



ATON - ENTERPRISE INFORMATION MINING SOLUTIONS

Da tempo assistiamo alla necessità crescente per le aziende di poter reperire informazioni "business critical" all'interno delle grosse moli di dati aziendali.

Se consideriamo poi i formati eterogenei (documenti pdf, excel, presentazioni, etc), le diverse dislocazioni (file-system, server di posta, repository documentali) e le fonti alternative (siti WEB, blog, forum), è evidente che non si può prescindere da strumenti specifici per rendere efficace la ricerca delle informazioni.

La capacità di ricercare informazioni ottenendole nel momento in cui sono utili, spesso a partire da grosse moli di dati distribuiti, rappresenta la possibilità per una organizzazione di rispondere rapidamente alle esigenze di cambiamento e di evoluzione dei servizi che offre.

La capacità poi di acquisire informazioni su fonti aperte agevola le attività di *Marketing Intelligence* e *Sentiment Analysis*, sempre più critiche in mercati altamente competitivi.

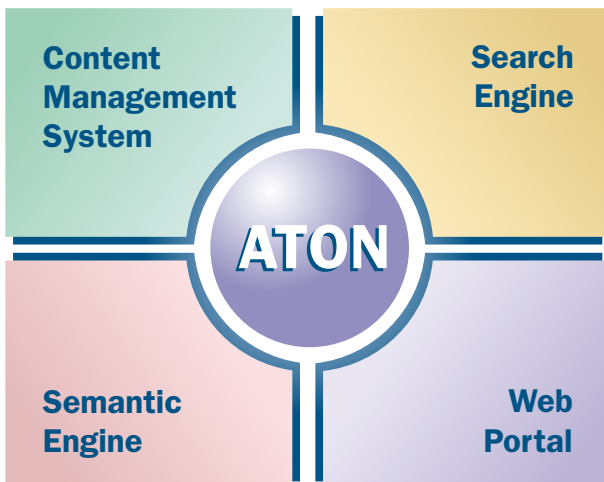
For some time now companies have been expressing an increasingly pressing need to be able to identify "business critical" information in huge amounts of corporate data. When we remember the multitude of possible file formats (pdf, spreadsheets, presentations, etc.), locations (file systems, mail servers, document repositories) and alternative sources (websites, blogs, forums), it is clear that specially designed tools are needed to improve the effectiveness of information searches.

The ability to search for information and find it when it's needed, often buried in huge amounts of distributed data, represents an opportunity for an organisation to respond rapidly to changes and developments in the services it offers.

The ability to acquire information from open sources facilitates the work of Marketing Intelligence and Sentiment Analysis, which are increasingly critical in today's highly competitive markets.

LA SOLUZIONE ATON

Gli strumenti di ricerca evoluti sono generalmente capaci di integrare caratteristiche utili ad una gestione *unificata delle informazioni*. Tali strumenti integrano le tecniche di ricerca tipiche della business intelligence tradizionale con la flessibilità di tool di search e di content management. Adottano, infatti, algoritmi e tecniche di ricerca flessibili, che meglio si adattano alla consultazione di informazioni distribuite e non strutturate, presenti all'interno ed al di fuori del contesto aziendale. Questo approccio molto spesso migliora il processo di ricerca grazie alla comprensione del contesto semantico in cui queste soluzioni operano. ATON è la soluzione di *Enterprise Information Mining* realizzata da Elsag Datamat per la gestione intelligente di tutta l'informazione di interesse aziendale. Aton si pone come obiettivo di organizzare la conoscenza di business attraverso informazioni presenti su applicazioni analitiche, completandola ed integrandola con quelle provenienti da *email, contact center, fogli excel, e blog*, utili a completare l'universo della conoscenza, organizzandola secondo categorie personalizzate di ricerca. La comprensione dei concetti rappresenta quindi un elemento chiave che migliora il processo di ricerca.



ATON è in grado non solo di “cercare” ma di “trovare” l'informazione, comprendendo il “significato” dei contenuti, grazie a metodologie di ricerca basate sull'analisi linguistica approfondita, i cosiddetti *motori semantici*. Attraverso l'analisi linguistica si effettua il text mining dei contenuti e vengono presentati i risultati in maniera strutturata. ATON va oltre poiché consente un migliore trattamento dell'informazione, sia nel caso di categorizzazione automatica (mediante tassonomie e criteri di classificazione dei contenuti) sia nell'ambito del text mining, effettuando correlazioni complesse. La tecnologia semantica aggiunge all'analisi linguistica la disambiguazione dei termini che possono assumere significati molteplici. Si pensi, ad esempio, alla ricerca di informazioni correlate alla parola tempo, nei suoi vari significati ed accezioni: successione di istanti, tempo meteorologico, musicale, parte di uno spettacolo, testata giornalistica e via dicendo.

