
02– La ricerca testuale

2.1. Gli operatori booleani

La stringa di ricerca può essere costituita da **un solo termine** o da **più termini**. Nel primo caso, l'interrogazione viene effettuata mediante la specificazione di una sola condizione, che viene soddisfatta dalla ricorrenza nel documento del termine che costituisce la stringa di ricerca. Così, una stringa di ricerca come “diffamazione” produrrà come risultato tutti e solo i documenti presenti nella base dati che soddisfino la condizione data dalla ricorrenza della parola “diffamazione”. L'uso di stringhe costituite da un solo termine, tuttavia, è solitamente sconsigliato, giacché la possibile presenza di sinonimi e di molteplici significati di quel termine espone la ricerca al rischio, congiunto, di silenzio e di rumore¹. Si pensi a una ricerca sull'istituto della separazione giudiziale operata immettendo come dato esca soltanto la parola “**separazione**”. Si otterrà un risultato al tempo stesso:

- a) rumoroso, giacché tra i documenti ottenuti in risposta ve ne saranno molti che alludono ad altri sensi del termine “separazione”, ad esempio al regime della separazione dei beni nel matrimonio, alla separazione dei beni del defunto da quelli dell'erede ex art. 512 c.c., alla separazione dei frutti naturali dalla cosa madre ecc.;
- b) silenzioso, giacché i documenti ottenuti in risposta non comprenderanno quelli che, pur non contenendo la parola “separazione”, contengono le locuzioni “coniuge separato”, “separati”, “separare” ecc., e che possono dunque essere rilevanti per gli obbiettivi della ricerca in quanto contengono informazioni relative all'istituto che si intende indagare.

Per evitare inconvenienti del genere è possibile formulare delle stringhe di ricerca facendo un uso accorto di *più* dati-esca, con lo scopo di precisare e rendere più completo il risultato dell'interrogazione. Questi dati esca possono essere combinati in vario modo attraverso alcuni termini speciali, noti come *operatori di relazione*. Tali termini vengono inseriti, uno per volta, *tra* le parole che compaiono nella stringa di ricerca, e consentono di specificare una serie di condizioni logiche relative ai documenti che ne costituiscono il risultato. I più importanti operatori di relazione sono i cosiddetti *operatori logici o booleani*²: AND, OR e NOT.

2.1.1. L'operatore AND

L'operatore AND viene inserito tra due o più termini quando si vuole che questi siano *tutti* presenti in ciascuno dei documenti che costituiscono il risultato della ricerca. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

¹ Un risultato può essere, al tempo stesso, silenzioso e rumoroso quando ricomprende solo una parte dei documenti rilevanti presenti nella base dati e tuttavia include una gran quantità di documenti non rilevanti.

² Da George Boole, filosofo e matematico inglese del XIX secolo, cui si deve l'invenzione del calcolo logico noto come *algebra proposizionale di Boole*. Boole notò che le combinazioni tra proposizioni obbediscono a leggi ben definite, di cui si può dare una formulazione precisa tanto quanto la definizione delle regole che governano le operazioni dell'aritmetica.

affitto AND locazione

produce in risposta tutti, e solo, i documenti presenti nella base dati in cui ricorrono *sia* il termine «affitto» *sia* il termine «locazione». Non vengono invece compresi nel risultato della ricerca:

- a) I documenti che non contengono né il termine «affitto» né il termine «locazione»;
- b) I documenti che contengono il termine «affitto» ma non il termine «locazione»;
- c) I documenti che contengono il termine «locazione» ma non il termine «affitto».



