



COGITO® Semantic Search

**La tecnologia linguistica nella ricerca
delle informazioni: la rivoluzione della
Semantic Intelligence**

Alla ricerca della conoscenza

La maggior parte delle informazioni che un'impresa possiede o può reperire non è strutturata, ma è in forma testuale. Ciò rappresenta il principale ostacolo per un'organica e costante fruizione delle conoscenze prodotte e acquisite ogni giorno: si percepisce il valore dei documenti, ma poi si fa fatica a trovare ciò di cui si ha bisogno.

Dalle informazioni al business. La **ricerca** come nuova opportunità di

- integrazione fra dati, conoscenze, capitale intellettuale
- valutazione e analisi strategica
- asset decisionale

La massa dei dati non strutturati, infatti, è in costante crescita esponenziale. L'eterogeneità di taglio e formato (doc, ppt di progetto, certificati, approfondimenti di settore, pagine web, e-mail, pdf di atti di convegni e seminari, articoli, notizie provenienti da network o istituti di ricerca, files xls o rtf con report, dati di bilancio...) e i diversi tipi di archiviazione rendono il problema ancor più complesso, anche perché spesso ci si limita a salvare i documenti su cui si lavora nel proprio pc senza depositarli in un luogo accessibile a tutti.

Come orientarsi in questo "universo di conoscenze"? Come capirlo rapidamente per ricavarne le informazioni giuste nel momento in cui se ne ha bisogno?

Dalla massa dei documenti alla "conoscenza". COGITO® Semantic Search

COGITO® Semantic Search è un'applicazione per la **ricerca e l'indicizzazione concettuale** delle informazioni.

Attraverso la tecnologia linguistica di Expert System per l'analisi semantica, COGITO® Semantic Search individua nella massa dei documenti non strutturati **solo le informazioni che servono**, distinguendo ciò che è utile da ciò che non lo è.

COGITO® Semantic Search



COGITO® Semantic Search oltrepassa i limiti dei comuni strumenti di ricerca, risolvendo nel contempo i problemi legati alla continua proliferazione delle conoscenze aziendali: in tempi rapidissimi e con ottimi risultati, si possono analizzare vasti ed eterogenei insiemi di testi, evitando che il recupero delle informazioni possa tradursi in un'ingestibile esplosione di dati (overload informativo) o, paradossalmente, in un'attività infruttuosa (underload informativo).

L'approccio tecnologico basato sulla comprensione semantica può migliorare l'efficacia di ricerca dei motori tradizionali.

Pur contemplando la ricerca per keyword, infatti, l'indicizzazione dei documenti attuata con COGITO® Semantic Search presuppone la **comprensione dei concetti**.

La ricerca con COGITO® Semantic Search

- analisi linguistica
- comprensione della funzione logica degli elementi delle frasi
- disambiguazione semantica dei termini
- espansione della ricerca per lemmi
- individuazione automatica dei concetti

Una vera e propria rivoluzione nel modo di fare ricerche: il recupero delle informazioni non è limitato alla determinazione della presenza/assenza delle parole-chiave, ma al significato che le parole esprimono nel contesto (accezione) e a prescindere da genere (maschile, femminile), numero (singolare, plurale) tempi e modi verbali (presente, passato, futuro... indicativo, congiuntivo, condizionale...).

