



## WHITE PAPER

---

### **Qualità e valore nella gestione dell'informazione non strutturata: gli strumenti basati sull'analisi semantica**

Sponsored by: **Expert System**

---

*Analista: Fabio Rizzotto*

*Settembre 2006*





# WHITE PAPER

---

## **Qualità e valore nella gestione dell'informazione non strutturata: gli strumenti basati sull'analisi semantica**

Sponsored by: **Expert System**

---

*Analista: Fabio Rizzotto*

*Settembre 2006*



## WHITE PAPER

---

# Qualità e valore nella gestione dell'informazione non strutturata: gli strumenti basati sull'analisi semantica

---

## IDC OPINION

Gli strumenti di Business Analytics rivolti all'analisi delle basi dati non riescono da soli a supportare la valorizzazione dell'intero patrimonio di conoscenza aziendale che, stando alla realtà attuale, è fortemente sbilanciato sulla componente destrutturata.

Si stima, infatti, che circa l'80% delle informazioni aziendali sia di tipo non strutturato, presente in differenti formati: documenti elettronici, fogli excel, presentazioni, pdf, contengono testi e dati di non facile reperibilità che custodiscono tuttavia un valore inestimabile per la comprensione delle dinamiche interne ed esterne alle imprese.

Il caos dell'informazione rappresenta la fonte primaria di inefficienze e sprechi, soprattutto se misurato in termini di tempo speso per attività poco produttive. Un nuovo corso per le politiche di gestione dei contenuti deve partire necessariamente dall'adozione di strumenti di Content Access che favoriscano l'accesso, l'analisi, l'estrazione e la visualizzazione delle informazioni realmente utili per il Knowledge Worker.

Nell'ambito di questi strumenti, l'innovazione tecnologica ha consentito il passaggio da strumenti di ricerca tradizionali a sistemi in grado non solo di "cercare" e di "trovare" l'informazione presentandola in un formato normalizzato, ma anche di andare oltre fino a cogliere il "significato" dei contenuti, grazie a metodologie di ricerca basate sull'analisi linguistica approfondita (i cosiddetti motori semantici).

Il White Paper IDC contestualizza le dinamiche suddette e descrive il valore di business delle soluzioni IT basate su questi tool di analisi linguistica, tra cui è possibile posizionare la piattaforma COGITO di Expert System, concepita per fornire "intelligence" ai processi di ricerca, estrazione, categorizzazione dell'informazione non strutturata per le esigenze di gestione interna piuttosto che di monitoraggio e analisi delle fonti esterne quali Internet.

## **L'informazione non strutturata: dal "caos totale" al "caos ragionato"?**

In un'epoca caratterizzata dal cambiamento, dalla globalizzazione e dall'intensificarsi della competizione in qualsiasi settore, la conoscenza è senza dubbio l'asset più importante che un'azienda ha a disposizione per acquisire un vantaggio competitivo.

A partire da metà degli anni novanta, in contemporanea con lo sviluppo di Internet e di tecnologie informatiche esponenzialmente più potenti, rilevanti investimenti sono stati fatti per cercare di trasformare in valore aggiunto il contenuto di centinaia di archivi sparsi all'interno delle organizzazioni o per integrare soluzioni diverse in modo da consentire un utilizzo più efficiente di quello che costituisce la base di conoscenza aziendale.

Ciononostante, risultati consistenti sono stati ottenuti solo quando sia stato possibile organizzare l'informazione in "dati", ovvero in database con oggetti (tabelle, colonne etc.) legati da un codice di relazione. Per il resto delle informazioni, disponibili solo in forma non strutturata e stimate essere l'80% del totale delle informazioni rilevanti presenti in azienda, **gli sforzi compiuti hanno prodotto risultati solo parziali** resi ancora più limitati dagli effetti generati dall'evoluzione delle tecnologie, dei formati, delle fonti che hanno portato a criticità che possono essere ricondotte a due principali tipologie:

- ☒ problematiche di ordine **quantitativo**, legate alla mole crescente di documenti e contenuti che ne rende complessa la gestione sia dal punto di vista IT (maggiore capacità hardware, perdite di efficienza e rendimento dei sistemi) sia a livello di funzioni utenti, dove i volumi fisici ed elettronici rischiano di diventare ingestibili;
- ☒ problematiche di natura **qualitativa**, connesse alla difficoltà di recuperare e gestire le informazioni in maniera automatizzata. Questo fa sorgere inefficienze e costi non facilmente misurabili ma a forte impatto sulla produttività.

