

# Smarter Content Summit

Roma, 4 aprile 2013

---

**Maurizio Gallotti**

*Technical Sales & Solutions*

IBM Software - Industry Solutions - ECM

**Smarter Content Business Solutions:**

*Nuova Conoscenza con l'Analisi dei Contenuti*  
*Casi d'uso*

# IBM Content Analytics

## Utilizzo delle informazioni

### Gli approcci tradizionali stanno convergendo

#### More than keyword search is needed

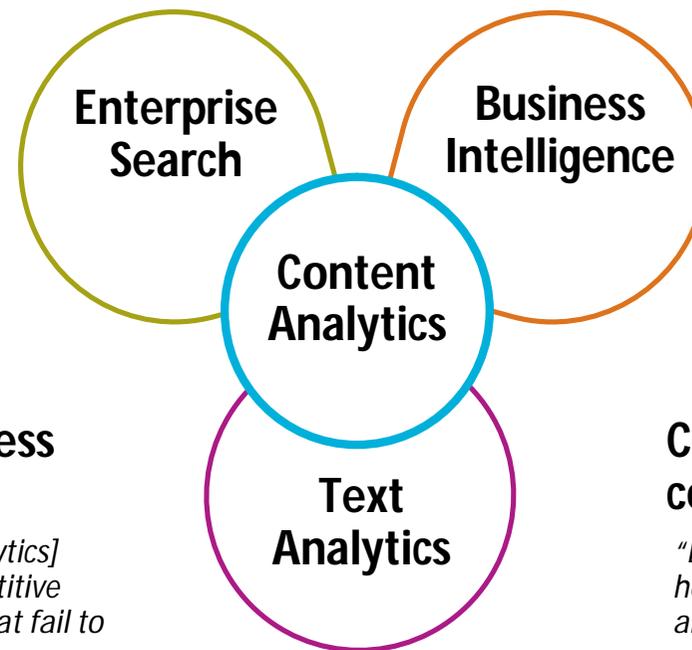
*"Making unstructured data searchable is now a presumed primary interface for applications of all kinds, as well as for intranets and content repositories."*

– Whit Andrews, Rita Knox Gartner

#### Increasing in business importance

*"Early adopters of [text analytics] are already gaining a competitive advantage. Organizations that fail to do so will be at risk."*

– Sue Feldman IDC



#### Analyzing unstructured content no longer optional

*"For many business process professionals, access to structured data, even when supported by BI or predictive analytics, lacks sufficient context for customer service, finance, and other areas where communications with customers involves many channels"*

– Craig Le Clair Forrester

#### Converging toward content analytics

*"Every enterprise should understand how content analytics can produce answers to its critical questions; understanding this now will make it possible to exploit these tools as their availability proliferates."*

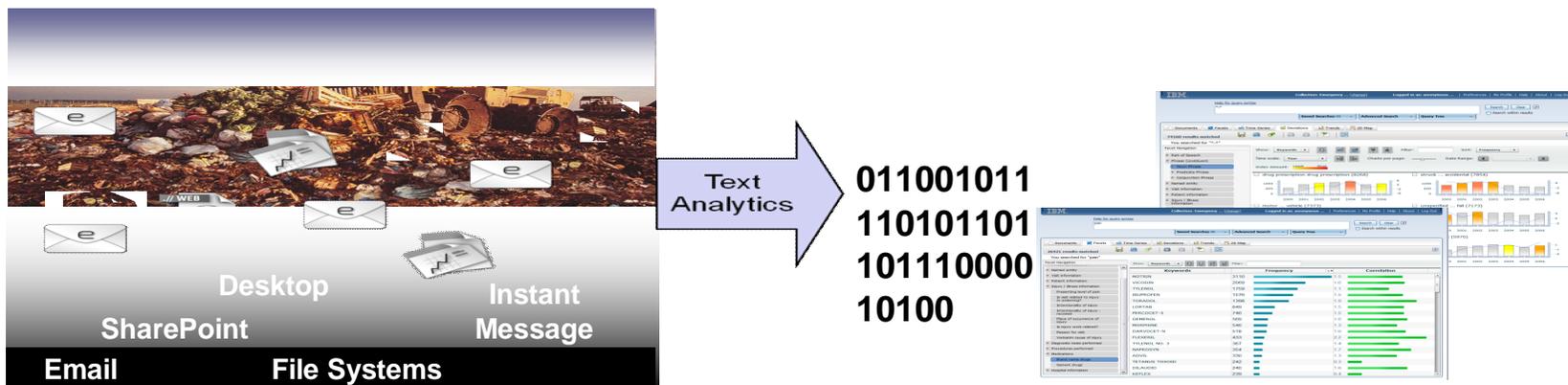
– Rita Knox Gartner



# IBM Content Analytics

*L'anello mancante tra la ricerca full-text e la Business Intelligence*

- Nella ricerca full-text i criteri sono espressi mediante termini o espressioni regolari
- La Business Intelligence analizza dati e informazioni ricavate dalle sorgenti dati strutturate
- **IBM Content Analytics** consente di:
  - estrarre, ordinare ed analizzare informazioni di valore contenute nei documenti
  - semplificare la condivisione del patrimonio informativo aziendale
  - facilitare e rendere efficace l'analisi di grandi quantità di documenti



# IBM Content Analytics

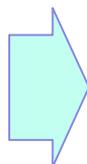
## Funzionalità di Text Analytics

I contenuti diventano dati strutturati...

Operatore CRC **PC143** ,  
contatto cliente del **15 Giugno**  
**2011 23:47**

Il cliente **Mario Rossi** lamenta  
la mancata ricezione  
dell'estratto conto per il c/c  
**123456** agenzia **ABCD**, che  
non arriva nè a lui e neppure  
al cointestatario **Paola Verdi** .  
Richiede invio a nuovo  
indirizzo

**Via Roma 22**, Secondigliano,  
**Napoli**.



Operatore_CRC	PC143
Contatto_Data_Ora	15/06/2011 : 23:47
Cliente_Nome	Mario
Cliente_Cognome	Rossi
Numero_Conto_Corrente	123456
Numero_Agenzia	ABCD
Cointestatario_Nome	Paola
Cointestatario_Cognome	Verdi
Cliente_Città	Napoli
Cliente_Quartiere	Secondigliano
Cliente_Indirizzo	Via Roma 22
Sentiment	Negativo