Rappresentazione grafica degli insiemi di documenti selezionati dalla stringa: affitto AND locazione

L'uso dell'operatore AND *restringe* la selezione dei documenti prodotti in risposta dal sistema: tanti più termini vengono combinati in AND, tanto meno numerosi saranno i documenti selezionati. Si tratta dunque di un operatore tipicamente adoperato per limitare il *rumore* della ricerca, e/o per contrastare gli effetti della *polisemia*. Una ricerca sulla nozione di “sana e robusta costituzione”, ad esempio, potrà essere resa meno rumorosa se la stringa di ricerca, in luogo del solo termine polisemico “costituzione”, contiene anche, combinati in AND, i termini “sana” e “robusta”. Una stringa come:

sana AND robusta AND costituzione

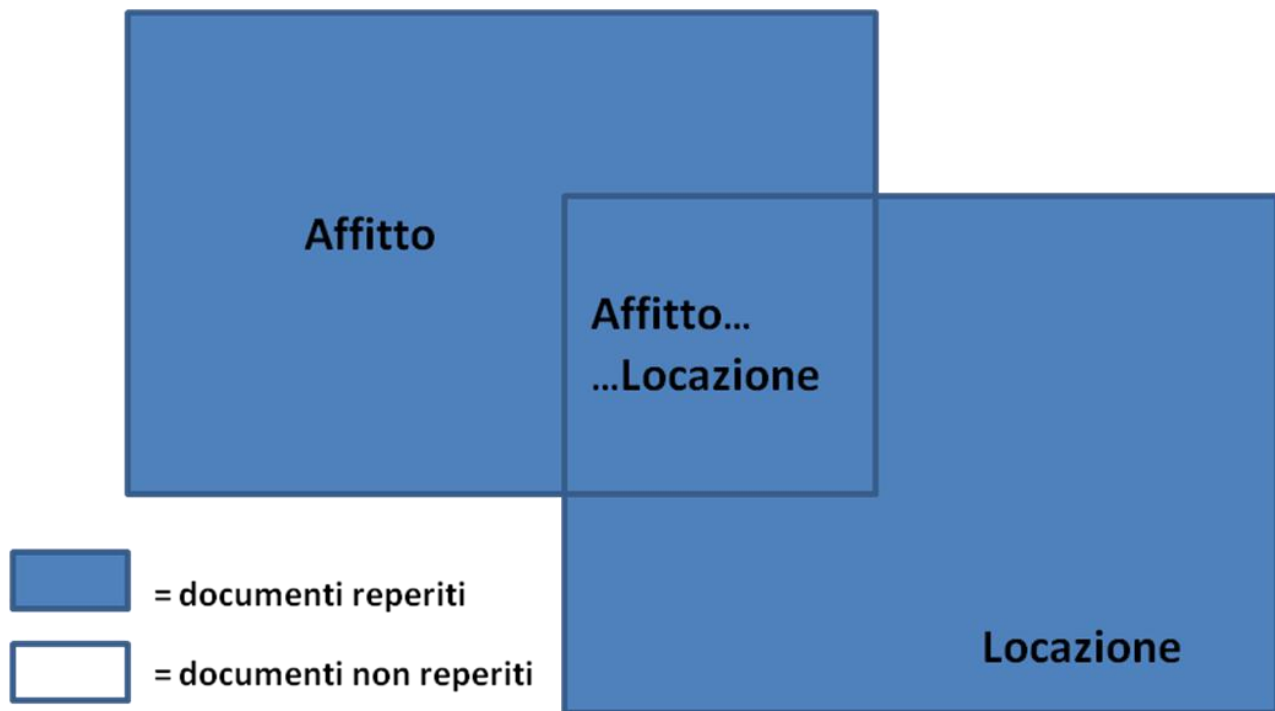
invero, produce in risposta tutti, e solo, i documenti in cui ricorrono contemporaneamente le parole “sana”, “robusta” e “costituzione”; pertanto si reperiranno documenti che riguardano la costituzione intesa come salute fisica dell'individuo, piuttosto che documenti che attengono alla costituzione intesa come atto processuale, legge fondamentale o creazione di una società.

2.1.2. L'operatore OR

L'operatore OR viene inserito tra due o più termini quando si vuole che *almeno uno di essi* sia presente in ciascuno dei documenti che costituiscono il risultato della ricerca. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

affitto OR locazione

produce in risposta tutti, e solo, i documenti presenti nella base dati in cui ricorre il termine «affitto», tutti i documenti presenti nella base dati in cui ricorre il termine «locazione» oltretché naturalmente tutti i documenti in cui ricorrono sia il termine «affitto» sia il termine «locazione». Vengono invece esclusi dal risultato della ricerca i documenti in cui non compaiono né il termine «affitto» né il termine «locazione».