Queste informazioni **non strutturate** presenti sotto forma, ad esempio, **di testo, fogli excel, presentazioni, email, pdf** rappresentano non soltanto la quota maggiore dell'informazione aziendale, ma sono spesso **depositarie delle chiavi interpretative** che insieme alla componente "dati" muovono il business dell'azienda. Inoltre, i documenti testuali spesso contengono veri e propri dati che seppelliti all'interno dell'informazione "non strutturata", perdono la propria valenza e significatività perchè non facilmente accessibili o connessi a un processo specifico.

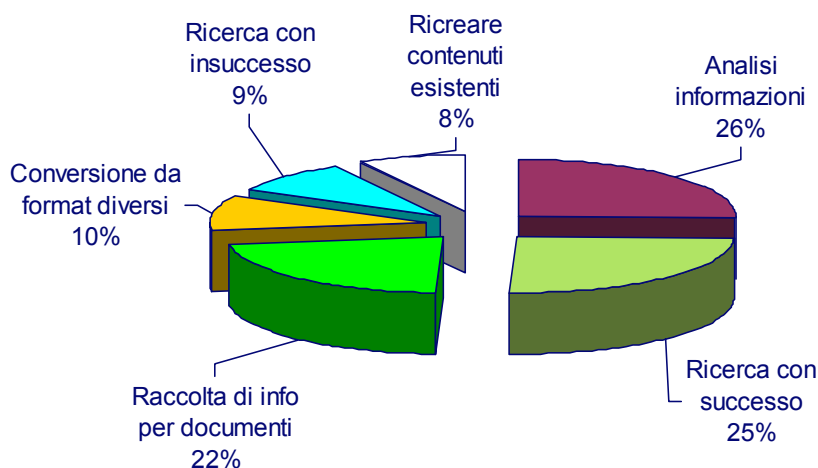
La contemporanea presenza quindi di un'organizzazione della conoscenza non ottimale, con silos di informazioni relative a progetti dipartimentali, content repository diversi etc. e di una crescente quantità di informazioni rilevanti e strategiche disponibili in formato non strutturato rendono sempre più complesso il compito degli utenti, che necessitano di adattare i contenuti ai workflow per costruire processi più strutturati. Inoltre, se si pensa alla necessità di integrare l'informazione interna con quella esterna (ad esempio su siti web o portali esterni) gli strumenti tradizionali non appaiono sempre adatti a fornire una "intelligence" reale, che individui l'informazione ricercata, estraiga solo quella rilevante e che lo faccia in maniera efficiente.

Come mostra la Figura 1, la situazione sopra descritta, con riguardo alla gestione dei contenuti, porta gli utenti a investire una certa quantità di tempo in attività poco efficienti. Posto pari a 100 il tempo speso dai knowledge worker per attività a forte connotazione "informativa", le stime presentate rappresentano una ripartizione

percentuale di questo tempo tra diverse procedure: si può osservare come a fronte di circa il 50% dedicato a ricerca con successo e analisi delle informazioni, il peso delle ricerche con insuccesso occupa il 9%, che a sua volta può essere la causa di una seconda distorsione, ovvero la riscrittura ex novo di materiale già esistente (8%), fino ad arrivare alla necessità di convertire in formati nuovi contenuti preesistenti in formati diversi (10%).

**FIGURA 1**

Attività "Information-related": come impiegano il tempo i Knowledge Worker



Fonte: IDC, 2006

### **Gli strumenti per accedere in maniera efficiente all'informazione non strutturata**

L'inadeguatezza dei meccanismi organizzativi rappresenta una delle cause delle problematiche sopra evidenziate. Metodologie consolidate, processi non integrati e scarsa pianificazione rendono spesso difficile l'accesso alle informazioni rilevanti da parte degli utenti. Tuttavia, nonostante gli sforzi sul piano organizzativo, appare ormai evidente la **difficoltà di superare** queste inefficienze **senza l'ausilio di strumenti tecnologici** adeguati.

Negli anni scorsi le aziende hanno investito in maniera prevalente nella gestione delle informazioni di tipo strutturato, migliorando l'infrastruttura (database) e adottando applicativi di **Business Intelligence** per supportare i processi di analisi dei dati.

