

---

# 01. L'oggetto della ricerca giuridica informatizzata

---

## 1.1. Cenni generali

---

L'*informatica giuridica documentale* ha come oggetto di studio la ricerca di documentazione giuridica effettuata con strumenti informatici. La fortuna di tale disciplina è cresciuta negli anni proporzionalmente alla diffusione delle tecnologie informatiche, in considerazione del fatto che gran parte dell'informazione di interesse giuridico è oggi disponibile in formato digitale e quindi consultabile o raggiungibile con strumenti informatici.

Tuttavia, la ricerca giuridica informatizzata soffre attualmente di alcuni inconvenienti, tra cui:

- a) Una parte della documentazione giuridica, e soprattutto quella relativa alla giurisprudenza di merito, non è pubblicamente disponibile in formato digitale, oppure è disponibile in quanto raccolta in banche dati accessibili soltanto a titolo oneroso.
- b) Considerata la massa documentale disordinata (soprattutto in Internet), sussistono rischi di reperire informazioni giuridiche non aggiornate, non rilevanti o inesatte.
- c) La ricerca documentaria condotta tramite strumenti digitali presuppone la padronanza di alcune competenze informatiche di base, pertanto non può essere effettuata da giuristi sprovvisti di una sia pur rudimentale alfabetizzazione informatica.

## 1.2. L'oggetto della ricerca

---

Oggetto della ricerca informatizzata di documentazione giuridica è ovviamente il *documento giuridico*, in versione digitale. Il documento giuridico può essere considerato come uno *strumento idoneo a veicolare informazioni giuridicamente rilevanti, indipendentemente dalla metodologia di rappresentazione e dal supporto utilizzato*.

L'informatica giuridica documentale riconosce una classica tripartizione dei documenti giuridici. I documenti giuridici vengono infatti tradizionalmente distinti in:

- 1) documenti normativi (chiamati anche, impropriamente, legislativi);
- 2) documenti giurisprudenziali;
- 3) documenti dottrinali.

### 1.2.1. I documenti normativi

---

I documenti normativi sono quelli che danno conto di norme giuridiche generali e astratte. Queste possono essere di emanazione internazionale, comunitaria, statale, regionale ecc. e provengono da autorità pubbliche dotate di poteri di produzione normativa.

Gli atti normativi sono quasi sempre soggetti a forme di pubblicazione ufficiale che ne determinano la formulazione canonica e ne condizionano l'entrata in vigore: Gazzetta ufficiale, bollettini di vario genere e, sempre più spesso, siti internet istituzionali. È dunque spesso possibile reperire nel giorno stesso della pubblicazione la normativa comunitaria, statale, regionale e gli atti degli organi amministrativi centrali e periferici.

Vediamo di seguito come si presenta un atto normativo tipico quale è un decreto legislativo nell'ordinamento giuridico italiano, e cerchiamo di individuare gli elementi più rilevanti ai fini della ricerca giuridica. Un decreto legislativo è strutturato secondo il seguente schema:

<b>epigrafe</b>	<b>Tipo</b>	<i>Decreto legislativo</i>
	<b>Data</b>	<i>7 marzo 2005</i>
	<b>Numero</b>	<i>n. 82</i>
	<b>Titolo</b>	<i>Codice dell'amministrazione digitale - CAD</i>
<b>Preambolo</b>	<p>IL PRESIDENTE DELLA REPUBBLICA</p> <p>Visti gli articoli 76, 87 e 117, secondo comma, lettera r), della Costituzione;</p> <p>Visto l'articolo 14 della legge 23 agosto 1988, n. 400, recante disciplina dell'attività di Governo e ordinamento della Presidenza del Consiglio dei Ministri;</p> <p>Visto l'articolo 10 della legge 29 luglio 2003, n. 229, recante interventi in materia di qualità della regolazione, riassetto normativo e codificazione - legge di semplificazione 2001;</p> <p>[...]</p> <p>Sentito il Garante per la protezione dei dati personali;</p> <p>Udito il parere del Consiglio di Stato, espresso dalla Sezione consultiva per gli atti normativi nell'adunanza del 7 febbraio 2005;</p> <p>[omissis]...</p> <p>EMANA</p> <p>il seguente decreto legislativo:</p>	
<b>Articolato</b>	<p>CAPO I</p> <p>Capo I</p> <p>PRINCIPI GENERALI</p> <p>SEZIONE I</p> <p>Sezione I</p> <p>Definizioni, finalità e ambito di applicazione</p> <p>Articolo 1</p> <p>Definizioni</p> <p>Art. 1. Ai fini del presente codice si intende per:</p> <p>Oa) AgID: l'Agenzia per l'Italia digitale di cui all'articolo 19 del decreto-legge 22 giugno 2012, n. 83, convertito, con modificazioni, dalla legge 7 agosto 2012, n. 134(1);</p> <p>[a) allineamento dei dati: il processo di coordinamento dei dati presenti in più archivi finalizzato alla verifica della corrispondenza delle informazioni in essi contenute;] (2)</p> <p>[b) autenticazione del documento informatico: la validazione del documento informatico attraverso l'associazione di dati informatici relativi all'autore o alle circostanze, anche temporali, della redazione;] (3)</p> <p>d) carta d'identità elettronica: il documento d'identità munito di elementi per l'identificazione fisica del titolare rilasciato su supporto informatico dalle amministrazioni comunali con la prevalente finalità di dimostrare l'identità anagrafica del suo titolare (4);</p> <p>[...]</p>	