Rappresentazione grafica degli insiemi di documenti selezionati dalla stringa: affitto OR locazione

L'uso dell'operatore OR *estende* la selezione dei documenti prodotti in risposta dal sistema: tanto maggiore è il numero dei termini combinati in OR, tanto più elevato è il numero dei documenti prodotti in risposta dal sistema. OR, pertanto, permette di contrastare il *silenzio* della ricerca e/o gli effetti della *sinonimia*. Una ricerca sull'istituto della compravendita effettuata mediante l'introduzione di una stringa di ricerca come

compravendita

infatti, è silenziosa nella misura in cui esclude dai risultati della ricerca tutti quei documenti in cui quell'istituto è designato tramite il termine sinonimo "vendita", e che pure erano rilevanti ai fini della ricerca. L'uso della combinazione

vendita OR compravendita

consente di eliminare questo problema, giacché produce un risultato che comprende tutti, e solo, quei documenti in cui ricorrono, anche disgiuntamente, il termine "vendita" e il termine "compravendita".

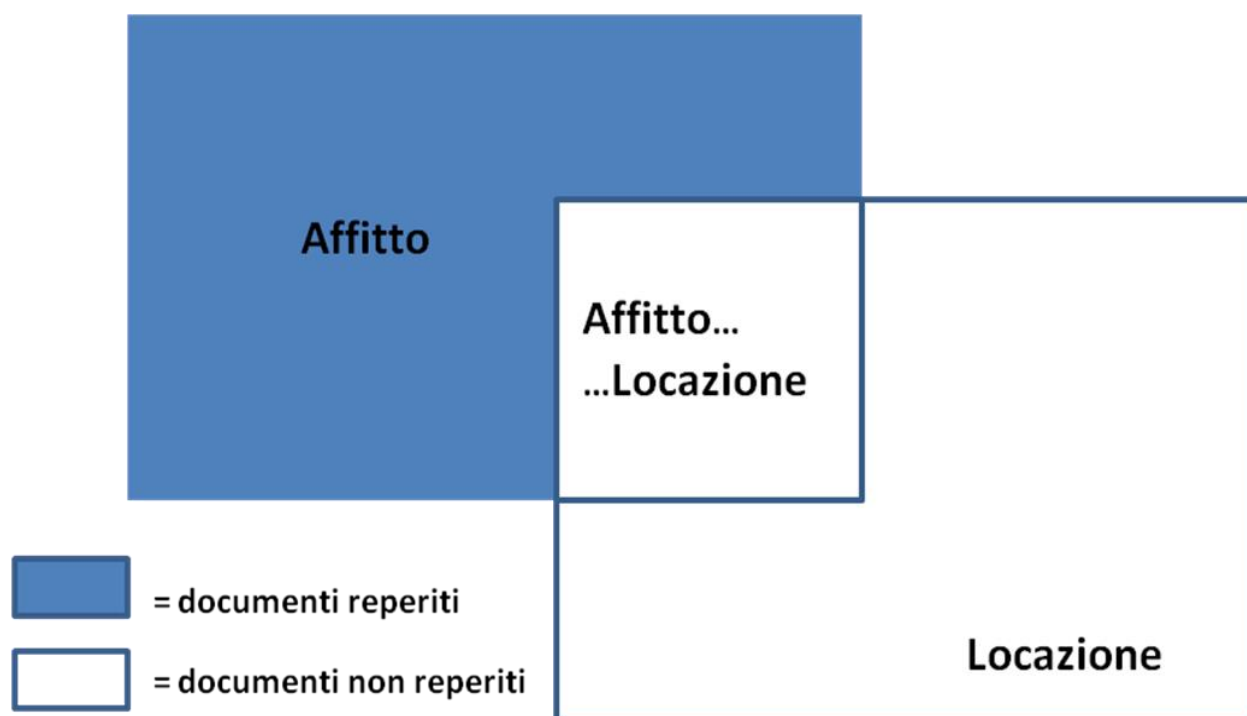
2.1.3. L'operatore NOT

L'operatore NOT viene inserito prima di un termine quando si vuole che la ricerca produca in risposta dei documenti in cui *non* compare quel termine. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

affitto NOT locazione

produce in risposta tutti, e solo, i documenti presenti nella base dati in cui ricorre il termine «affitto» ma non il termine «locazione». Vengono invece esclusi dal risultato della ricerca:

- I documenti in cui non compaiono né il termine «affitto» né il termine «locazione»;
- I documenti in cui compaiono sia il termine «affitto» sia il termine «locazione»;
- I documenti in cui non compare il termine «affitto» ma compare il termine «locazione».



Rappresentazione grafica degli insiemi di documenti selezionati dalla stringa: affitto NOT locazione

Anche l'operatore NOT viene solitamente adoperato per contrastare il rumore delle ricerche e/o il fenomeno della polisemia delle parole. Si supponga ad esempio che un ricercatore intenda reperire documenti riguardanti la nozione di inquinamento inteso come fonte di degrado ambientale. Una stringa di ricerca come:

inquinamento

tuttavia, produrrebbe un risultato che include anche documenti in cui il termine «inquinamento» viene associato al termine «probatorio», e quindi inteso come attentato

all'acquisizione o alla genuinità della prova in sede processuale. Per evitare questo rumore, è possibile adoperare una stringa come:

inquinamento NOT probatorio

che esclude dal risultato della ricerca tutti i documenti in cui ricorre il termine "probatorio", che assai probabilmente contengono informazioni non rilevanti ai fini della ricerca.

2.2. La sintassi delle stringhe di ricerca

La sintassi corretta delle stringhe di ricerca prevede che due termini possano essere combinati da un solo operatore booleano. È dunque sintatticamente scorretta (e produce risultato nullo) la stringa di ricerca

affitto OR AND locazione

mentre sono sintatticamente corrette le stringhe:

affitto AND locazione

affitto OR locazione

affitto OR locazione AND immobiliare

2.2.1. L'uso delle parentesi

È possibile elaborare stringhe di ricerca complesse, che includono molti dati-esca combinati in vario modo. Sorge però l'esigenza di comprendere l'ordine in cui il software di interrogazione combina i termini inseriti nella stringa di ricerca. In mancanza di una specificazione di tale ordine, infatti, una stringa come:

affitto OR locazione AND immobiliare

sarebbe ambigua, in quanto potrebbe essere interpretata come:

- 1) "trova tutti e solo i documenti in cui compare il termine «affitto», i documenti in cui compaiono sia il termine «locazione» sia il termine «immobiliare», e i documenti in cui compare sia il termine «affitto», sia il termine «locazione» sia il termine «immobiliare»";
- 2) "trova tutti e solo i documenti in cui compaiono sia il termine «affitto» sia il termine «immobiliare», i documenti in cui compaiono sia il termine «locazione» sia il termine «immobiliare», e i documenti in cui compare sia il termine «affitto», sia il termine «locazione» sia il termine «immobiliare»;

La differenza tra le due interpretazioni è evidente: nel secondo caso, infatti, vengono esclusi dal risultato della ricerca i documenti in cui compare il termine «affitto» ma non compare il termine «immobiliare».

Per evitare ambiguità del genere, i software di interrogazione, prendono in considerazione le operazioni secondo un ordine prestabilito per *default*. Solitamente, l'ordine in cui il sistema

raggruppa le operazioni è il seguente: prima vengono considerate le combinazioni in AND, poi le combinazioni in NOT e infine le combinazioni in OR ³. Più precisamente:

- a) Tra AND e OR prevale AND. Il sistema, cioè, combina per prima cosa i termini combinati in AND, trattandoli come se fossero inclusi tra parentesi. Ciò indipendentemente dalla posizione che la combinazione occupa nella stringa di ricerca. Ad esempio, una stringa come:

atto AND vendita OR compravendita

viene interpretata dal sistema come:

(atto AND vendita) OR compravendita

e cioè come un'istruzione volta a far reperire al sistema tutti e solo i documenti che contengono:

- 1) Sia la parola "atto" sia la parola "vendita";
- 2) La parola "compravendita";
- 3) Sia la parola "atto," sia la parola "vendita" sia la parola "compravendita.