# IBM Content Analytics

## Funzionalità di Text mining

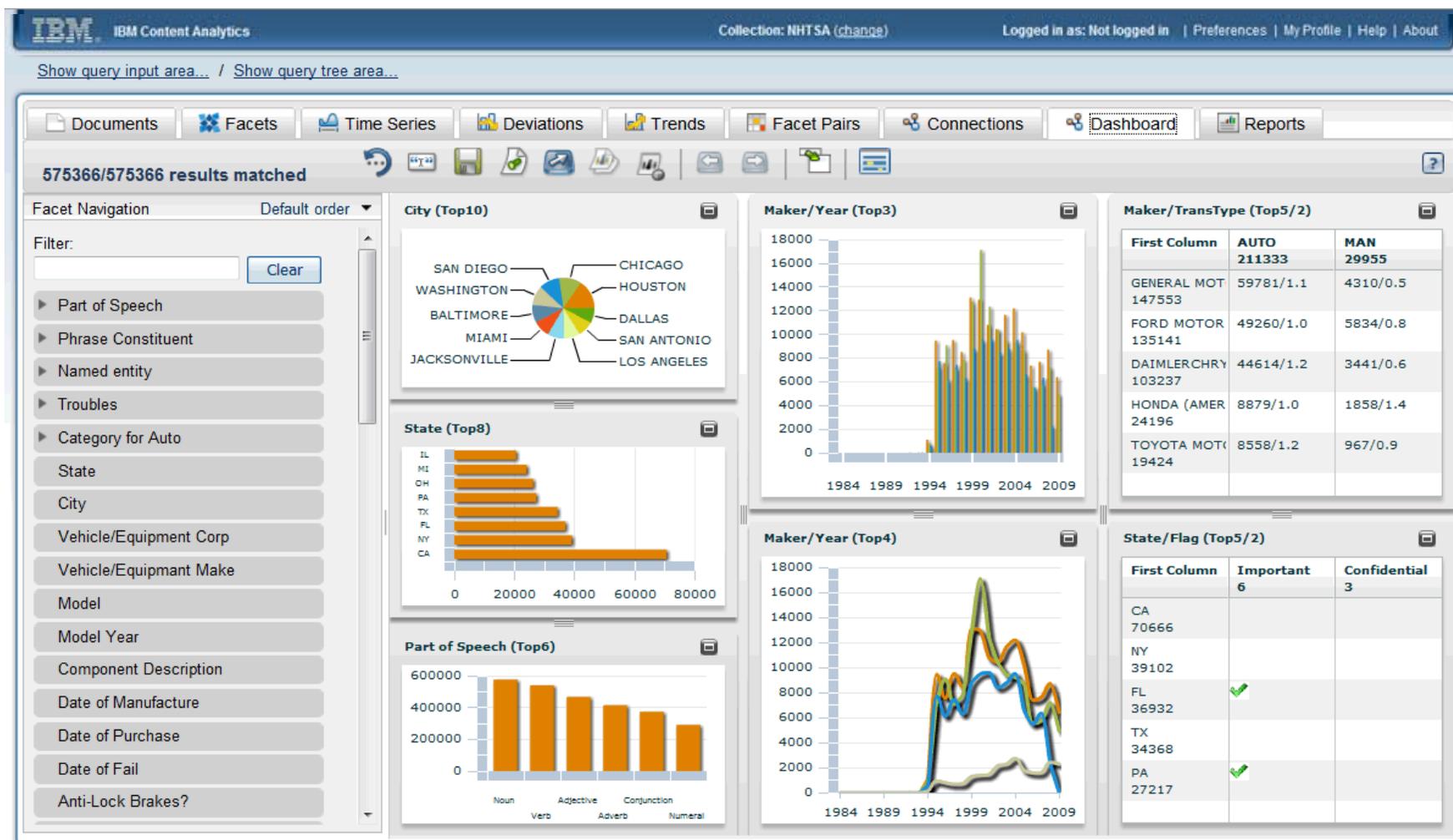
... navigabili ed analizzabili

The image displays several screenshots of the IBM Content Analytics interface, illustrating various text mining capabilities:

- Documents:** Shows a list of search results with document titles and snippets.
- Sentiment:** Displays a bar chart representing sentiment analysis results over time.
- Time Series:** Shows a line chart with data points over a period of years.
- Facets:** Shows a list of filters (facets) for search results, such as 'Category for Auto'.
- Deviations:** Shows a bar chart representing deviations from a baseline.
- Facet Pairs:** Shows a table of relationships between different facets.
- Trends:** Shows a line chart with data points over a period of years.
- Dashboard:** Shows a comprehensive overview of search results and analytics, including a list of keywords and a bar chart.
- Connections:** Shows a network diagram with nodes and edges, representing relationships between entities.

# IBM Content Analytics

## Dashboard per executive summaries



# IBM Content Analytics

## Sentiment analysis

637/637 results matched

Facet Navigation: Default order

Filter: product

Sentiment Table

Sentiment	Values	Positive	Ambivalent	Negative
Admirable 6.5 cu. ft. Super Capacity Gas Dryer (34)		28 82%	2 5%	2 5%
LuthorCorp 7.3 Cu. Ft. 7-Cycle Electric Dryer - White (18)		14 77%	3 16%	1 5%
Extensive Enterprise 4.5 Cu. Ft. 14-Cycle Ultra Capacity High-Efficiency Washer - White (10)		6 60%	2 20%	2 20%
Extensive Enterprise 4.5 Cu. Ft. 14-Cycle Ultra Capacity High-Efficiency Washer - Graphite Steel (10)		9 90%	0 0%	1 10%
Extensive Enterprise SteamDryer 7.3 Cu. Ft. 14-Cycle Ultra Capacity Electric Dryer - Graphite Steel (10)		9 90%	1 10%	0 0%
Extensive Enterprise 3.5 Cu. Ft. 7-Cycle High-Efficiency Washer - White (10)		8 80%	1 10%	0 0%
EC 4.0 Cu. Ft. 26-Cycle King-Size Washer - Silver Metallic (10)		5 50%	2 20%	3 30%
EC 4.0 Cu. Ft. 26-Cycle King-Size Washer - White (10)		0 0%	4 40%	6 60%
EC 3.2 Cu. Ft. 9-Cycle Super Capacity Wash				

Positive Expressions Negative Expressions

Analyze more in the Facets view

great good happy well Great excellent faster super does t quickly delivered pleased with satisfied I love old price sen NOT complain wise bells and whistles great job a must pro can't be beat NOT waste like a lot simple to use Thanks th does everything better easily Thank You greaat I love it eff more than enough efficient prepared can't beat Best of all NOT happier fast Highly recommend perfectly love this NOT problems can't beat

I was looking for n inexpensive dryer, and I needed to have it quickly. I never hear of Admirable

Selezionando una facet viene visualizzata la valutazione del "sentiment" per ciascun valore