<b>(eventuale) Allegati</b>	ALLEGATO 1 TABELLA DI CORRISPONDENZA DEI RIFERIMENTI PREVIGENTI AL CODICE DELLE DISPOSIZIONI LEGISLATIVE E REGOLAMENTARI IN MATERIA DI DIGITALIZZAZIONE DELLE PUBBLICHE AMMINISTRAZIONI. <i>[omissis]</i>
-----------------------------	---

Fig. 1: struttura standard di un decreto legislativo

- **l'epigrafe.** Contiene gli *estremi* dell'atto, e dunque una serie di dati necessari all'individuazione dello stesso. Si tratta dei seguenti elementi:
  - il **tipo** di *atto normativo* (legge, decreto legge, decreto legislativo ecc.), elemento che in certi casi rivela anche l'ente emanante (decreto del Presidente della Repubblica, decreto del Presidente del Consiglio dei ministri eccetera);
  - la **data**, che indica il giorno in cui l'atto normativo è stato *promulgato*. La promulgazione è l'atto formale con il quale si dichiara valido ed efficace un atto normativo. È diversa dalla data della pubblicazione nella Gazzetta Ufficiale o in altri bollettini aventi funzione di pubblicità; la data di pubblicazione, infatti, indica il giorno a partire dal quale la norma si considera vigente a tutti gli effetti. Nel caso della legge ordinaria italiana, la pubblicazione sulla Gazzetta Ufficiale avviene subito dopo la promulgazione, e comunque entro 30 giorni dalla stessa.
  - il **numero** dell'atto (se presente). Gli atti normativi seguono solitamente una numerazione progressiva annuale: fanno eccezione molti atti normativi secondari, come ad esempio i decreti ministeriali e i decreti del Presidente della Repubblica, che possono o meno essere numerati. Si noti che tipo, data e, qualora presente, il numero sono gli estremi identificativi essenziali per il reperimento e la menzione degli atti normativi (vedi *infra*).
  - Il **titolo** dell'atto, che identifica e spesso indica in forma sintetica il contenuto ovvero l'oggetto dell'atto. Dal titolo si può evincere la pertinenza dell'atto normativo agli obiettivi della ricerca, ma vi sono anche atti che vengono pubblicati con cadenze periodiche, come ad esempio le leggi di stabilità o i c.d. "decreti milleproroghe", i cui titoli, più che al contenuto normativo dell'atto, fanno riferimento agli scopi che con l'atto si vogliono perseguire o alla funzione generale assegnata alla norma. In questi casi, la valenza informativa del titolo rispetto agli specifici obiettivi di ricerca è ridotta.
- Il **preambolo**. In tale sezione sono indicati gli organi che hanno partecipato alla formazione e alla promulgazione dell'atto normativo, anche con funzioni consultive, nonché, in taluni casi, le precedenti norme, le ragioni e le circostanze che ne hanno reso giuridicamente doverosa o politicamente opportuna l'emanazione (c.d. "*motivazione*" dell'atto normativo).
- **L'articolato**. Si tratta del contenuto dell'atto, il testo normativo vero e proprio, che può essere più o meno esteso anche in considerazione delle diverse materie trattate. Il testo normativo si chiama "articolato" appunto perché suddiviso in un numero variabile di *articoli*, che sono delle porzioni di testo normativo che includono una o più disposizioni. Gli articoli sono talora raggruppati in **sezioni e/o capi**, che sono degli insiemi di articoli di contenuto omogeneo (più sezioni, se presenti, possono formare un capo); più capi possono essere riuniti in **titoli** tematici, uno per ogni materia disciplinata, e più titoli possono essere raggruppati in **parti**, ad esempio, una parte prima comprende i principi generali della materia regolata, una parte le norme che disciplinano specifici settori o ambiti, una parte le norme in materia di tutela giuridica e le sanzioni ecc. Parti, titoli, capi e sezioni sono numerati secondo il sistema di numerazione romano (I, II, III, IV ecc.).

Gli articoli sono numerati secondo il sistema di numerazione decimale e spesso intitolati per mezzo di una "rubrica", che indica sinteticamente l'oggetto dell'articolo.

Gli articoli possono a loro volta essere divisi in **commi**, che possono essere ulteriormente divisi in **numeri e/o lettere**. Avviene a volte che una legge contenga soltanto due o tre articoli suddivisi però in decine o centinaia di commi, a loro volta suddivisi in numeri e lettere.

Ricapitolando: l'articolato è strutturato, procedendo dal generale al particolare, in parti, titoli, capi, sezioni, articoli, commi, numeri e lettere, tutte porzioni del testo normativo che possono essere prese in considerazione singolarmente in fase di ricerca.

### Testo dell'atto normativo

PARTE I

TITOLO I

CAPO I

SEZIONE I

Articolo 1

Comma 1.