Questi strumenti (di **Business Intelligence** o, più in generale, di **Business Analytics**) hanno dimostrato un grande valore nel supportare i processi decisionali. Reporting, analisi dei dati storici, simulazioni previsionali sono solo alcune delle tecniche che forniscono alle imprese indicazioni per la comprensione dei risultati e la messa a punto delle strategie evolutive.

Tuttavia, il **peso crescente** dei contenuti non strutturati e la difficoltà di **accedere** all'informazione rilevante hanno evidenziato la necessità di guardare a un **concetto più ampio** di Business Intelligence, non limitato alla sola componente strutturata ma in **grado di far emergere** anche il valore della componente non strutturata dell'informazione.

La "data analysis" può rivelarsi infatti **incompleta** nello spiegare le dinamiche di business e nel fornire le chiavi interpretative, quel "perché" che spesso risiede all'interno di contenuti non strutturati e che non è estraibile direttamente dalla fonte numerica. A loro volta, gli stessi dati, se riportati all'interno ad esempio di un testo, sfuggono agli strumenti di "data analysis" e rendono ancora più evidente l'esigenza di tool per l'analisi dell'informazione non strutturata (cosiddetti strumenti di "Content Access", Figura 2).

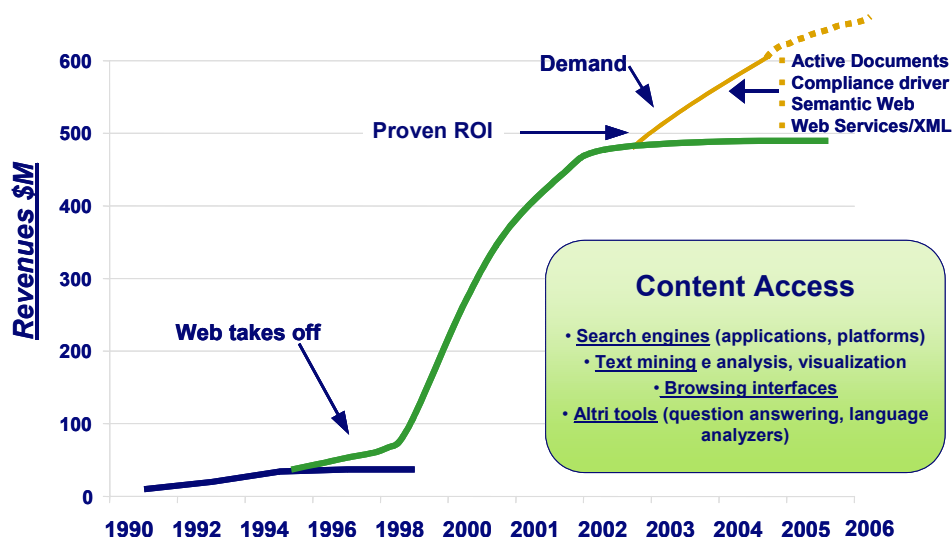
Una gestione dei contenuti allargata a questa prospettiva può consentire di "normalizzare" l'informazione non strutturata, di automatizzare processi e metodologie favorendo la crescita della produttività individuale.

L'accesso tramite strumenti di ricerca, la visualizzazione, la categorizzazione automatica del contenuto non strutturato o l'estrazione mirata di questi contenuti (**text mining**) non rappresentano temi nuovi nel panorama informatico ma hanno subito un'impennata negli ultimi anni grazie alla crescente sofisticazione delle tecnologie (Figura 2). Al boom determinato dall'esplosione della rete (che nella seconda metà degli anni novanta ha creato il business dei motori di ricerca) ha fatto seguito una fase di breve stabilizzazione prima della seconda ondata di sviluppo che stiamo vivendo in questi anni. Oggi la domanda per questo tipo di soluzioni è sostenuta da:

- ☒ **driver interni:** come già evidenziato, cresce la consapevolezza di porre rimedio al "caos" interno e di migliorare l'accesso all'informazione **esterna** (ad esempio quella residente su Internet);
- ☒ **driver esterni:** la domanda cresce anche grazie all'introduzione di tecnologie di Information Management più avanzate e all'esigenza di rendere i processi conformi a nuovi requisiti normativi (compliance).