La stringa non viene invece interpretata dal sistema come:

atto AND (vendita OR compravendita)

giacché la combinazione in AND è prioritaria rispetto alla combinazione in OR. Quest'ultima interpretazione avrebbe prodotto un risultato di ricerca diverso giacché, diversamente dall'interpretazione precedente, avrebbe escluso i documenti in cui ricorre la parola "compravendita" ma non la parola "atto".

Allo stesso modo, una stringa come:

atto OR vendita AND compravendita

viene interpretata dal sistema come:

atto OR (vendita AND compravendita)

e cioè come un'istruzione volta a far reperire al sistema tutti e solo i documenti che contengono:

- 1) Sia la parola "vendita" sia la parola "compravendita";
- 2) La parola "atto";
- 3) Sia la parola "atto," sia la parola "vendita" sia la parola "compravendita.

- b) Tra NOT e AND, prevale quello più a sinistra. Il sistema, cioè, combina per prima cosa le coppie di termini che compaiono più a sinistra nella stringa di ricerca, trattando i termini come se fossero inclusi tra parentesi. Ad esempio, una stringa come:

atto AND vendita NOT compravendita

³ Ciò, almeno, è quanto accade nella maggior parte delle banche dati in commercio. Poiché peraltro questa regola ha numerose eccezioni, è opportuno consultare la guida della singola banca dati in ordine alla priorità adoperata. In alternativa, è naturalmente possibile disambiguare le stringhe di ricerca inserendo manualmente le parentesi (vedi *infra*, questo stesso paragrafo).

viene interpretata dal sistema come:

(atto AND vendita) NOT compravendita;

e cioè come un'istruzione volta a far reperire al sistema tutti e solo i documenti che contengono sia la parola "atto" sia la parola "vendita", a condizione che non compaia anche la parola "compravendita". D'altro canto, una stringa come:

atto NOT vendita AND compravendita

viene interpretata dal sistema come:

(atto NOT vendita) AND compravendita

e cioè come un'istruzione volta a far reperire al sistema tutti e solo i documenti che contengono sia la parola "atto" sia la parola "compravendita", a condizione che non compaia anche la parola "vendita".

- c) Tra NOT e OR prevale NOT. Il sistema, cioè, combina per prima cosa le coppie di termini combinati in NOT, trattando i termini come se fossero inclusi tra parentesi. Ad esempio, una stringa come:

atto OR vendita NOT compravendita

viene interpretata dal sistema come:

atto OR (vendita NOT compravendita)

e cioè come un'istruzione volta a far reperire al sistema tutti e solo i documenti che contengono:

- 1) I documenti in cui compare la parola "atto";
- 2) I documenti in cui compare la parola "vendita", a patto che non compaia anche la parola "compravendita";
- 3) i documenti in cui compaiono sia la parola "atto" sia la parola "vendita", a patto che non compaia anche la parola "compravendita".

Similmente, una stringa come:

atto NOT vendita OR compravendita

viene interpretata dal sistema come:

(atto NOT vendita) OR compravendita

e cioè come un'istruzione volta a far reperire al sistema tutti e solo i documenti che contengono:

- 1) I documenti in cui compare la parola "atto", a patto che non compaia anche la parola "vendita";
- 2) I documenti in cui compare la parola "compravendita";

- 3) i documenti in cui compaiono sia la parola “atto” sia la parola “compravendita”, a patto che non compaia anche la parola “vendita”.

Per specificare un ordine diverso da quello predefinito dal software di interrogazione, il ricercatore può fare uso delle *parentesi*: i termini racchiusi tra parentesi vengono infatti combinati con un ordine prioritario rispetto alle altre combinazioni presenti nella stringa di ricerca. Ad esempio, si è detto che una stringa come:

atto AND vendita OR compravendita

viene interpretata dal sistema come un’istruzione volta a selezionare tutti, e solo:

- 1) I documenti che contengono sia il termine “atto” sia il termine “vendita”;
- 2) I documenti che contengono il termine “compravendita”
- 3) I documenti che contengono sia il termine “atto” sia il termine “vendita” sia il termine “compravendita”.