Lettera a)

Numero 1.

Fig. 2: Struttura dell'articolato

- **Allegati.** Questi elementi sono meramente eventuali. Gli allegati integrano o precisano il contenuto normativo della norma con delle disposizioni di dettaglio, talvolta oggetto di autonoma revisione periodica: consistono spesso in tabelle, elencazioni, riferimenti esplicativi della norma o in norme tecniche d'attuazione.

Vediamo ora le modalità attraverso cui gli atti normativi dell'ordinamento giuridico italiano vengono citati in altri documenti o nel parlato. Gli atti normativi possono citarsi o menzionarsi in forma estesa o in forma abbreviata. La forma estesa è così composta:

1. Tipo di atto: legge, decreto legislativo, decreto-legge o altro;
2. Data di promulgazione dell'atto, per esteso "GG mese AAAA" (giorno, mese in lettere, anno di 4 cifre) seguita da una virgola;
3. Numero dell'atto, preceduto da "n.";
4. Titolo dell'atto, fra virgolette e preceduto da un trattino o dalla locuzione "in materia di".

Ad esempio: Legge 22 maggio 1978, n. 194 - Norme per la tutela sociale della maternità e sull'interruzione volontaria della gravidanza, o Decreto legislativo 7 marzo 2005, n. 82 - Codice dell'amministrazione digitale

La forma abbreviata è composta da:

1. Tipo dell'atto, scritto in forma abbreviata (vedi sotto);
2. Numero dell'atto, seguito da una barra /;

### 3. Anno di promulgazione.

Ad esempio, la legge citata sopra viene indicata in forma abbreviata scrivendo: L. 194/1978, o, nel parlato, dicendo “legge 194 del 1978”.

Qui di seguito si riportano alcune sigle da usare per le abbreviazioni relative ai vari tipi di leggi e altri atti aventi valore di legge nell’ordinamento giuridico italiano:

Cost. = Costituzione della Repubblica;  
D.L. = decreto legge;  
D.Lgs. = decreto legislativo;  
D.M. = decreto ministeriale;  
D.P.C.M. = decreto del Presidente del Consiglio dei Ministri;  
D.P.R. = decreto del Presidente della Repubblica;  
L. = legge;  
L. costituzionale = legge costituzionale;  
L.R. = legge regionale;  
R.D. = regio decreto;  
R.D.L. = regio decreto legge;  
R.D.Lgs. = regio decreto legislativo;  
D.G.R. = deliberazione della giunta regionale;  
Reg. = regolamento.

#### 1.2.2. I documenti giurisprudenziali

---

I documenti giurisprudenziali sono quelli che danno conto di provvedimenti di organi dotati di potestà giurisdizionale. Tipici provvedimenti di questo tipo sono le sentenze giudiziali, ma sono talora pure rilevanti per la ricerca documentaria ordinanze e decreti. Tradizionalmente, si afferma che mentre gli atti normativi sono perlopiù generali e astratti, gli atti giurisdizionali sono perlopiù individuali e concreti. Ciò significa che i provvedimenti giurisdizionali non dettano una disciplina riguardante una classe aperta di individui e di situazioni, ma contengono delle prescrizioni relative a soggetti e condotte individuati singolarmente (classi chiuse), sulla base di un’attività applicativa di norme generali e astratte (leggi, regolamenti ecc.).

Tra i provvedimenti giurisprudenziali sono inclusi quelli emessi da tribunali internazionali o comunitari, dalle corti superiori nazionali (Corte costituzionale e Corte di cassazione), e dai tribunali di ogni ordine e grado. Possono riguardare la giustizia ordinaria, quella amministrativa, quella tributaria, quella militare ecc.

Vediamo di seguito come si presenta un documento giurisprudenziale tipico quale una sentenza della Corte di Cassazione italiana. La struttura standard di una sentenza di tale corte prevede i seguenti elementi:

<b>Estremi del provvedimento</b>		Sentenza Cassazione civile sez. III, 24/09/2008, (ud. 24/09/2008, dep. 05/10/2008), n.29030
<b>intestazione</b>	<b>Ente giudicante</b>	Repubblica Italiana In nome del Popolo Italiano La Corte Suprema di Cassazione Sezione terza civile
	<b>Collegio giudicante</b>	Composta dagli Ill.mi Sigg.ri Magistrati: Dott. Giuseppe X - primo presidente Dott. Antonio Y - presidente di sezione Dott. Gennaro Z – consigliere (omissis) Ha pronunciato la seguente SENTENZA
	<b>Parte Attrice</b>	sul ricorso iscritto al n. 35566/18 proposto da: YOGHI s.r.l., in persona dei rispettivi legali rappresentanti pro-tempore...
	<b>Parte convenuta</b>	Parte ricorrente: Bubu s.r.l., in persona del legale rappresentante, e domiciliato in Roma, Piazza della Repubblica presso lo studio dell'avvocato F.P.
<b>SVOLGIMENTO DEL PROCESSO</b>	<b>Fatto (spesso indicato con la formula "ritenuto in fatto")</b>	FATTI DI CAUSA  Con la sentenza ora denunciata, la Corte d'appello di Milano confermava la sentenza del Tribunale della stessa sede, che aveva rigettato la domanda proposta da C.F., lettore di madrelingua inglese, contro l'Università degli studi di Milano e contro l'INPS - per ottenere, previo accertamento della unitarietà del dedotto rapporto di lavoro subordinato con l'Università, a far tempo dalla data della prima assunzione, il riconoscimento del proprio diritto al trattamento economico corrispondente al livello retributivo del professore associato a tempo in base, essenzialmente, ai rilievi seguenti: <i>(omissis)</i>
	<b>Diritto (spesso indicato con la fomrula "considerato in diritto")</b>	RAGIONI DELLA DECISIONE  1. Con il primo motivo di ricorso - denunciando violazione e falsa applicazione di norme di diritto (art. 2909 c.c.; e artt. 324, 116 e 416 c.p.c.), nonché vizio di motivazione (art. 360 c.p.c., nn. 3 e 5) - C.F. censura la sentenza impugnata - per avere "negato la rilevanza dei precedenti giudicati intercorsi tra le stesse parti ed aventi ad oggetto, oltre alla natura subordinata ed unitaria ab origine del rapporto di lavoro di cui è causa (...)". Il primo motivo di ricorso è fondato. L'accoglimento - che ne consegue - assorbe gli altri motivi dello stesso ricorso, in quanto suppongono - talora esplicitamente - il rigetto, nel merito, del primo motivo... <i>(omissis)</i>
<b>Dispositivo</b>		P.Q.M. La Corte accoglie il ricorso; Cassa la sentenza impugnata con rinvio alla Corte d'appello di Brescia, anche per il regolamento delle spese di questo giudizio di cassazione.
<b>Data udienza</b>		Così deciso in Roma, nella Camera di consiglio, dalla Sezione Terza Civile della Corte di Cassazione, il 24 settembre 2008
<b>Data deposito</b>		Depositato in Cancelleria il 5 ottobre 2008

Fig. 2: struttura standard di una sentenza della Cassazione

- **Estremi del provvedimento**, sono le informazioni minime necessarie all'identificazione del provvedimento, ossia il tipo di provvedimento, l'ente giudicante, il numero e l'anno di

deposito.

- L'**intestazione**, che reca informazioni più estese su:
  - L' *ente* giudicante, con l'indicazione della circoscrizione territoriale e/o della sezione;
  - i *nomi dei componenti del collegio* se trattasi di un organo collegiale, con indicazione dei ruoli rivestiti nell'ambito del processo, o il nome del magistrato designato in presenza di una sentenza emessa dal giudice monocratico; si tratta di un elemento di rilevanza pressoché nulla in sede di ricerca documentaria.
  - L'indicazione del *tipo di provvedimento* (sentenze, ordinanze, decreti ecc.).
  - Sempre nell'intestazione sono contenute le informazioni relative alle *parti del processo*: parte attrice e parte convenuta, coloro cioè che intentano una causa o che resistono in giudizio, con l'indicazione dei rispettivi legali rappresentanti. Anche questo, è solitamente un elemento di importanza nulla ai fini della ricerca.
- Lo **svolgimento del processo**, a sua volta suddiviso in:
  - *fatto*, ovvero la rappresentazione sintetica della fattispecie concreta sottoposta all'esame del giudice,
  - *diritto*, che contiene le argomentazioni giuridiche seguite dal giudice nel prendere la decisione e le norme e i principi di diritto applicati.
- Il **dispositivo** contiene il comando con cui il giudice esprime il proprio potere giurisdizionale e può consistere in una condanna, in un rigetto, o in una diversa intimazione o disposizione, che varia in funzione dell'autorità giudicante. Il dispositivo segue di solito all'acronimo "p.q.m.", che sta per "per questi motivi".
- le **date**: quella dell'udienza, relativa quindi al giorno in cui s'è stato celebrato il dibattimento, e quella del deposito, giorno in cui la cancelleria (o la segreteria, nel caso della Corte Costituzionale), acquisito l'atto dal giudice, lo rende pubblico attraverso la sua registrazione. Quest'ultima data è quella più rilevante per la ricerca del documento.

Le sentenze si menzionano per esteso riportando il tipo di provvedimento, l'ente giudicante, la data del deposito in formato GG/MM/AAAA e il numero. Ad esempio:

Sentenza Cassazione Civile, sez. III, 05 settembre 2008, n. 17768

Sentenza Corte Costituzionale, 23 novembre 2007, n. 401

La menzione abbreviata segue diversi standard che includono comunque l'ente giudicante, il numero e almeno le ultime due cifre dell'anno in cui il provvedimento è stato depositato. Ad esempio:

Cass. Civ. I sez. n. 8806/17

La quasi totalità delle sentenze della Corte di Cassazione e della Corte Costituzionale viene resa pubblica in formato **massima** da uffici che operano in seno agli stessi enti giudicanti o da editori privati. Le massime enunciano ed evidenziano in forma estremamente sintetica alcune delle ragioni che hanno guidato il giudice nel prendere la sua decisione. Le massime sono utili in fase di ricerca giacché consentono di prendere rapidamente cognizione dei contenuti di certe sentenze, eventualmente riconoscendole come rilevanti per la definizione della questione giuridica di cui ci si sta occupando. Il ricercatore potrà in questi casi procedere al reperimento dell'intero provvedimento da cui la massima è tratta. Una massima presenta la seguente struttura:

<b>Estremi</b>	Autorità: Cassazione penale sez. I Data: 14 febbraio 2008 Numero: n. 13410
<b>Testo della massima</b>	L'attività medico-chirurgica, per essere legittima, presuppone il «consenso informato» del paziente, giacché il paziente non solo deve poter scegliere tra le diverse possibilità di trattamento medico, ma può anche eventualmente rifiutare la terapia e decidere consapevolmente di interromperla, in tutte le fasi della vita, anche in quella terminale.
<b>Relazioni con sentenze precedenti</b>	Cassazione penale sez. I, 06 Feb 1998 n. 3147

Fig. 3: struttura standard di una massima della Cassazione

- gli **estremi**, in cui sono presenti tutti i riferimenti che servono per individuare il provvedimento da cui è tratta la massima, dunque i dati relativi all'ente giudicante, sezione giudicante, al numero della sentenza e alla data del deposito.
- Il **testo della massima**.
- La massima si conclude sovente con la segnalazione di eventuali provvedimenti o altre massime di contenuto conforme o comunque rilevante secondo il redattore della massima.

### 1.2.3. I documenti dottrinari

La dottrina giuridica è quell'insieme di teorie sviluppato dai giuristi a commento della regolazione giuridica e della sua applicazione ai casi concreti.

A differenza di quanto avviene per i documenti legislativi e giurisprudenziali, l'informazione dottrina non si presenta in forma standardizzata, ma trova stanza in una serie variegata di pubblicazioni che vanno dai manuali e trattati alle opere monografiche, alle raccolte collettanee, alle riviste, ai commentari legislativi e giurisprudenziali.

Ne sono esempio i *manuali* i quali offrono al lettore, espone in forma sintetica, le nozioni basilari di una determinata disciplina giuridica: vengono solitamente utilizzati come libri di testo per la preparazione degli esami universitari o concorsi, e presentano nel titolo locuzioni quali "Istituzioni (principi, corso) di diritto ... (materia)". I *trattati*, di contenuto più dettagliato dei manuali, analizzano invece singoli istituti giuridici o determinate partizioni di una disciplina.

Tra le *opere collettanee*, redatte da più autori, incontriamo le *enciclopedie* (e i *digesti*), opere organizzate per voci che contengono il contributo di più autori, di solito tante quante sono le voci e le sotto-voci sviluppate nei volumi: classico esempio è rappresentato dall'*Enciclopedia del diritto* che si articola in circa 40 volumi più gli indici cumulativi. Anche i *commentari* (e le *rassegne*), in quanto contengono contributi diversi autori a margine di articoli di legge o di codice, possono essere annoverati tra le opere collettanee. Insieme alle opere appena menzionate, che contengono scritti inseriti in progetti editoriali ad ampio raggio, troviamo anche volumi frutto della collaborazione di più autori che possono confluire in *atti di convegno*, *annali di facoltà*, *studi*, o *scritti in onore* (o in memoria).

Le *riviste* consistono in raccolte di articoli dottrinari (ma anche, talora, di atti legislativi e provvedimenti giurisprudenziali), che vengono pubblicate con cadenza periodica (settimanale, mensile, trimestrale, semestrale ed in qualche caso annuale). I contributi pubblicati nelle riviste sono sovente sottoposti a rigorosi processi di *peer review*. Vale la pena di ricordare che le riviste possono essere a contenuto interdisciplinare e raccogliere soltanto una tipologia documentale, o

riguardare un intero settore disciplinare (esempio: *Rivista di diritto civile*), una particolare materia (esempio: *Rivista di diritto tributario*) o addirittura argomenti specifici (esempio: *Il diritto di famiglia*).

Infine, le pubblicazioni contenenti la cosiddetta informazione secondaria (*repertori e bibliografia*) hanno un carattere informativo strumentale. A differenza delle riviste, che contengono documenti per esteso, qui troviamo una esposizione sistematica e ragionata delle materie giuridiche ordinate alfabeticamente dove per ogni argomento vengono fornite le notizie bibliografiche relative a pubblicazioni pertinenti. In pratica, i redattori delle bibliografie spogliano gli indici di tutte le riviste giuridiche e riportano nelle loro opere tutti i titoli dei contributi editoriali pubblicati e le informazioni necessarie per pervenire al testo per esteso dell'articolo o della monografia. Classico esempio di opera contenente informazione secondaria è il *Dizionario bibliografico delle riviste giuridiche italiane*, che viene pubblicato a cadenza annuale.

---

# 02 – Le banche dati, la ricerca per estremi e la ricerca testuale

---

## 2.1. Le banche dati: definizione e struttura

---

Ai fini di questa introduzione all'informatica giuridica documentale, le banche dati possono essere definite come "insieme di informazioni relative ad un preciso dominio di conoscenze, registrate in formato digitale e consultabili attraverso un software di ricerca"<sup>1</sup>. Si tratta in particolare, nel nostro caso, di sistemi informativi atti all'archiviazione e al reperimento di *informazioni giuridiche*.