**FIGURA 2**

Il "boom" degli strumenti di Content Access



Fonte: IDC, 2006



## "Text mining": l'analisi semantica come strumento efficace di intelligence

L'informazione **non strutturata** trasferisce **valore ai processi di business** se le tecniche utilizzate per renderla disponibile risultano efficaci. Nel corso del tempo si è assistito all'introduzione di diverse soluzioni a supporto della ricerca e dell'accesso all'informazione. È possibile individuare tre fasi principali in questo percorso:

- ☒ una **prima generazione** di strumenti, nata prima dell'avvento di Internet, era basata essenzialmente su tecniche di ricerca full text tramite parole chiave (keywords). Le banche dati e raccolte documentali diffuse fin dagli anni ottanta fornivano queste funzionalità;
- ☒ lo sviluppo tecnologico ha successivamente portato a un affinamento delle tecniche tradizionali nella direzione di nuove funzionalità: si sono affermati quindi i **motori di ricerca**, strumenti più avanzati in grado di combinare la ricerca con parole chiave all'uso di algoritmi statistici per un accesso più mirato all'informazione;
- ☒ la **terza generazione** di strumenti va oltre gli approcci sopra descritti. Negli ultimi anni si è fatta strada una nuova classe di prodotti e strumenti basati sull'uso della **linguistica** come strumento di ricerca e analisi. All'interno di quest'ultima categoria assumono particolare rilevanza gli strumenti basati **sull'analisi linguistica approfondita (motori semantici)**. La tecnologia semantica aggiunge all'analisi linguistica di base (grammaticale e morfologica) l'analisi logica e del periodo, e la disambiguazione dei termini polisemici, ovvero l'identificazione del significato corretto di tutte le parole che in una lingua possono assumere significati diversi. In questo modo, si ottiene una comprensione univoca e superiore del significato del contenuto di un documento avvicinando lo strumento informatico ai processi utilizzati dalle persone.

Le **tecnologie linguistiche, in particolare quelle basate sulla semantica**, si contraddistinguono per la capacità di superare i **limiti** che gli strumenti di **ricerca tradizionali** hanno mostrato nel corso del tempo in termini di:

- ☒ scarsa precisione, ovvero incapacità di identificare SOLO i documenti rilevanti in funzione dei criteri di selezione (si pensi all'effetto dei diversi significati associati ad una parola);
- ☒ limitato "recall", ovvero incapacità di estrarre TUTTI i documenti rilevanti per la ricerca in essere. Con i sistemi tradizionali, per esempio, è impossibile estrarre tutti i documenti che trattano il concetto su cui si vuole effettuare la ricerca a meno di non inserire tutti i termini legati al concetto stesso ad uno ad uno come parole chiave.

Queste problematiche rendono **difficile "trovare"** l'informazione ricercata. Inoltre, laddove disponibile, il trattamento è complesso in quanto la comprensione del significato è nulla o limitata e i risultati **non sono presentati in formato strutturato**. Gli strumenti basati sull'analisi linguistica si prefiggono di ridurre queste distorsioni; permettono di effettuare il text mining di contenuti (fornendo all'utente un "codice di interrogazione" più vicino ai processi mentali) e di presentare i risultati in maniera strutturata. Gli strumenti basati sulla semantica vanno oltre poiché consentono un migliore trattamento dell'informazione, sia nel caso di categorizzazione automatica (ovvero l'utilizzo di tassonomie e criteri di classificazione dei contenuti) sia del text

mining; sono in grado di comprendere in modo preciso il significato di ogni parola basandosi sul contesto in cui ogni termine è utilizzato.

La maggiore capacità di comprensione del contenuto consente anche l'utilizzo della tecnologia linguistica semantica per tool di "Question Answering". Questi permettono di interrogare il sistema, anche utilizzando il linguaggio naturale, per ottenere un risultato che non è più dettato dalla frequenza o dalla "casualità" (cerca tutte le informazioni che contengono la parola o frase digitata) ma punta a dare una risposta a una query specifica dell'utente.