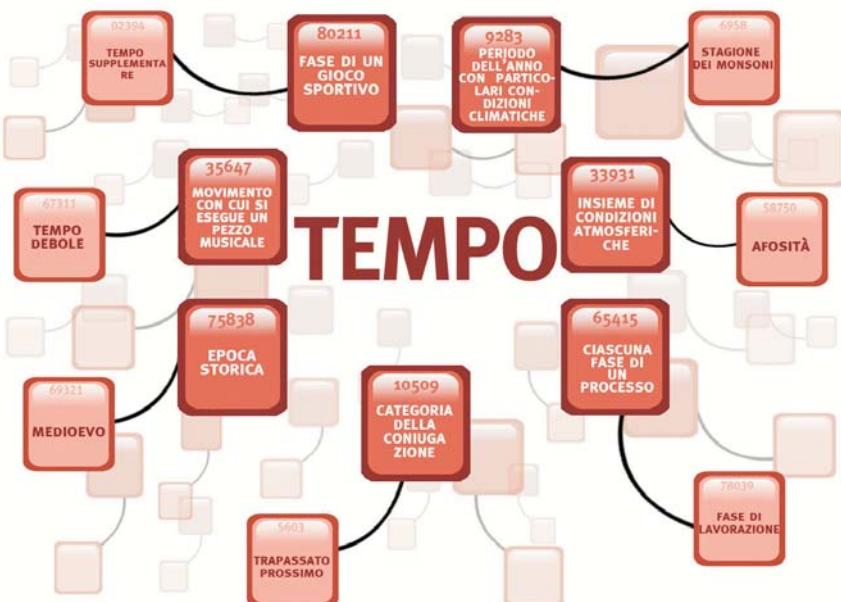
Ad esempio, qualora si esegua una ricerca per il verbo "portare", saranno selezionate anche le frasi contenenti la parola "porta" intesa come terza persona singolare del presente indicativo del verbo; e saranno invece esclusi i casi in cui "porta" è sostantivo.

È possibile "comprendere" i termini che compongono i testi, associando quelli con più accezioni a un significato piuttosto che a un altro.

Una parola può avere significati diversi, pur essendo scritta allo stesso modo; la tecnologia linguistico-semantica è l'unico mezzo per risolvere i casi di ambiguità perché si basa anche sull'**analisi** e sulla **comprendere del contesto** e determinare, così, il corretto significato delle parole.

Un esempio è il termine "tempo" che può essere inteso come epoca storica, successione di istanti, condizioni meteorologiche, parte di uno spettacolo (il primo tempo), categoria della coniugazione verbale (tempo passato remoto), velocità di esecuzione di un brano (tempo adagio), *Il Tempo* nome del quotidiano...

Oppure di "calcio" che può indicare un colpo dato col piede, il gioco di squadra, l'impugnatura di una pistola, l'elemento chimico...



- ricerca per keyword, lemmi, concetti
- indicizzazione su base semantica dei contenuti
- ricerca per "soggetto, verbo, oggetto"

Andando oltre la "forma" della keyword (parola intesa come semplice sequenza di lettere) e arrivando al "contenuto" (parola intesa come entità concettuale), si ottengono risultati ottimi in termini di **Recall** (capacità di trovare più informazioni possibili attinenti a ciò che si sta cercando), **Precision** (capacità di individuare con maggiore precisione le informazioni utili) e **Ranking** (capacità di ordinare i risultati nel modo più corretto: all'inizio i documenti più rilevanti e alla fine quelli meno interessanti perché più "lontani" da quanto cercato).

Una volta delimitato l'ambito della ricerca si può stabilire di espandere l'analisi per l'indicizzazione a tutti i contenuti in cui sono presenti eventuali sinonimi. A "macchina" intesa come veicolo a motore, per esempio, saranno associati anche auto, automezzo e autoveicolo; ma non macchina fotografica, congegno meccanico, ecc. La funzione di ricerca, inoltre, può essere estesa anche agli iponimi, cioè a termini più specifici: berlina, utilitaria, diesel, cabriolet; o, addirittura, ferrari, smart, xsara picasso...

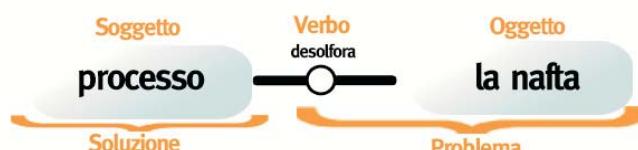
La semantic intelligence nell'individuazione di problemi e soluzioni

Trovare subito le informazioni che servono svolge un ruolo fondamentale nella creazione del valore: più è rapido l'accesso al patrimonio informativo, più snello ed efficace sarà il processo decisionale.

In particolare quando si stanno cercando risposte a domande precise, ovvero quando si cercano soluzioni per problemi ben determinati.