Se invece, tramite l’uso delle parentesi, si istruisce il sistema a combinare per primi i termini collegati in OR, scrivendo:

atto AND (vendita OR compravendita)

allora il risultato della ricerca comprenderà tutti, e solo:

- 1) I documenti che contengono sia il termine “atto” sia il termine “vendita”;
- 2) I documenti che contengono sia il termine “atto” sia il termine “compravendita”;
- 3) I documenti che contengono sia il termine “atto” sia il termine “vendita” sia il termine “compravendita”.

La seconda stringa, diversamente dalla prima, lascia fuori dalla selezione i documenti che contengono la parola “compravendita” ma non contengono la parola “atto”.

Alla luce di quanto detto sopra, le parentesi sono superflue quando i termini tra esse racchiusi sono collegati mediante AND, dato che il sistema tratta i termini collegati in AND come se fossero comunque compresi tra parentesi. Pertanto, una stringa come:

inadempimento AND (impossibilità AND sopravvenuta)

individua esattamente gli stessi risultati di una stringa come:

inadempimento AND impossibilità AND sopravvenuta.

2.3. Altri operatori per la ricerca

Gli operatori AND, OR e NOT sopra esaminati sono gli operatori di relazione classici. Vi sono peraltro degli altri operatori di relazione, denominati *operatori di prossimità*, che operano come

degli AND rafforzati, nel senso che non si limitano a richiedere la compresenza dei dati-esca nei documenti che costituiscono il risultato della ricerca, ma vengono utilizzati per chiedere al sistema di fornire soltanto i documenti in cui i dati-esca sono presenti ad una certa distanza tra loro, cioè adiacenti o vicini. Questi operatori di prossimità sono:

ADJ (operatore di adiacenza).

Viene inserito tra due o più termini quando si vuole che questi siano *tutti* presenti, *in posizione di adiacenza*, in ciascuno dei documenti che costituiscono il risultato della ricerca. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

ordine ADJ pubblico

produce in risposta tutti, e solo, i documenti presenti nella base dati in cui ricorrono i termini «ordine» e «pubblico» l'uno accanto all'altro, ossia i documenti in cui compare la frase «ordine pubblico»⁴. Se i due termini dell'esempio fossero stati combinati in AND, la ricerca avrebbe selezionato tutti i documenti che contengono in qualsiasi parte del testo i termini «ordine» e «pubblico». sarebbero dunque stati inclusi nel risultato della ricerca non solo documenti che contengono riferimenti al concetto di ordine pubblico, ma anche documenti relativi, ad esempio, all'*ordine* impartito da un *pubblico* ufficiale, all'*ordine* del disbrigo delle pratiche in un ufficio *pubblico*, e così via.

NEAR (operatore di vicinanza)⁵. Viene inserito tra due o più termini quando si vuole che questi siano *tutti* presenti, *in posizione di vicinanza*, in ciascuno dei documenti che costituiscono il risultato della ricerca. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca:

impossibilità NEAR sopravvenuta

produce in risposta tutti, e solo, i documenti presenti nella base dati in cui ricorrono i termini «impossibilità» e «sopravvenuta» in posizione di vicinanza, ossia i documenti in cui quei termini compaiono intervallati da un certo numero di altre parole (solitamente 5, 7 o 15: dipende dalla banca dati utilizzata)⁶. Saranno pertanto compresi nel risultato della ricerca documenti relativi alla "sopravvenuta impossibilità", all'"impossibilità sopravvenuta", all'"impossibilità non sopravvenuta", all'"impossibilità di adempiere per causa sopravvenuta" e via dicendo.