---

## **Gli strumenti di analisi linguistica e i benefici di Knowledge Management**

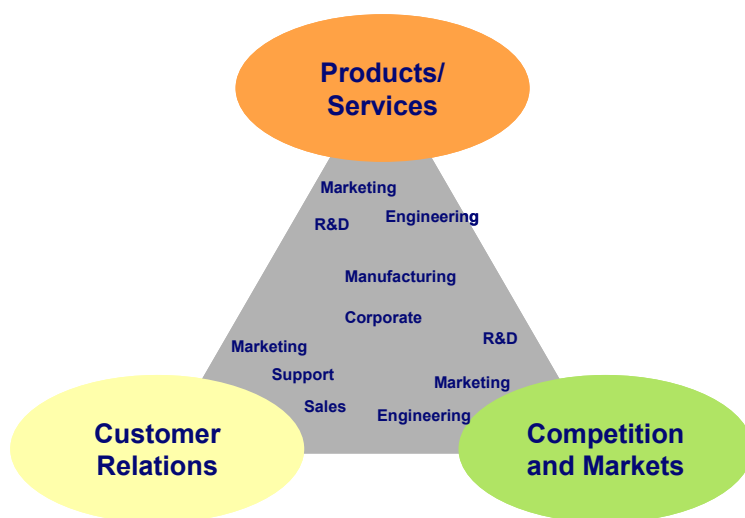
Gli strumenti di terza generazione - basati su tool di analisi linguistica approfondita o semantica - sono in grado di **aggiungere valore** all'azienda andando ad agire sui tre elementi fondamentali che costituiscono l'azienda:

- ☒ **Tecnologie:** grazie alla capacità di comprendere, evidenziare ed eventualmente estrarre i contenuti non strutturati più rilevanti, questi strumenti rendono possibile la valorizzazione dell'informazione non strutturata. In questo modo aumentano l'efficacia e la produttività degli investimenti in strumenti più tradizionali quali per esempio le applicazioni di Business Intelligence o gli strumenti per il Customer Relationship Management;
- ☒ **Persone:** l'utilizzo di strumenti di analisi linguistica semantica consentendo, per esempio, una migliore precisione e recall nell'attività di ricerca e una più precisa categorizzazione automatica, sia delle informazioni interne che esterne all'azienda, aumenta la qualità del lavoro e la produttività individuale del knowledge worker;
- ☒ **Processi:** la qualità dei risultati dell'analisi delle informazioni non strutturate, derivanti dall'utilizzo degli strumenti di analisi linguistica approfondita, consente una maggiore automazione nei processi di gestione dell'informazione.

**Vantaggi** ulteriori possono essere ritrovati nella possibilità di ampliare il numero degli utenti coinvolti in processi e progetti, di incentivare la collaborazione e la condivisione della conoscenza, di tenere maggiormente sotto controllo il processo di market intelligence. La Figura 3 rappresenta graficamente i contesti organizzativi nei quali possono essere ottenuti ritorni significativi da iniziative di questo genere.

**FIGURA 3**

La portata "enterprisewide" degli strumenti di Text Mining



Fonte: IDC, 2006

La valorizzazione della conoscenza come asset strategico ottenuta grazie agli strumenti di analisi linguistica può interessare diversi contesti:

- ☒ **Customer relation:** i tool di text mining sono in grado di portare **benefici** alla gestione della relazione con il cliente. Le aziende possono, infatti, adottare tecniche di analisi dell'informazione non strutturata per migliorare l'efficacia delle politiche e degli strumenti di Customer Relationship Management (**CRM**). Si pensi, ad esempio, agli appunti e alle note sotto forma di testo frutto dell'interazione degli operatori di call center con i clienti: un sistema di text mining può aiutare a individuare eventuali criticità o elementi che spesso sfuggono ai sistemi operazionali. In altri casi questi strumenti possono fornire all'operatore soluzioni in "real-time" da proporre al cliente in funzione della particolare tipologia di richiesta. Al di fuori dei confini aziendali, l'analisi semantica di siti web, blog o altre fonti può fornire indicazioni su specifici clienti migliorando i processi di pre-sale e post-sale.
- ☒ **Eccellenza di prodotto/servizio:** i processi che portano alla realizzazione di un prodotto o alla delivery di un servizio possono avvalersi di strumenti di analisi linguistica. L'accesso efficiente a contenuti e informazioni **non strutturati** può supportare ad esempio le fasi di **studio** e di **progettazione** di un prodotto, consentendo ai responsabili (di progetto o di produzione) di valorizzare l'esperienza acquisita e di stare al passo con i cambiamenti del mercato. In generale, in questa sfera è possibile far rientrare tutte le attività di "Search&Retrieval" avanzato, quali il monitoraggio delle fonti informative interne, la ricerca e individuazione di correlazioni fra documenti storici etc.
- ☒ **Market Intelligence:** il text mining può supportare le organizzazioni nell'interpretazione degli **scenari** competitivi e nella messa a punto delle strategie di business. Gli strumenti di **Semantic Web**, ad esempio, consentono l'analisi della cosiddetta "Internet-invisibile" favorendo la "risalita" di informazioni utili che possono riguardare concorrenti, mercati, informazioni finanziarie etc. L'analisi intelligente delle fonti esterne non consente soltanto di intercettare le dinamiche tecnologiche e settoriali; più in generale, il monitoraggio permette alle imprese di proteggere maggiormente le informazioni sensibili tutelando la sfera della **proprietà intellettuale**.