La struttura tipica di una banca dati si compone di due parti fondamentali:

- 1) una base di dati;
- 2) un software di interrogazione della base di dati, corredato di un'*interfaccia*.

La *base di dati* è costituita da uno o più insiemi di documenti, strutturati e suddivisi in archivi in ragione del tipo di informazione contenuta. Le basi dati odierne sono spesso caratterizzate dalla c.d. **indicizzazione totale**; ciò significa che *tutti* i termini del documento costituiscono specifici elementi di richiamo di quel documento in fase di ricerca. Ciò distingue l'indicizzazione di una banca dati da quella degli archivi tradizionali, in cui gli indici prendono in considerazione solo alcuni elementi del documento. Si pensi ad uno schedario bibliografico dove gli elementi per ricercare un documento sono l'autore del documento oppure la materia indicata attraverso descrittori o parole chiave tratti da uno schema di classificazione. Abbiamo in questo caso solo due indici: i nomi degli autori e i descrittori utilizzati per delineare il tema trattato dal documento. L'accesso alle informazioni può avvenire solo attraverso questi due dati. Si tratta, ovviamente, di un criterio di indicizzazione molto parziale e limitato. Nei sistemi informatici, il criterio di indicizzazione è invece *totale* nel senso che *tutti* gli elementi del documento, dal titolo alle singole parole del testo, costituiscono chiavi di accesso e reperimento del documento stesso. È dunque possibile utilizzare come "dato esca" una qualsiasi parola, nome o numero che si ritenga siano contenuti nei documenti da reperire.

Il **software di interrogazione**, corredato da un'interfaccia grafica che facilita la comunicazione utente-computer, è finalizzato al reperimento dell'informazione ricercata dall'utente. Tale software opera confrontando i dati immessi nella ricerca (c.d. "dati-esca") con quelli presenti nella base di dati e producendo in risposta tutti, e solo, i documenti che contengono quei dati (c.d. **combaciamento**). Oggi, tuttavia, alcuni software di ricerca funzionano grazie ad algoritmi assai più sofisticati, che non si limitano a reperire ed elencare i documenti che "combaciano" con i dati-esca inseriti, ma li individuano e ordinano secondo un criterio statistico di rilevanza rispetto ai

---

<sup>1</sup> L'ordinamento giuridico italiano prevede anche una definizione legislativa di "banca dati": l'art. 2, comma 1, numero 9, della legge sul diritto d'autore (L. 633/1941) come modificata dal d. lgs. 196/1999 definisce la banca dati come «raccolta di opere, dati o altri elementi indipendenti sistematicamente o metodicamente disposti ed individualmente accessibili mediante mezzi elettronici o in altro modo».

presumibili obiettivi di ricerca dell'utente. **L'interfaccia comprende di solito una maschera di ricerca** per ciascun sottoinsieme di documenti su cui si intende svolgere la ricerca (documenti normativi, giurisprudenziali o dottrinari). La maschera di ricerca è data da una serie di *campi*, ossia di spazi nei quali è possibile immettere i dati-esca, e possono essere di tipologia variabile: parole che si ritenga ricorrano nel documento da reperire, numeri e date identificativi di leggi e sentenze, nomi di autori e così via.

## 2.2. Tipologia delle banche dati

---

In base ai *contenuti*, le banche dati giuridiche possono essere classificate come:

- a) *Testuali* (o *primarie*), se contengono unità documentali nel testo completo (leggi, sentenze, massime, eccetera);
- b) *Bibliografiche* (o *secondarie*), se contengono soltanto i riferimenti ad unità documentali, pubblicate per esteso su supporti tradizionali (archivi cartacei), su altre banche dati o in siti Internet.

Le banche dati primarie consentono di leggere direttamente e, solitamente, salvare, stampare ecc. i documenti che vengono reperiti; le banche dati secondarie contengono soltanto indicazioni utili al reperimento di quei documenti, indicandone la collocazione o altre informazioni rilevanti per il loro acquisto o consultazione,

In base alla *localizzazione* dei dati, le banche dati elettroniche si dicono:

- a) *online* quando le basi di dati risiedono fisicamente in un server remoto rispetto al dispositivo usato dall'interrogante, il quale può accedervi via internet;
- b) *offline* quando le basi di dati sono contenute in supporti di memorizzazione inseriti o interni al dispositivo usato dall'interrogante.

Le banche dati *online*, oggi decisamente più diffuse, presentando una base di dati aggiornata in modo tempestivo e costante, tuttavia la loro consultabilità è subordinata alla possibilità di accedere a Internet. Per questo motivo, possono risultare poco pratiche nelle situazioni in cui tale accesso è indisponibile o precario.

Le banche dati *offline* viaggiano col supporto su cui è memorizzata la base di dati, dunque sono consultabili indipendentemente dalla disponibilità di un accesso a Internet; tuttavia esse sono soggette ad aggiornamenti meno frequenti (di solito l'abbonamento alla banca dati prevede l'invio di aggiornamenti con cadenza che varia dai 2 ai 12 mesi), e sono quindi esposte a rischi di obsolescenza della base dati.