### Vista Sentiment analysis

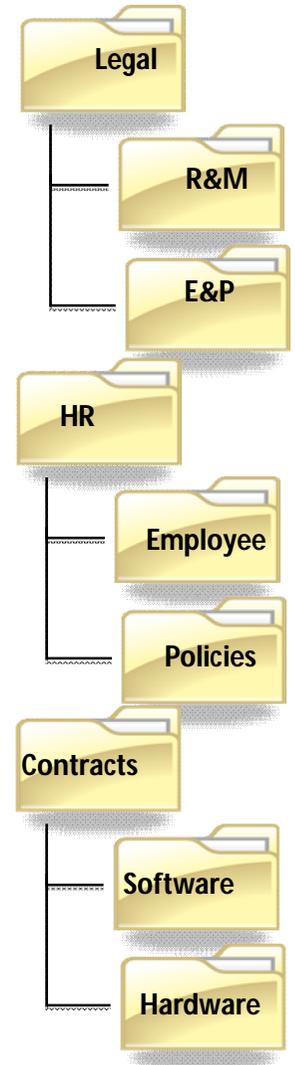
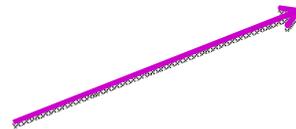
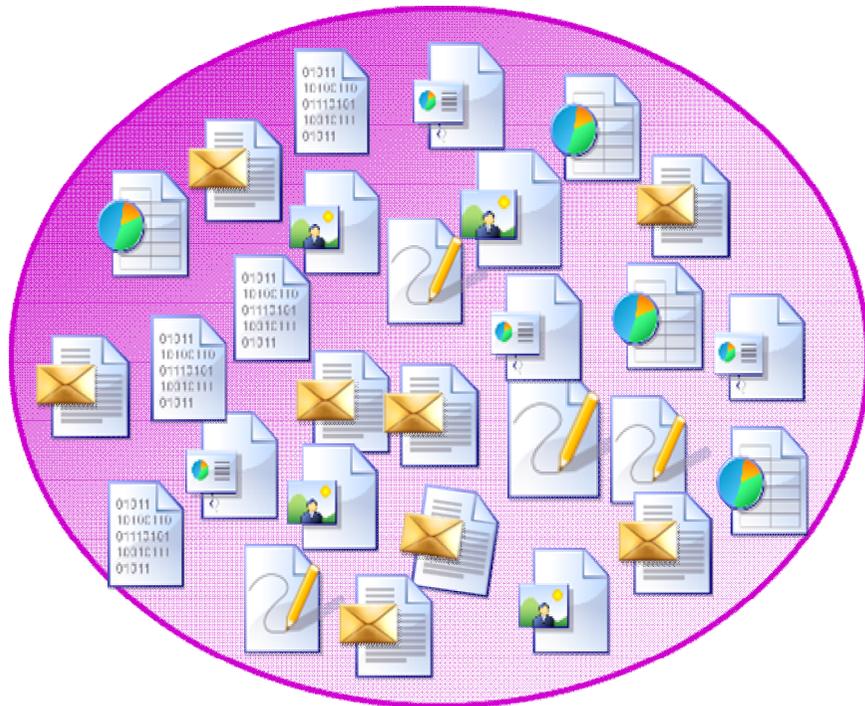
- Elenca i valori della facet selezionata con la valutazione del sentiment per ciascuno
- Mostra il numero di documenti con accezioni positive/ambivalenti/negative e la loro percentuale



# IBM Content Classification

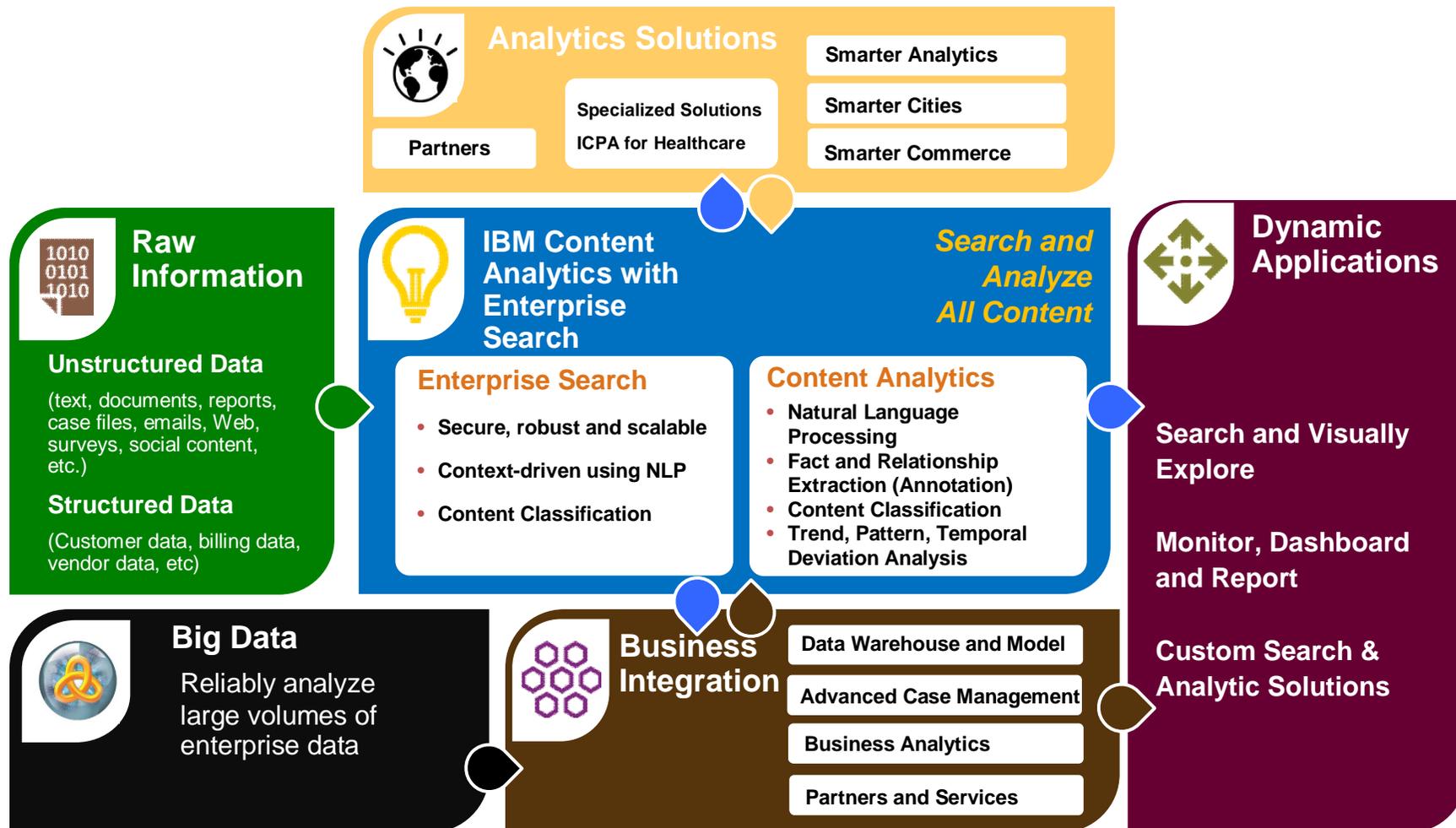
## *Classificazione dei contenuti*

- Elimina i costi della classificazione manuale dei documenti
- Classificazione automatica delle email
- Classifica repository e filesystem preesistenti
- Routing automatico di documenti in base al testo contenuto
- Supporto al processo di dematerializzazione del cartaceo
- Estende la “facet” navigation



