distribuita, dislocata cioè in più punti della Rete.

I vantaggi e gli svantaggi delle due soluzioni sono in realtà per così dire speculari. Le principali caratteristiche degli archivi centralizzati sono rappresentate da informazioni strutturate, sistemi di reperimento omogenei e più sofisticati, registrazione uniforme dei dati.

Nel sistema ad informazione diffusa mancano gli elementi sopra descritti, e questo comporta sicuramente dei problemi da risolvere, ma sono presenti altri vantaggi quali, innanzi tutto, la ricchezza di informazione, l'ampiezza della diffusione, la responsabilità dei dati (l'informazione risiede dove nasce), la possibilità di aggiornamento tempestivo, l'agilità nella creazione degli archivi.

Ormai la soluzione tradizionale (centralizzata) è per molti aspetti superata e di fatto impraticabile, se ragioniamo in termini di larga scala: è impensabile, ad esempio, che presso un sistema informativo si raccolga tutta la normativa italiana sia a carattere nazionale che a carattere locale (regionale, provinciale, comunale, ecc.), per l'impossibilità pratica (più che tecnica) di gestire una mole di dati di questa portata, mentre appare ragionevolmente realizzabile l'ipotesi che ogni realtà territoriale renda disponibile, senza eccessivo sforzo, la propria produzione normativa.

Ciò non ci esime dal constatare che l'informatica giuridica con le sue banche ordinate ha prodotto notevoli benefici, mentre il regno di Internet si presenta come un mondo di anarchia, dove ciascuno mette in linea ogni tipo di materiale nelle forme più diverse: elenchi, banche dati, testo integrale, riferimenti. Il panorama si presenta così variegato da appesantire spesso le operazioni di reperimento.

Sembrerebbe quindi che l'utilizzo di reti telematiche segni un passo indietro, ma se si entra in un nuovo tipo di filosofia e si realizzano quegli accorgimenti che superino i lati negativi appena considerati i vantaggi che emergeranno saranno sicuramente di grande portata.

Difatti, l'esigenza di sfruttare al meglio le potenzialità di una rete telematica come Internet ha consentito di ricorrere a strumenti, come guide, motori di ricerca specifici, intelligent agent che consentano all'utente di orientarsi e ottenere delle informazioni pertinenti, ristabilendo un ordine fra il disordine.

Ma affinché tali strumenti possano essere efficienti è necessario che siano coordinati fra di loro: l'elemento indispensabile che sta alla base di questa nuova filosofia è infatti proprio quello della cooperazione che non significa solo comunicarsi dati, ma significa, in prospettiva, concordare strategie, uniformare certi canoni di base (per esempio rendere omogenee le citazioni) in modo da evitare per quanto possibile la dispersività che può divenire fonte di disinformazione.

Nel settore giuridico, però, la ricerca dell'informazione presenta maggiori complessità, in quanto a differenza di Internet, nell'ambito di una banca dati giuridica siamo in presenza di dati strutturati, sia grazie all'unitarietà del contenuto, sia grazie all'opera di catalogazione e indicizzazione effettuata da operatori specializzati. L'efficienza della ricerca, quindi, è data da diversi fattori quali: ambito circoscritto, omogeneità dei materiali, possibilità di combinare diverse metodologie di ricerca, possibilità di effettuare ricerche per codici identificativi numerici. Tutti questi fattori non esistono in Internet.

Ulteriori ricerche sono state fatte nel campo dei metamotori di ricerca, che si presentano all'utente come pagine Web in cui occorre inserire un'unica stringa di ricerca e indicare quali motori di ricerca consultare. Il metamatore si occuperà di inviare contemporaneamente la query ai vari motori selezionati, adattando la sintassi della query a quelle richieste dai singoli motori: la ricerca sarà effettuata quindi in parallelo, i risultati saranno poi presentati insieme, dopo aver eliminato le ridondanze.

Sulla base dello stesso principio che ispira i metamotori sono stati creati i cd. *metadata* che possono rivelarsi particolarmente utili per il reperimento di informazioni giuridiche in reti distribuite come appunto Internet. I metadata possono essere utili in quanto possono fornire all'utente un metodo più affidabile delle attuali ricerche di tipo full text o tramite meri soggettari. La possibilità di reperire informazioni rilevanti in Internet "sembra" decrescere proporzionalmente alla quantità di informazioni disponibili: per risolvere tale problema è emersa recentemente la proposta di "arricchire" i materiali pubblicati in Internet affiancando loro una descrizione mediante "informazioni concernenti altre informazioni": i *metadata* appunto.

Risulterebbe così possibile descrivere la natura di una risorsa Internet: la sua ubicazione, l'oggetto di cui tratta, il formato con cui tale oggetto è presentato, ecc. Un'adozione estesa dei *metadata* nella descrizione-catalogazione delle risorse Internet faciliterebbe, quindi, il reperimento delle informazioni e in casi particolari anche una loro valutazione qualitativa.

Una nuova concezione della Rete

A prescindere dall'adozione di questi nuovi strumenti bisogna riconoscere che fino a tre anni fa, comunque, il panorama dell'informazione su Internet era desolante: uffici pubblici, ministeri, enti o agenzie governative neppure lontanamente pensavano di usare questo mezzo come strumento di diffusione del sapere.

Oggi, come è noto, le cose sono sensibilmente cambiate, anche se ci appare che la fase attuale sia una fase di transizione o, meglio, di assestamento.

Si sta facendo largo, in numerosi siti, una particolare attenzione alla qualità dell'informazione offerta ed alla sua sistemazione organica, tale da garantire un approccio il più "indolore" possibile all'utente che si avvicina alla rete.

L'opera sarà probabilmente completa quando anche l'informazione utilizzerà in pieno le potenzialità offerte dalla rete Internet e si trasformerà in informazione interattiva, tale da permettere non solo la consultazione ma anche la partecipazione dell'utente a determinati procedimenti, sia in fase di ricerca che nell'utilizzo stesso dell'informazione.

Sarà allora necessario un passaggio da una staticità dell'informazione, assai diffusa, ad un utilizzo dinamico delle risorse che la rete offre. È il passaggio, ad esempio, da un sito puramente informativo – anche se accurato ed autorevole – ad un sito che, accanto all'informazione, offra una sorta di interattività, in poche parole la partecipazione dell'utente ad un qualche processo, dal più semplice al più complesso.

Alcuni siti pubblici stanno già mettendo in pratica questo modo nuovo di concepire l'utilizzo di Internet, ma è necessario ancora lavorare molto per ottenere un effettivo aumento e miglioramento dei servizi disponibili in rete, così da rendere quest'ultima appetibile per quanti siano interessati ad ottenere prestazioni a valore aggiunto, vale a dire senza code agli sportelli, senza spostarsi da casa, con possibilità di verifica costante dello stato delle pratiche, con la possibilità di avvalersi di sistemi di pagamento elettronico ecc.

Il Governo ed in particolare il Ministero per l'Innovazione e le Tecnologie ha dettato regole fondamentali per lo sviluppo di Internet che hanno inciso non solo in ambito pubblicitario ma in tutti i settori dell'informazione e dei servizi.

Lo stesso codice dell'Amministrazione digitale disciplina anche la presenza delle pubblica amministrazione on line ed in particolare agli articoli 53 e 54 viene previsto l'obbligo per gli enti pubblici di riorganizzare i propri siti Internet in modo da individuare una serie di contenuti minimi e necessari, compresa la disponibilità di moduli e formulari per via telematica.