## 2.3. Gli indici di prestazione nella ricerca documentaria

---

Un sistema informativo, o una particolare ricerca, si dicono *efficaci* quando sono in grado di reperire *tutta e soltanto* l'informazione pertinente agli obiettivi di ricerca. Gli indici per misurare tale efficacia sono due:

- 1) La *precisione*, che indica la capacità di riportare tra i risultati della ricerca *solo* i documenti pertinenti, escludendo quelli non pertinenti. È calcolabile sulla base del rapporto tra il numero dei documenti *pertinenti* ritrovati e il totale dei documenti ottenuti in risposta dal sistema. Ad

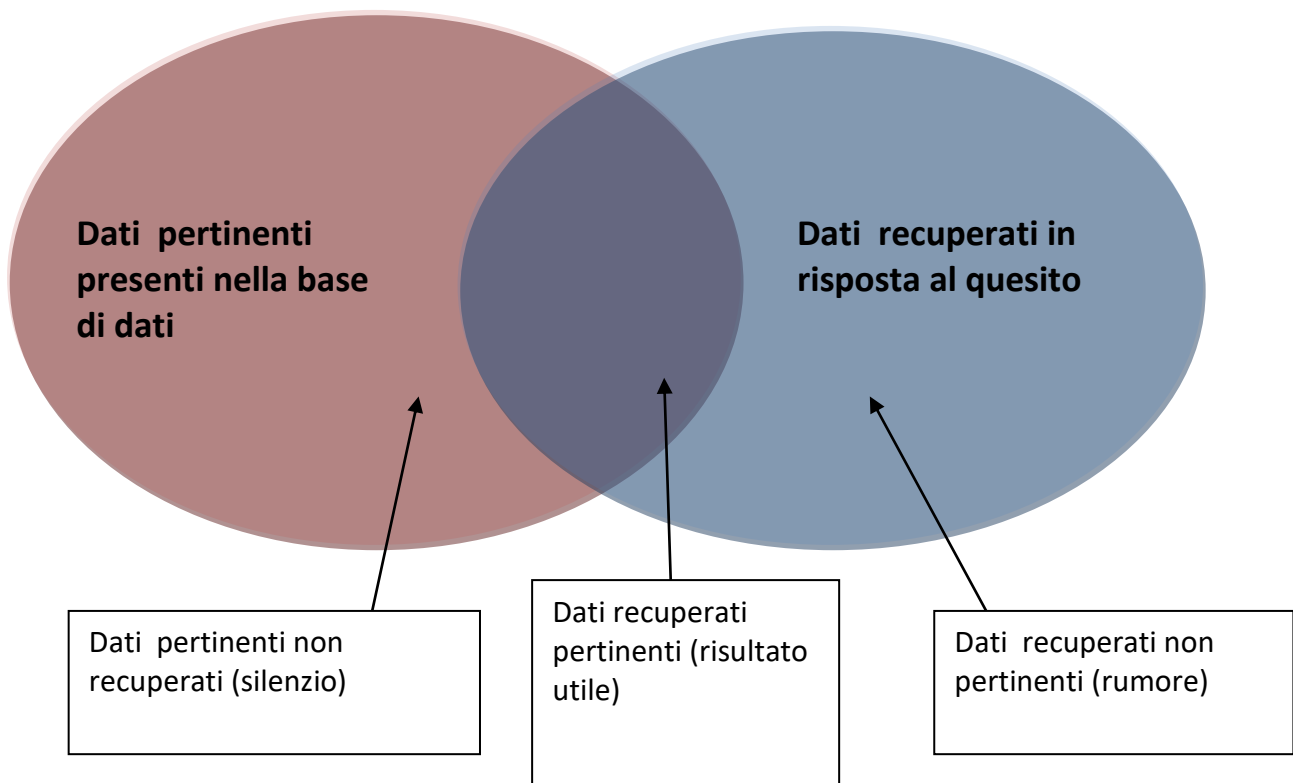
esempio, documenti ottenuti pertinenti alla ricerca: 60, totale dei documenti ottenuti in risposta: 100; indice di precisione  $60/100 = 0,6$ ;

- 2) il *richiamo*, che indica la capacità di reperire *tutti* i documenti pertinenti tra quelli presenti nell'archivio. È dato dal rapporto tra il numero dei documenti pertinenti ritrovati e il totale dei documenti pertinenti contenuti nella base di dati. Ad esempio, documenti pertinenti ottenuti: 60, documenti pertinenti presenti nella base dati: 100; indice di richiamo  $60/100 = 0,6$ .