# IBM Content Analytics

## Sommario



# ***Alcune esperienze italiane***



# IBM Content Analytics

## Classificazione ed analisi trouble tickets



70.000 utenti interni

3.000 trouble tickets aperti in media ogni giorno

Servizio  
TECNOLOGICO  
APPARECCHIATURE  
STAMPANTE  
LASER / GETTO  
NON STAMPA / STAMPA CARATTERI STRANI

ICM

4-5 campi da riempire per classificare il trouble ticket, dall'utente se utilizza l'applicazione web, o dall'operatore di Help Desk se ticket aperto dall'utente per telefono; Causa complessità classificazione, 90% tickets aperti da utenti via telefono



Analista

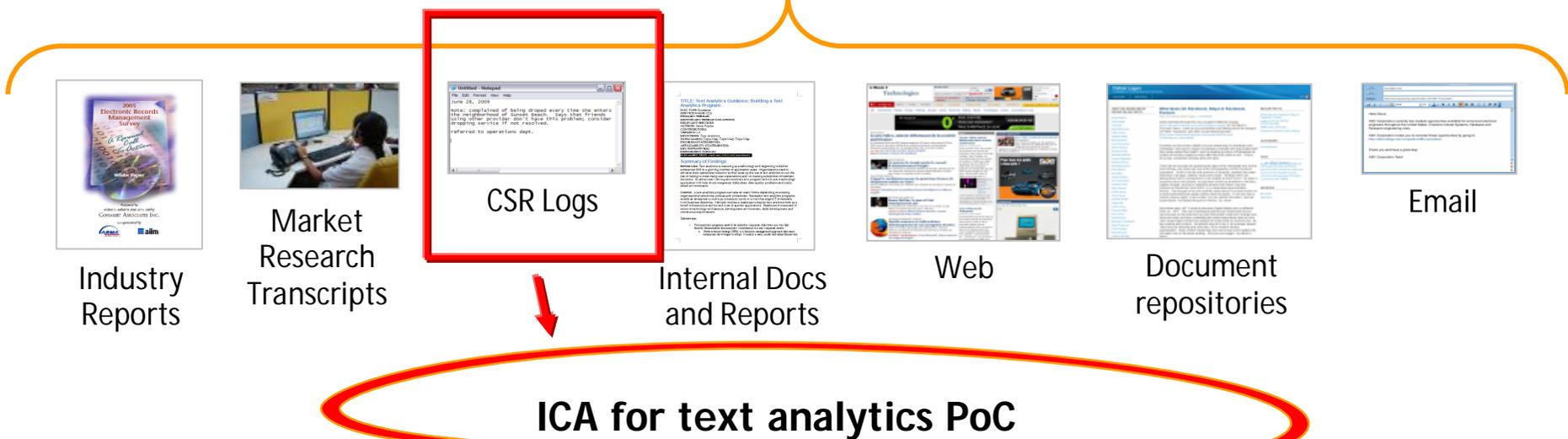
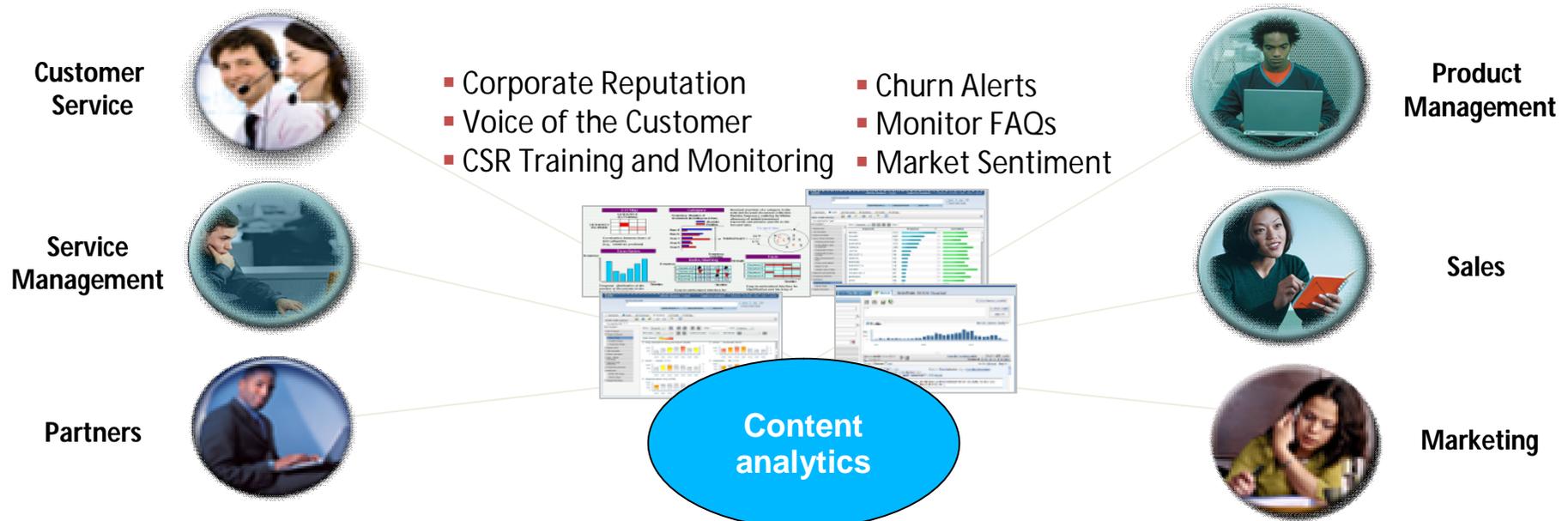
Keywords	Frequency
APPLICAZIONI	289
APPARECCHIATURE	183
MANUTENZIONE	181
BANCOMAT	139
RICHIESTE DIVERSE	77
SOLLECITO	43
Abilitazioni	17
Mail sospetta	17