"..." (virgolette, ricerca per frase esatta). Quando nel campo di ricerca testuale viene inserita una frase racchiusa da virgolette, il risultato della ricerca comprenderà tutti, e solo, i documenti che contengono quella frase così com'è stata digitata, comprensiva di articoli e preposizioni. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

"cessazione degli effetti civili del matrimonio"

⁴ In alcune banche dati, l'operatore ADJ non tiene conto dell'ordine in cui vengono inseriti i termini combinati; di conseguenza, se si impiega una stringa come quella riportata nell'esempio, i documenti prodotti in risposta sono sia quelli in cui compare la frase «ordine pubblico» sia quelli in cui compare la frase «pubblico ordine». È buona norma consultare il manuale della banca dati per accertare questa modalità di funzionamento dell'operatore ADJ.

⁵ Talvolta l'operatore di vicinanza è denominato WITH. Consultare il manuale della banca dati per accertare la notazione in uso.

⁶ È buona norma controllare il manuale della banca dati per accertare il numero di parole che determina la distanza massima dei termini combinati in NEAR.

produce in risposta a tutti, e solo, documenti che contengono la frase «cessazione degli effetti civili del matrimonio». La ricerca per frase esatta è molto utile per evitare il fenomeno del rumore, tuttavia produce qualche rischio di silenzio, giacché il sistema produce in risposta solo i documenti in cui la frase ricorre nella sua esatta formulazione. Un'interrogazione che operi tramite la stringa di ricerca appena riportata non includerà nella risposta quei documenti in cui, in luogo della frase “cessazione degli effetti civili del matrimonio”, compare, ad esempio, la frase “cessazione dell’efficacia civile del matrimonio”, che pure, evidentemente, si riferisce allo stesso istituto oggetto dell’interesse del ricercatore.

2.3.1. la ricerca per radice e il mascheramento

Alcuni strumenti di ricerca consentono di ridurre i rischi di silenzio dati dal modo in cui i termini immessi nelle stringhe di ricerca vengono *declinati* o *coniugati*. Se per esempio si inserisce in una stringa di ricerca il termine «prescrizione» con l'intento di selezionare tutti i documenti che trattano di quest'istituto, si ottiene una ricerca silenziosa nella misura in cui il risultato non comprende quei documenti, potenzialmente rilevanti, che contengono le parole «prescritto», «prescritti», «prescrittibile» ecc. Per ovviare a questa difficoltà è possibile utilizzare lo strumento della *ricerca per radice*, che opera attraverso l'immissione dei primi caratteri che compongono il termine, seguiti dal simbolo “*”. La ricerca produrrà in risposta tutti e solo i documenti in cui compaiono termini con radice (cioè la parte iniziale della parola) identica a quella che precede l'asterisco. Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

prescri*

produce in risposta a tutti, e solo, documenti che contengono delle parole che cominciano con la sequenza di caratteri «prescri», e dunque «prescrizione», «prescritto», «prescritti», «prescrittibile», «prescrittività», ecc. È ovvio che più caratteri sono compresi nella radice, più mirata è la ricerca; un troncamento limitato alla sillaba «pre» sarebbe ben poco utile a un ricercatore che volesse reperire documenti sulla prescrizione, giacché questa radice è presente in parole come ad esempio «prelazione», «precetto», «pretore», «prevenzione», «previdenza» ecc. e quindi determina la selezione di documenti circa istituti e organi affatto eterogenei rispetto agli obiettivi euristici di partenza.

Per molti versi analoga alla ricerca per troncamento è la tecnica del “mascheramento”, che opera tramite l'immissione del simbolo “?”. Sarà così possibile ricercare un termine nelle sue diverse forme, ossia al singolare, plurale, maschile e femminile, oppure nelle varie sue possibili dizioni (almeno quando queste differiscono soltanto per un carattere). Ad esempio, un'interrogazione che operi tramite la stringa di ricerca

fide?ussione

produce in risposta a tutti, e solo, i documenti che contengono termini come “fideiussione” e “fidejussione”; un'interrogazione che operi tramite la stringa di ricerca

azion?

produce in risposta tutti, e solo, i documenti che contengono termini come “azione”, “azioni”, “azionà”, “azionò”, e così via. Anche in questo caso, si tenta di estendere il risultato di una ricerca

fino a ricomprendere tutti i documenti che contengono un certo termine, indipendentemente dal fatto che esso compaia scritto al singolare o al plurale, o scritto in una variante piuttosto che in un'altra.