I fenomeni del richiamo e della precisione possono essere colti dalla prospettiva complementare di chi sia interessato a considerare non già l'efficacia, bensì l'*inefficacia* di un sistema informativo o di una ricerca particolare. Si ottengono così altri due indici:

- 3) Il *rumore*, ossia il difetto di precisione, indica la quantità di documenti *non pertinenti* che il sistema presenta come risultato della ricerca. Risulta dal numero di documenti *non pertinenti* ottenuti diviso per il numero totale dei documenti ottenuti in risposta dall'elaboratore. Ad esempio: documenti non pertinenti: 40, totale documenti ottenuti in risposta: 100; indice di rumore  $40/100 = 0,4$ .
- 4) Il *silenzio*, ossia il difetto di richiamo, esprime l'incapacità di reperire i documenti pertinenti contenuti nell'archivio consultato. Può calcolarsi dividendo il numero di documenti pertinenti che, pur contenuti nell'archivio consultato, non sono stati selezionati per il numero dei documenti pertinenti presenti nell'archivio stesso. Ad esempio: documenti pertinenti non compresi nella risposta 40, totale documenti pertinenti contenuti nella base dati: 100; indice di silenzio  $40/100 = 0,4$ .

È possibile impiegare i termini "rumore", "rumorosa" e "silenzio", "silenziosa" per designare/connotare il risultato indesiderabile di una ricerca che ha prodotto, rispettivamente, troppi risultati non pertinenti e pochi risultati pertinenti. Segue una rappresentazione grafica dei fenomeni del rumore e del silenzio.



## 2.4. *La ricerca per estremi e la ricerca testuale*

---

Le odierne banche dati giuridiche consentono due tipi di ricerca:

- a) la *ricerca per estremi*, che opera attraverso l'immissione dei dati alfanumerici che identificano univocamente il documento da reperire;
- b) la *ricerca testuale*, che opera attraverso l'immissione di una o più parole che si presume ricorrano nel documento da reperire.

La ricerca per estremi non richiede particolari conoscenze tecniche: sarà sufficiente inserire nei campi appositi il tipo, il numero e la data dell'atto o del provvedimento che si intende reperire (ad esempio, indicando nell'apposito campo il tipo "legge", il numero "300" e l'anno "1970"). Si ricorre a questo tipo di ricerca quando si dispone già delle conoscenze relative agli estremi che identificano il documento da trovare. Molte odierne banche dati facilitano le ricerche per estremi consentendo l'individuazione di alcuni documenti, specie normativi, mediante l'immissione delle denominazioni con cui tali atti vengono solitamente designati nella prassi. È dunque possibile ad esempio ricercare la legge 300/1970 indicando nell'apposito campo (spesso denominato "nome volgare" o simili) "statuto dei lavoratori", o trovare il d. lgs. 196/2003 indicando "codice della privacy" e così via.

Quando si intende reperire della documentazione di cui si ignorano gli estremi, oppure quando si intende estendere la selezione dei documenti non limitandola quelli già noti, si ricorre alla *ricerca testuale*, che opera attraverso l'immissione in appositi campi di termini che si ritiene siano presenti nel titolo o nel testo dei documenti da reperire. I dati-esca che vengono così introdotti, eventualmente combinati attraverso gli operatori che saranno descritti nei prossimi paragrafi, costituiscono la c.d. *stringa di ricerca*. La formulazione di un'adeguata stringa di ricerca è la chiave per svolgere ricerche che ottengano successo sia sotto il profilo della *precisione* sia sotto il profilo del *richiamo*.

### 2.4.1. *Polisemia, sinonimia, infettibilità*

---

La *polisemia* è il fenomeno linguistico che si verifica quando una parola viene intesa in più di un significato. Ad esempio, il termine "costituzione" assume diversi significati a seconda che sia usato per designare la legge fondamentale di uno stato (es. "Costituzione della Repubblica"), la complessione fisica di un individuo (es. "Gianni è un ragazzo di costituzione robusta") l'atto con cui la parte fa ingresso nel processo (costituzione in giudizio), l'atto che determina la nascita di una società (es. costituzione di una s.p.a), ecc. L'immissione in fase di ricerca del termine "costituzione", produrrà dunque in risposta tutti, e solo, i documenti che contengono quel termine, indipendentemente dai significati che esso assume nei vari contesti. Il risultato della ricerca è pertanto in questi casi esposto al rischio di *rumore*, giacché comprende documenti in cui quel termine ricorre con significati diversi da quello rilevante ai fini della ricerca (ad esempio, si cercavano documenti relativi alla costituzione di una società ma si reperiscono anche molti documenti relativi alla "costituzione in giudizio" o alla costituzione italiana).

La *sinonimia* è il fenomeno linguistico che si verifica quando più parole indicano un significato identico in ogni contesto (sinonimia totale, es: "fideiussione", "fidejussione") o in alcuni contesti (sinonimia parziale, es. "vendita", "compravendita"). L'immissione in fase di ricerca di un termine

come “fideiussione” produrrà in risposta i documenti che contengono quel termine, scritto esattamente in quel modo, ma *non* i documenti che comprendono il termine “fidejussione”, giacché il sistema prende in considerazione i termini in quanto mere sequenze di caratteri, e non in quanto entità linguistiche dotate di un certo significato. Il risultato della ricerca è pertanto in questi casi esposto a un elevato rischio di *silenzio*, poiché non comprende alcuni documenti che pure erano rilevanti ai fini della ricerca stessa.

Per ovviare alle difficoltà derivanti dalla polisemia e dalla sinonimia dei dati-esca, è possibile formulare stringhe di ricerca complesse, combinando più termini attraverso gli operatori di cui si parlerà nella prossima dispensa.

---