Servizi di analisi, classificazione e ricerca per:

- Ricerca per fatti specifici
- Problemi più frequenti
- Soluzioni più frequenti
- Correlazioni tra i due suddetti
- Verifica esatta categorizzazione
- Qualità del servizio dei fornitori HD
- Navigazione ed analisi per scoprire informazioni utili a migliorare il servizio HD

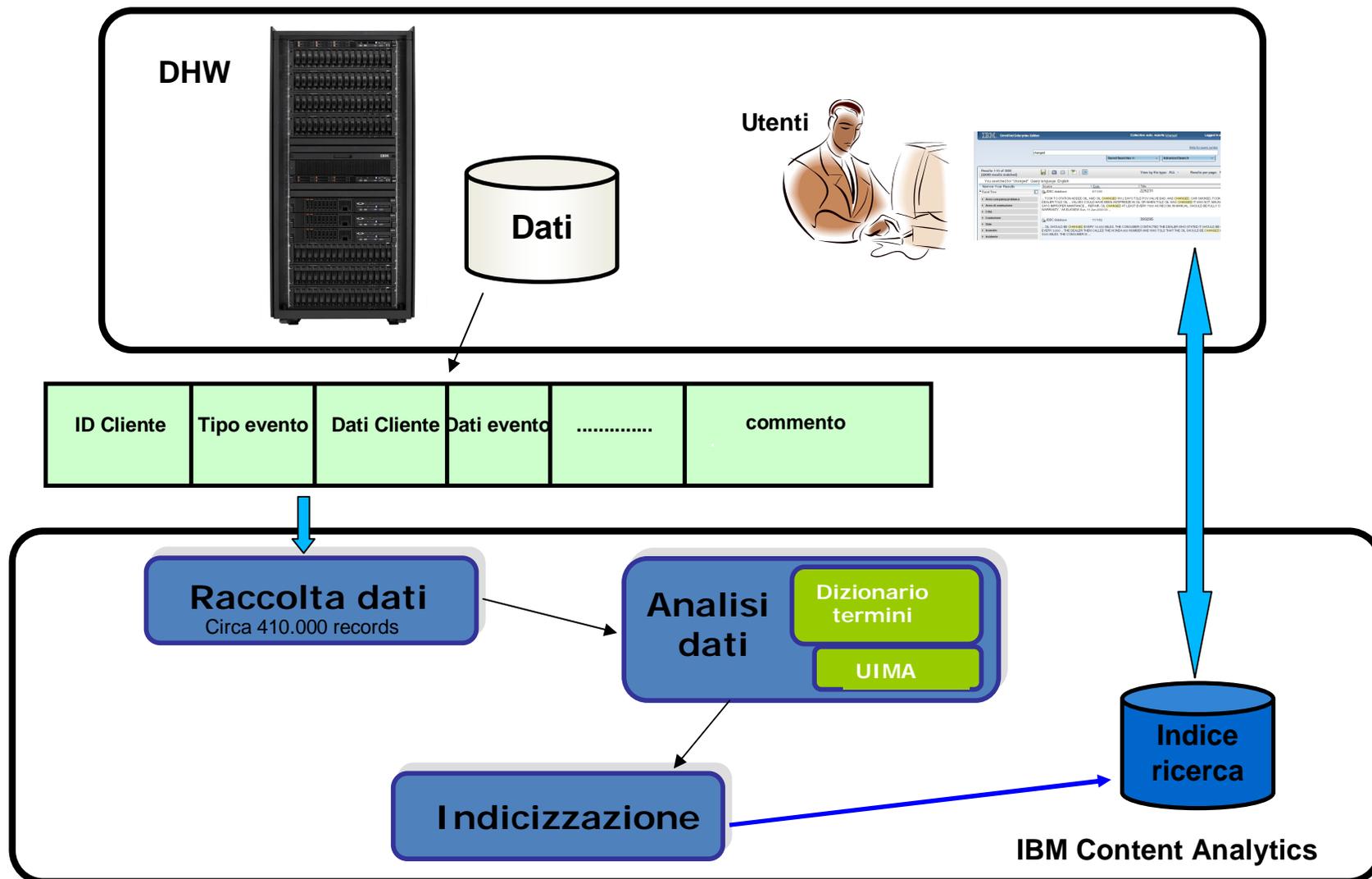
# IBM Content Analytics

## Analisi Contact Center: il contesto



# IBM Content Analytics

## Analisi Contact Center: schema funzionale



# IBM Content Analytics

## Caso d'uso: analisi lamenti e reclami clienti – 1 di 7

1. Analizzati 412000 records circa, relativi a contatti con i clienti da parte dei Contact Center e delle Agenzie, contenenti commenti a testo libero

Raccolta: bank\_demo (modifica)

Documenti Facet Serie temporali Deviazioni Tendenze Coppie di facet Connessioni Dashboard

Risultati 1-10 di 3000  
(412454/412454 risultati associati)

Risultati per pagina

Origine	Data	Titolo
csv	19/03/12	commenti_anag2.txt?id=2
aperto c/c, rilasciato carte bancomat su tutti i cointestatari, no problem su cosentino francesca, e.family su Franco Carne		
csv	07/03/12	commenti_anag2.txt?id=1
valutati gli investim re su obbligazioni bnpp con cedola trimestrale. mi parla de		
csv		commenti_anag2.txt?id=0
IL CLIENTE HA RICHIESTO AGG.POSIZ. FINANZIARIA		
csv	13/03/12	commenti_anag2.txt?id=3
bene fondi su assegni versati nei giorni scorsi, al fine di passare add. carta di credito aziendale.		

Navigatione facet Ordine predefinito

Filtro:  Cancella

- Parte della frase
- Frase costituente
- Codice Cliente
- Dati Cliente
- Data di nascita
- Dati Evento
- Termini di interesse

2. Configurare facets utili all'analisi mappandole a campi presenti nei records

3. In collaborazione con il Cliente identificato il dizionario dei termini di interesse da estrarre dai commenti



# IBM Content Analytics

Caso d'uso: analisi lamentele e reclami clienti – 2 di 7

1. Selezione vista Facet

2. Selezione facet

3. Selezione tutti i valori estratti dai commenti che fanno riferimento ad insoddisfazione clienti

4. Aggiungo i valori selezionati alla query

	Parole chiave	Frequenza
<input type="checkbox"/>	pac	4752
<input type="checkbox"/>	index	4649
<input type="checkbox"/>	fondo	4192
<input type="checkbox"/>	salute	3988
<input type="checkbox"/>	bonifico	2655
<input checked="" type="checkbox"/>	lamentela	761
<input checked="" type="checkbox"/>	reclamo	703