## **COGITO: la language technology di Expert System per la "Semantic Intelligence"**

Tra le soluzioni di mercato, la piattaforma tecnologica **COGITO di Expert System** si basa sulla semantica per ricercare, categorizzare, analizzare ed estrarre testi e informazioni non strutturate. **COGITO** è il risultato dell'esperienza maturata da Expert System nell'applicazione della linguistica alle tecnologie informatiche. La mappa concettuale, tipica di questa tipologia di strumenti poggia, nel caso di **COGITO**, su una rete semantica denominata "Sensigrafo". I moduli della soluzione **COGITO** coprono aree molto differenti del mondo dell'informazione, spaziando trasversalmente tra contenuti interni alle organizzazioni e contenuti esterni come il web.

**COGITO di Expert System** è una tecnologia che si adatta a diversi contesti aziendali e si prefigge di portare i vantaggi della "Semantic Intelligence" nelle principali aree evidenziate in Figura 3:

- ☒ i moduli di "**Semantic Search**", "**Discover**" e "**Categorizer**" rappresentano il nucleo centrale della piattaforma **COGITO** e forniscono una serie di funzionalità di "Content Management e Access" quali indicizzazione e ricerca, estrazione e normalizzazione, categorizzazione e distribuzione;
- ☒ il modulo "**Contact**" è il modulo dedicato alla gestione del customer care sia via web sia via SMS sfruttando al massimo la base di conoscenza aziendale;
- ☒ il modulo "**Intelligence**", invece, è la componente dedicata all'analisi delle fonti aperte esterne, anche in real time, al fine di supportare sia attività di difesa militare, anti-terrorismo e pubblica sicurezza sia le iniziative strategiche aziendali in tema di analisi e monitoraggio dello scenario competitivo e di mercato.

Le caratteristiche della tecnologia **COGITO di Expert System** appaiono allineate alle esigenze di un mercato in rapida trasformazione, che necessita di mettere ordine e trarre beneficio dalla gestione delle informazioni non strutturate che popolano sia i sistemi IT interni sia l'ecosistema di relazioni esterne all'organizzazione.

Se si osservano le tre aree più critiche per la gestione dell'informazione non strutturata, l'applicazione della Semantic Intelligence genera vantaggi evidenti anche a confronto degli strumenti di analisi linguistica di base. I benefici possono essere riassunti nei punti seguenti:

- ☒ la **ricerca semantica** permette di contenere gli impatti negativi dell'esplosione dell'informazione, consentendo all'utente di bypassare il problema della quantità per confrontarsi direttamente con le **questioni più qualitative della materia**;
- ☒ la categorizzazione basata sulla semantica - grazie alla maggiore comprensione del contenuto e alla disambiguazione dei termini - permette di raggiungere una qualità notevole, paragonabile a quella ottenuta manualmente, anche nel caso di classificazioni basate su migliaia di categorie e sottocategorie;
- ☒ anche l'attività di estrazione automatica beneficia di questa maggiore comprensione, consentendo l'identificazione non solo di entità specifiche, ma anche delle relazioni a prima vista non evidenti tra concetti e informazioni presenti in parti diverse di uno stesso documento o anche in documenti diversi.