COGITO® Semantic Search determina la funzione logica degli elementi delle frasi e identifica ogni gruppo "soggetto-verbo-oggetto".

Questi tre elementi, fondamentali in qualsiasi frase, sono ricwendibili a uno schema che rappresenta una buona approssimazione di un **problema** (verbo+oggetto) per il quale esiste una **soluzione** (soggetto).



La procedura desolfora anche il gasolio
L'ammoniaca trasforma il b-naftolo in b-naftilammina in condizioni che lasciano il fenolo inalterato. La **procedura** descritta **desolfora** anche il **gasolio**. L'ossidazione nitrica o cromica trasforma il naftalene in un di acido

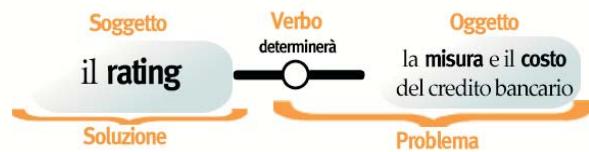
Vari criteri desolforano la nafta
Si può r2 eliminare zolfo dai combustibili derivati del petrolio. **Vari criteri**, già e0 da tempo studiati, **desolforano la nafta**. Diversamente dalla benzina

Il procedimento desolforò il gasolio
Gli idrocarburi di maggior peso molecolare vennero frammentati in presenza di un catalizzatore. Il **procedimento desolforò** il **gasolio**: per la forte capacità e0 di detonazione

Determinando soggetto-verbo-oggetto che le frasi dovrebbero avere per essere considerate degne d'attenzione, si restringe e nel contempo si perfeziona l'ambito di ricerca.

Si ottiene così un **livello di precision impensabile per un tradizionale strumento di ricerca**.

Gli esiti, ovvero le "soluzioni" individuate per i "problem" indagati, forniscono immediatamente indicazioni corrette su dove indirizzare eventuali analisi più approfondite.



Vantaggi

- Analisi linguistica approfondita e comprensione semantica
- Non solo ricerca per keyword, ma anche per lemma, concetto, soggetto-verbo-oggetto
- Riduzione di costi e tempi nel reperimento delle informazioni
- Risultati migliori in termini di recall, precision, ranking
- Facile integrazione con qualsiasi architettura software
- Implementazione su diverse piattaforme e integrabilità con altre applicazioni di Knowledge Management

COGITO®

Language Technology

COGITO® language technology ergo... Semantic Intelligence

COGITO® è l'esclusiva piattaforma semantica di Expert System per un'efficace gestione della conoscenza: dalla ricerca ed estrazione all'analisi, classificazione e trasformazione delle informazioni non strutturate. Insieme di tecnologie e risorse frutto di centinaia di anni/uomo di ricerca e sviluppo, COGITO® è la migliore tecnologia per la comprensione semantica disponibile sul mercato perché supera qualsiasi altro approccio di trattamento automatico della lingua ed è dotato di una rete semantica - chiamata Sensigrafo® - che non teme confronti:

- più di 520.000 concetti
- più di 2.800.000 connessioni (in riferimento a contesti, tipi di costruzione, argomenti e domini, locuzioni, fraseologie...)
- più di 400.000 collegamenti di iponimia e iperonimia
- più di 55.000 collegamenti di iperonimia e troponimia
- più di 370.000 collegamenti per il corpus e decine di migliaia di collegamenti tra soggetti, oggetti, meronimi...

Expert System S.p.A.

Via Virgilio 56/Q
41100 Modena - Italy

tel: +39 059 894011
fax: +39 059 894099

Via Machiavelli 47
00185 Roma - Italy

www.expertsystem.it

info@expertsystem.it

Expert System S.p.A. è leader di mercato nella realizzazione di soluzioni avanzate di **Semantic Intelligence** per aziende ed enti governativi.

Unica realtà italiana a fornire a Microsoft tecnologie avanzate integrate in tutti i suoi principali prodotti, Expert System fa della **gestione "intelligente" delle informazioni non strutturate** il proprio core business.