# IBM Content Analytics

## Caso d'uso: analisi lamentele e reclami clienti – 3 di 7

The screenshot displays the IBM Content Analytics interface. At the top, the breadcrumb trail shows 'records' and 'Raccolta: bank\_demo (modifica)'. The navigation bar includes 'Documenti', 'Facet', 'Serie temporali', 'Deviazioni', 'Tendenze', 'Coppie di facet', 'Connessioni', and 'Dashboard'. The 'Connessioni' button is circled in yellow. Below the navigation bar, the text '1452/412454 risultati associati' is circled in yellow. The 'Navigazione facet' section on the left shows 'Dati Cliente' and 'Dati Evento' facets. The 'agenzia' facet under 'Dati Cliente' and the 'oggetto evento' facet under 'Dati Evento' are both circled in yellow. The 'Chiave' section in the center provides a legend for 'agenzia' (dark blue) and 'oggetto evento' (light blue), and a frequency scale from 'Alta' to 'Bassa'. The 'Correlazione Quantità' section is also visible. On the right, a network graph shows nodes for 'AG.ROMA', 'FINANZIAMENTO - MUTUO', 'AG.3 ROMA', and 'NUOVO CONTO CORRENTE'. The 'AG.3 ROMA' node is circled in yellow. A yellow callout box points to the 'Connessioni' button with the text: '2. Passo a vista Connessioni'. Another yellow callout box points to the '1452/412454 risultati associati' text with the text: '1. Drill down che porta a restringere a soli 1452'. A third yellow callout box points to the 'agenzia' and 'oggetto evento' facets with the text: '3. Selezione le due facets che mi consentono di verificare la correlazione di lamentele e reclami rispetto ad agenzie e tipo di richiesta cliente'. A fourth yellow callout box points to the network graph with the text: '4. Le funzioni statistiche della soluzione mettono in evidenza un'alta correlazione tra due specifiche coppie di valori. Per analizzare i motivi di lamentele e reclami all'apertura di nuovi conti correnti sulla specifica agenzia, la seleziono con un click e la aggiungo alla query.'



# IBM Content Analytics

Caso d'uso: analisi lamenti e reclami clienti – 4 di 7

1. Drill down che porta a restringere a soli 13 records

13/412454 risultati associati

Mostra: Parole chiave

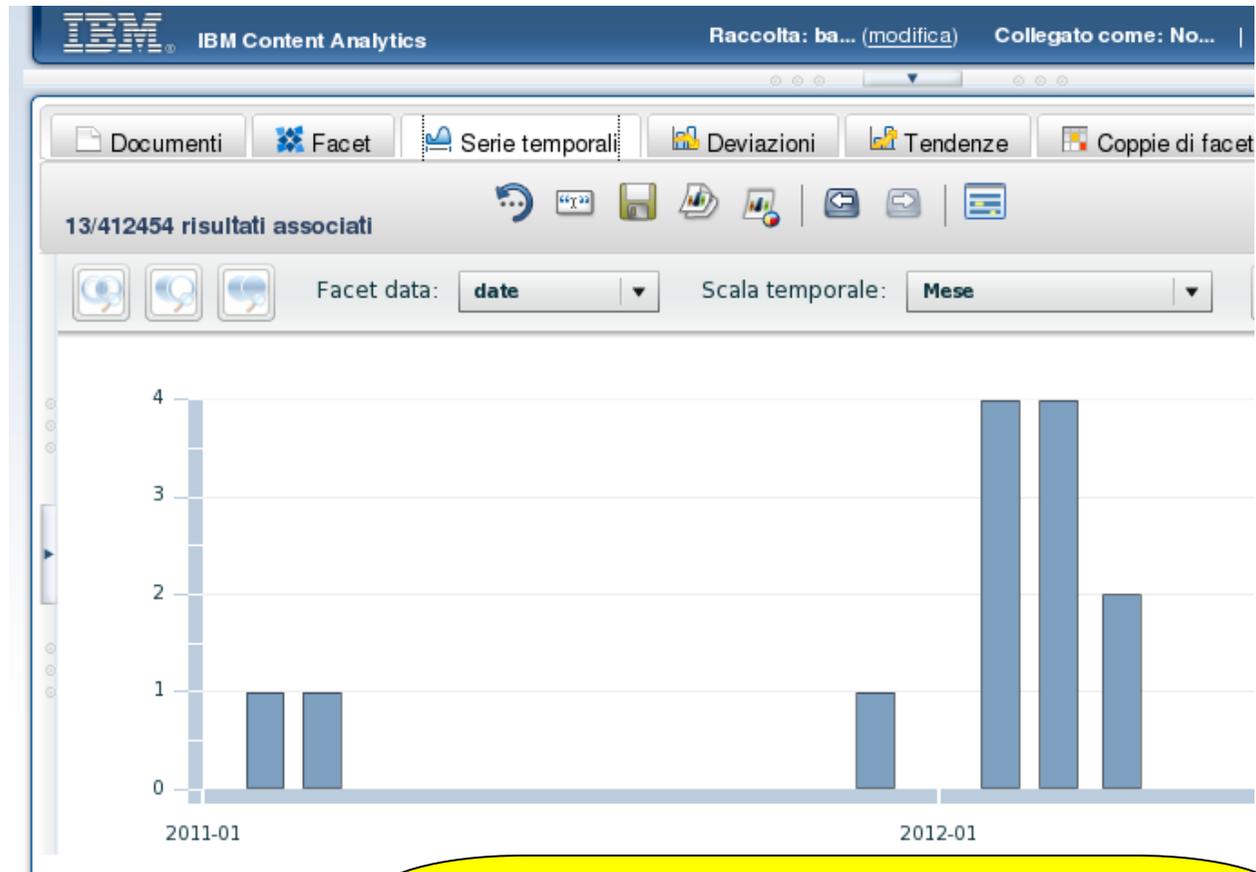
Parole chiave	Frequenza
4629088	2
301220976	2
301258641	2
3725873	1
10537606	1
2386334	1
8677368	1
5837744	1
2423112	1
7792058	1

2. Per comprendere se l'insoddisfazione è concentrata su pochi critici clienti, passo alla vista facet e seleziono Codice Cliente; il problema è oggettivamente distribuito su più clienti



# IBM Content Analytics

*Caso d'uso: analisi lamentele e reclami clienti – 5 di 7*



Il malcontento dei clienti potrebbe essere dovuta a problemi contingenti sorti in agenzia in uno specifico periodo (problemi ambientali, logistici o dipendenti insoddisfatti); la distribuzione nel tempo di lamentele e reclami nella vista Serie Temporali mostra che il problema dell'Agenzia è permanente, e nello specifico legato ai primi mesi dell'anno.