## Case Study: EniTecnologie

EniTecnologie, in quanto Corporate Technology Company del Gruppo ENI, rappresenta uno dei maggiori centri di eccellenza per la ricerca industriale. EniTecnologie svolge attività di innovazione tecnologica per il mantenimento della competitività di business nel breve, medio e lungo periodo. A tale fine opera in coerenza con le strategie Eni su tutti gli aspetti della catena dell'innovazione, dal monitoraggio tecnologico agli studi di scenario fino alla ricerca applicata, dallo sviluppo e valutazione di tecnologie fino al trasferimento tecnologico dell'innovazione al business.

Il presidio della tecnologia e la generazione di valore attraverso attività di innovazione sono parte integrante della vision Eni.

Per EniTecnologie è di fondamentale importanza il monitoraggio dei competitor e la percezione dei segnali deboli di cambiamento colti sui mercati finali di riferimento. Altrettanto importante è la gestione efficace ed efficiente di dati e prodotti della conoscenza generata all'interno.

Il principio della "**conoscenza come asset**" rappresenta pertanto uno dei pilastri della strategia di gestione della conoscenza di EniTecnologie che da molti anni ha attribuito particolare attenzione alle metodologie e alle tecniche di analisi dei contenuti e all'evoluzione delle piattaforme tecnologiche di supporto.

Intorno alla metà degli anni novanta, l'avvento di Internet ha rappresentato un momento di discontinuità nella politica di Content Management di EniTecnologie, che si è arricchita della componente web in maniera sempre più determinante. Il secondo momento decisivo, più recente, è costituito dall'avvento degli strumenti di analisi linguistica, che hanno rivoluzionato gli approcci alla comprensione e all'estrazione delle informazioni.

Componente importante delle "knowledge activities" di EniTecnologie è l'accesso e l'analisi di documenti di contenuto tecnologico e brevettuale; da questo punto di vista l'attività può essere definita "**document-intensive**".

**EniTecnologie ha individuato in COGITO di Expert System** una soluzione particolarmente incisiva per venire incontro ad esigenze di ricerca testuale di tipo "non convenzionale". L'applicazione di un modello di ricerca basato sulla semantica ha permesso di avvicinare sensibilmente lo strumento agli utenti, grazie alla possibilità di elaborare e presentare i risultati secondo il principio delle mappe concettuali.

COGITO ha permesso di **avvicinare** lo strumento alla "mente" dell'utente per portare alla luce i nessi e le relazioni esistenti nell'informazione non strutturata. Ad esempio, il sistema consente di interrogare i documenti di brevetto seguendo un approccio simile al processo mentale dell'utilizzatore, che tipicamente cerca di individuare all'interno del documento "soluzioni" a "problemi". Questa logica collima, a sua volta, con le caratteristiche del documento brevettuale che, per propria natura, tratta l'innovazione con il criterio "problema verso soluzione".

Accanto alle caratteristiche funzionali di semantica e text mining, a favore della selezione della piattaforma COGITO hanno giocato anche la flessibilità e l'adattabilità alle esigenze di EniTecnologie, a differenza di prodotti concorrenti più chiusi e meno parametrizzabili.

Le **forti potenzialità** dello strumento COGITO hanno portato EniTecnologie a **espandere rapidamente il progetto**: da una prima fase sperimentale concentrata su pochi utenti, l'iniziativa si è diffusa grazie a meccanismi collaborativi che hanno alimentato la costruzione di una significativa "**user experience**".

I benefici di processo percepiti hanno portato EniTecnologie al successivo sviluppo di un progetto per un Portale Intranet che svolgesse, tra l'altro, le funzioni di collaboration, comunicazione interna, archiviazione e ricerca della documentazione per metadati (parole chiave, autore, etc.) e, in particolare, archiviazione e ricerca basate su classificazioni della conoscenza (tassonomie) e riconoscimento di concetti, significati e strutture linguistiche complesse.

La collaborazione tra il team Expert System e il gruppo di lavoro di EniTecnologie è stata proficua lungo tutto il processo che ha portato non solo all'implementazione di COGITO ma soprattutto al go-live e al mantenimento, caratterizzati da un forte commitment interno teso al superamento dei problemi e all'instaurazione di una condizione di apprendimento permanente.

I **benefici** dell'adozione di **COGITO** come piattaforma di Knowledge Management spaziano dall'efficacia della **ricerca semantica** - capace di fornire **risposte mirate** e di estrarre "**significato**" dai documenti testuali - a una più completa **organizzazione** del patrimonio informativo, dalle garanzie di sicurezza e compliance allo sharing di conoscenza e collaborazione tra gli utenti, ma soprattutto al supporto tangibile ai **processi di business** che si sviluppano lungo l'asse della mission strategica.