THE ATON SOLUTION

Advanced search tools are usually integrated with unified information management functions. These tools integrate the search techniques typically used in traditional business intelligence activities, with the flexibility of search and content management tools. They use flexible search algorithms and techniques, which are best suited to the consultation of distributed, unstructured information stored both inside and outside the corporate perimeter. This approach frequently improves the search process by understanding the semantic context in which these solutions operate. ATON is the Enterprise Information Mining solution developed by Elsag Datamat for the intelligent management of all information of interest to the company. The aim of ATON is to organise the business knowledge contained in data used by analytical applications, to integrate this with supplementary information acquired from emails, contact centres, spreadsheets and blogs, and to organise it into personalised search categories. Concept comprehension therefore represents a key factor in improving the search process.

ATON is able not only to “look for” information, but also to “find” it, by understanding what content “means” using search techniques based on in-depth linguistic analysis, also known as semantic engines. Linguistic analysis is used to perform content text mining and the results are presented in a structured form. ATON also improves information handling both in the case of automatic categorisation (based on taxonomy and content classification criteria), as well as in the framework of text mining, by performing complex correlations. Semantic technology supplements linguistic analysis by providing disambiguation of terms with multiple meanings. Just think, for example, about searching for information correlated with the word cold, which has a variety of meanings: an illness, a temperature sensation or a natural phenomenon.

THE TECHNOLOGY APPROACH

The approach adopted by ATON, which incorporates complementary search engine and semantic engine technologies, increases the number of search results and improves effectiveness. This overcomes limitations caused by the constant evolution of corporate knowledge, extracting only useful information by being able to “understand” the context and decide on the correct meaning. This is performed by a semantic network which organises knowledge about the domain and which can be constantly updated and developed. Large quantities of information can be managed using a highly scalable search index, which is essential in large enterprises. Generally speaking, traditional full-text indexing only makes a portion of the available information accessible, because it presupposes that the user knows exactly what to look for, often based on keywords, but much information may escape the search procedure.

L'APPROCCIO TECNOLOGICO

L'approccio di ATON, che ingloba tecnologie complementari di un motore di ricerca e un motore semantico, consente di incrementare i risultati della ricerca migliorandone l'efficacia. Oltrepassa i limiti legati alla continua evoluzione delle conoscenze aziendali, estraendo solo le informazioni utili, grazie anche alla "comprensione" del contesto per determinare il corretto significato. Questo avviene grazie alla implementazione di una rete semantica che organizza la conoscenza del dominio e che può essere continuamente aggiornata ed evoluta. Consente la gestione di grosse moli di dati con la possibilità di implementare un indice di ricerca largamente scalabile, fondamentale nelle grandi realtà aziendali. In genere la comune indicizzazione full-text rende accessibile solo una parte delle informazioni perché presuppone che l'utente sappia esattamente cosa cercare utilizzando spesso parole chiavi (keyword): così facendo però buona parte dell'informazione può sfuggire alla ricerca. In ATON l'indicizzazione full-text viene integrata con l'aggiunta di metadati a contenuto informativo lessicale, sottoponendo il testo ad analisi grammaticale. Ciascun elemento della frase, anche composto da più parole, viene ricondotto ad un lemma del lessico della lingua di riferimento.

Questo consente enormi vantaggi in termini di:

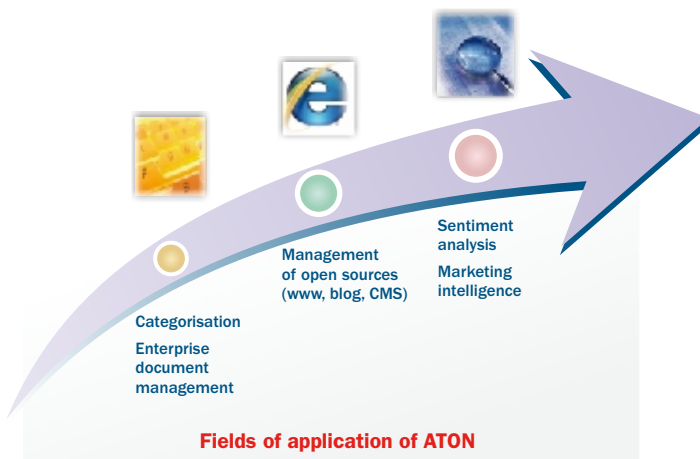
- **Recall** in quanto consente di migliorare l'attinenza dei risultati.
- **Precision** incrementando la selettività della ricerca
- **Ranking** posizionando il singolo risultato nell'elenco tenendo conto anche degli aspetti lessicali e semantici.