# IBM Content Analytics

## Caso d'uso: analisi lamentele e reclami clienti – 6 di 7

1. Passando alla vista Documenti posso scorrere il sommario dei documenti selezionati per capire le cause di lamentele e reclami

The screenshot displays the IBM Content Analytics interface. At the top, the IBM logo and 'IBM Content Analytics' are visible, along with the collection name 'Raccolta: bank\_demo (modifica)'. Below this is a navigation bar with tabs for 'Documenti', 'Facet', 'Serie temporali', 'Deviazioni', 'Tendenze', 'Coppie di facet', 'Connessioni', and 'Dashboard'. The 'Documenti' tab is selected and circled. The main area shows search results for 'Risultati 1-10 di 13 (13/412454 risultati associati)'. On the left, there is a 'Navigazione facet' section with a 'Filtro:' input field and a 'Cancella' button. Below this are several facet categories: 'Parte della frase', 'Frase costituente', 'Codice Cliente', 'Dati Cliente', 'Data di nascita', 'Dati Evento', and 'Termini di interesse'. The main results table has columns for date and document ID. The first row shows a date of 04/02/11 and ID 'commenti\_anag2.txt?id=359948'. The second row shows 07/03/11 and ID 'commenti\_anag2.txt?id=210880'. The third row shows 21/12/11 and ID 'commenti\_anag2.txt?id=275648'. The fourth row shows 15/02/12 and ID 'commenti\_anag2.txt?id=10880'. A 'csv' icon is circled next to the third row. Callouts provide instructions: 1. 'Passando alla vista Documenti posso scorrere il sommario dei documenti selezionati per capire le cause di lamentele e reclami' (pointing to the 'Documenti' tab); 2. 'E' possibile salvare la query complessa man mano costruita, per rieseguirla quando necessario' (pointing to the 'Facet' and 'Serie temporali' icons); 3. 'E' possibile esportare i documenti risultato dell'analisi per elaborazioni su altri sistemi' (pointing to the 'csv' icon); 4. 'Con un click è possibile visualizzare la sintesi del documento originale per analisi di dettaglio (vedi slide successiva)' (pointing to the 'csv' icon).

2. E' possibile salvare la query complessa man mano costruita, per rieseguirla quando necessario

3. E' possibile esportare i documenti risultato dell'analisi per elaborazioni su altri sistemi

4. Con un click è possibile visualizzare la sintesi del documento originale per analisi di dettaglio (vedi slide successiva).



# IBM Content Analytics

Caso d'uso: analisi lamentele e reclami clienti – 7 di 7

Sintesi del documento originale.

Analisi documenti

Titolo: commenti\_anag2.txt?id=122685  
Data: 27/02/12

Facet analisi	
Nome	Valore
Sostantivo - Predicato	cliente ... chiedere
Sostantivo generale	cliente
Verbo - Sostantivo	chiedere ... appuntamento
Verbo	chiedere
Sostantivo generale	appuntamento
Termini di interesse	appuntamento
Preposizione con sostantivo	per ... informazioni
Termini di interesse	informazione
Sostantivo generale	informazioni
Preposizione con sostantivo	sullo ... richiesta
Sostantivo generale	richiesta
Preposizione con sostantivo	di ... apertura c/c
Sequenza sostantivi	apertura c/c
Sostantivo generale	apertura
Termini di interesse	cc

Facet metadati	
Nome	Valore
Codice Cliente	
agenzia	
area	
CAP residenza	
capogruppo	
classificazione eurisko	
codice agenzia	

cliente chiede appuntamento per informazioni sulla sua richiesta di apertura c/c , **lamentando** di attendere da 2 settimane.

Annulla



# IBM Content Analytics

## *Analisi bonifici bancari*

### **OBIETTIVO**

Realizzazione di un ambiente demo con dati reali cliente per mostrare in che misura questa soluzione può essere di ausilio alle attività di analisi; nello specifico analisi dei dati relativi ai bonifici per scoprire eventuali anomalie.

### **DESCRIZIONE**

- Installazione & configurazione ICA per raccolta record bonifici bancari da ambiente RDBMS di test
- Indicizzazione ed analisi di circa 500.000 record
- Estratte informazioni specifiche dai campi “descrizione” dei bonifici
- Configurato il prodotto per poter correlare informazioni estratte con dati strutturati (es.:IBAN, dati anagrafici, ecc)