Oggi il patrimonio di knowledge interno viene gestito dai moduli **COGITO Categorizer** e **COGITO Semantic Search**, già integrati nel Portale Intranet di EniTecnologie.

Contestualmente è stata avviata l'integrazione della componente di analisi delle fonti esterne. La mappa della conoscenza creata grazie alla piattaforma COGITO è stata completata con il modulo "**Intelligence**" che opera nel monitoraggio delle **informazioni esterne**. I **knowledge workers di EniTecnologie si avvalgono di "COGITO Intelligence" per cogliere elementi altrimenti non facilmente estraibili senza l'uso** dell'approccio semantico, che consente una completa e rapida organizzazione del patrimonio informativo: siti esterni e portali web possono essere "scandagliati" secondo criteri standard, parametrizzati o definiti mediante interrogazioni specifiche, frutto dell'esperienza e dell'autonomia dell'utente, che può aggirare la problematica "dell'eccesso di informazione" per puntare a contenuti di valore per l'analisi e l'elaborazione di tutti gli elementi che possono impattare sull'analisi dei trend tecnologici anche di lungo periodo e sull'individuazione delle strategie tecnologiche e di innovazione dei competitor.

## Considerazioni finali: capitalizzare la conoscenza

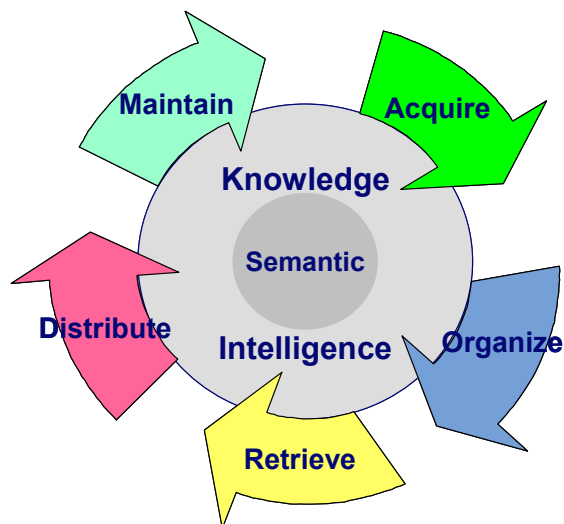
Una gestione efficace della conoscenza aziendale, che non può prescindere da una gestione ottimale delle informazioni non strutturate, consente alle aziende di generare valore sfruttando al massimo le proprie risorse. Si viene infatti a creare una sorta di "ambiente di apprendimento" che alimenta un circolo virtuoso che facilita ed amplifica questa creazione di valore: una nuova conoscenza, a sua volta, contribuirà ad alimentare la conoscenza già presente e distribuita.

Semplificando, le componenti operative di una gestione delle informazioni non strutturate secondo il principio della "capitalizzazione della conoscenza" possono essere così sintetizzate (Figura 4):

- ☒ acquisizione;
- ☒ organizzazione e archiviazione;
- ☒ accesso e recupero;
- ☒ diffusione e distribuzione;
- ☒ mantenimento, in un'ottica di valore, rilevanza e aggiornamento.

### FIGURA 4

Un nucleo semantico alimenta il circolo virtuoso della conoscenza



Fonte: IDC, 2006

Soluzioni basate su tool di analisi semantica, ottimizzando la gestione dell'informazione non strutturata, giocano un ruolo importante nel favorire l'attivazione di un processo di apprendimento, che ricavi valore da contenuti disseminati nei vari strati delle architetture IT. Questa importanza è destinata a crescere in futuro proporzionalmente alla quantità di informazioni non strutturate presenti in azienda. Da non trascurare infine è il processo di **personalizzazione** che consenta di adattare lo strumento alle effettive esigenze e caratteristiche dell'azienda stessa.

---

## **Copyright Notice**

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.

Copyright 2006 IDC. Reproduction without written permission is completely forbidden.





IDC Italia - Viale Monza, 14 - 20127 Milano ITALY  
Tel: +39 02 284571 - Fax: +39 02 28457.333

[www.idc.com/italy](http://www.idc.com/italy)