Chiaramente l'utente può anche limitarsi a scrivere una o più keyword, ma questo porterà in genere a criteri di ricerca meno precisi, perché la mancanza di contesto renderà più difficile il processo di disambiguazione dei termini con più significati.

IL FRONT-END DELL'APPLICAZIONE

Il front-end di ATON è stato studiato per permettere all'utente la composizione del criterio di ricerca con la massima precisione. L'utente ha la possibilità di decidere se vuole cercare keyword, lemmi o concetti anche su frasi articolate con operatori booleani di sua scelta. Dispone inoltre di una ricerca evoluta proponendo una finestra di navigazione nella rete semantica, nella quale l'utente può selezionare una o più accezioni aiutando così il processo di disambiguazione, eventualmente spostandosi su concetti diversi in relazione con il concetto iniziale. Può inoltre navigare attraverso i contenuti della ricerca filtrando in base alla categoria di interesse.

Questa interfaccia permette all'utente di definire in maniera autonoma e personalizzata il rapporto tra recall e precision. Consente inoltre all'utente la ricerca per concetti correlati. Per andare incontro poi alle esigenze dell'utente è possibile integrare un interprete di linguaggio naturale. In fase di ricerca, l'utente è invitato ad esprimere la sua richiesta con una frase o una domanda articolata in forma discorsiva.



In ATON, full-text indexing is integrated with lexical metadata by submitting the text to grammatical analysis.

Each element in the multi-word phrase is associated with a headword in the reference language.

This delivers huge advantages in terms of:

- **Recall**, as it improves the pertinence of results;
- **Precision**, by increasing the selectivity of the search;
- **Ranking**, by positioning individual results based on lexical and semantic criteria.

Clearly, users can also simply enter one or more keywords, but in general the resulting search criteria are much less precise, because lack of context makes the disambiguation of terms with multiple meanings more difficult.

APPLICATION FRONT-END

The front-end of ATON has been designed to allow users to define search criteria with the greatest precision based on keywords, headwords or concepts, even in complex phrases, and any Boolean operators.

There is also an advanced search mode, with a window in which to browse through the semantic network and select one or more meanings to help the disambiguation process and modify the original concepts where necessary.

Users can also browse through research results, filtering them based on categories of interest.

This interface enables users to fine tune the recall/precision balance themselves, as well as allowing them to search for correlated concepts. A natural language interpreter can also be incorporated if required, in which case users are invited to define their search criteria using a phrase or question expressed in natural language. This is then analysed to identify the headwords and concepts, and so provide users with structured answers.

SERVICE ORIENTED ARCHITECTURE

The service oriented architecture adopted uses an integration bus to facilitate platform component interoperability and make evolution easier by simplifying the introduction of new services.

La frase viene sottoposta ad una serie di analisi allo scopo di trovare i lemmi e i concetti espressi e fornire all'utente risposte strutturate.

L'ARCHITETTURA ORIENTATA AI SERVIZI

L'architettura di riferimento "Service Oriented" utilizza un bus di integrazione che facilita l'interoperabilità delle componenti della piattaforma e ne agevola le sue evoluzioni semplificando l'introduzione di nuovi servizi. Offre servizi che possono essere invocati da altri sistemi attraverso i protocolli standard di rete.

Questa architettura rende possibile inoltre l'integrazione con altre soluzioni di knowledge management eventualmente presenti in azienda. ATON dispone di connettori per acquisire automaticamente contenuti sia su file system, che sulle cosiddette "fonti aperte" come il WEB, le newsgroup ed i blog.

TECNOLOGIE DI BASE

La versione Enterprise di ATON integra due tecnologie complementari: un *Enterprise Search Engine* basato su tecnologia FAST ESP (*) ed un *Motore Semantico* basato su tecnologia COGITO Expert System (*) per l'analisi linguistica. ATON è interfacciabile con le migliori soluzioni disponibili sul mercato, tra cui la soluzione ArcheaEnterprise (*) di Secure Content Management System sviluppata da Elsag Datamat.

It provides services which can be called by other systems using standard network protocols.

This architecture also makes integration possible with other existing knowledge management solutions.

ATON has connectors to acquire content automatically both from file systems and so-called "open sources", such as the web, newsgroups and blogs.

BASIC TECHNOLOGY

The enterprise version of ATON integrates two complementary technologies: an Enterprise Search Engine based on FAST ESP () and a Semantic Engine based on COGITO Expert System (*) technology for linguistic analysis. ATON can be interfaced with leading solutions available on the market, including the ArcheaEnterprise (*) Secure Content Management System developed by Elsag Datamat.*

KEY FEATURES:

Content migration

Complex searches

Automatic categorisation

Content indexing (keywords, concepts, headwords)

Natural language interface

Semantic intelligence

(*) All above marks are registered trademarks and property of relevant companies.