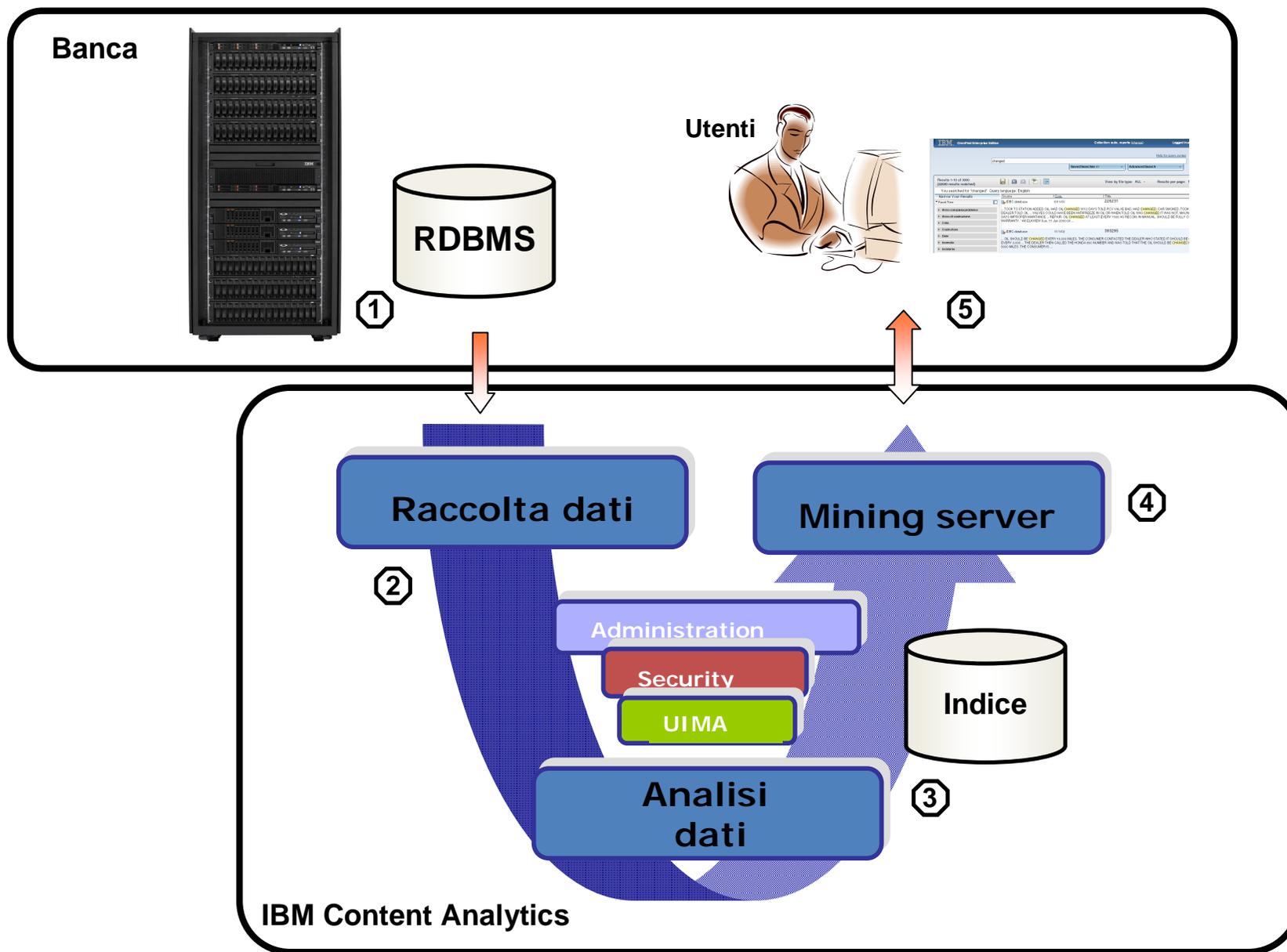
### **BENEFICI**

Possibilità di far emergere comportamenti anomali, fraudolenti o non aderenti alle politiche aziendali nell’ambito delle operazioni di bonifico, sia da parte degli clienti che degli operatori bancari, attraverso una applicazione intuitiva fruibile direttamente dagli utenti di business.



# IBM Content Analytics

## Analisi bonifici bancari : schema funzionale



# IBM Content Analytics

## *Analisi bonifici bancari : vista facets*

467335/467335 results matched

Facet Navigation Default order

Filter:

- Part of Speech ?
- Phrase Constituent ?
- Termini Sospetti
  - Per Categoria
  - Per Termine**
- Dati Ordinante
- Dati Beneficiario
- Dati di Carico
- Dati Movimento
- Importi per valore

Search type:

Facet Path:

Keyword:

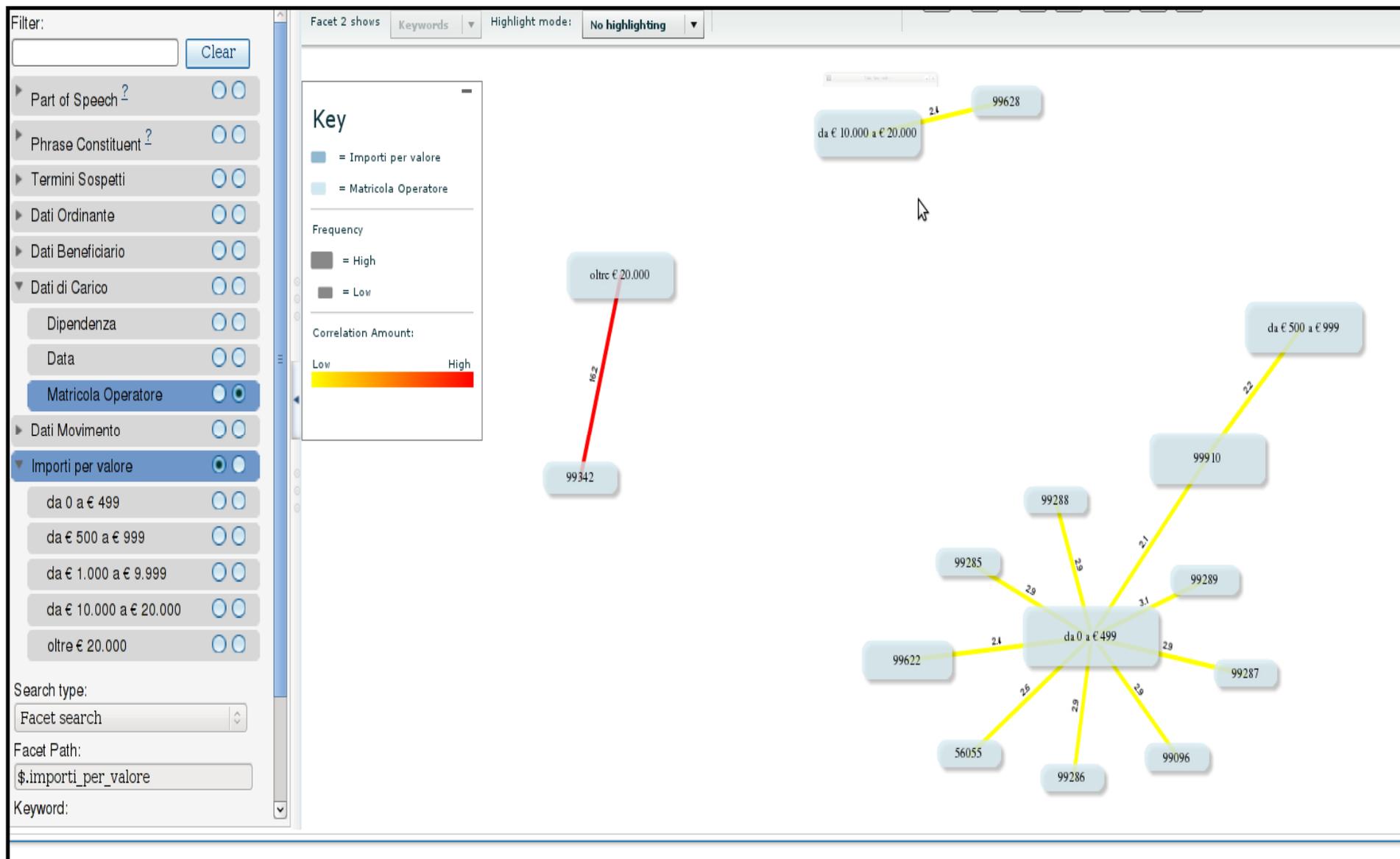
New search  
 Add to search

<input type="checkbox"/>	Keywords	Frequency	
<input type="checkbox"/>	acconto	10325	<div style="width: 100%;"></div>
<input type="checkbox"/>	giroconto	3055	<div style="width: 30%;"></div>
<input type="checkbox"/>	rimborso spese	1209	<div style="width: 12%;"></div>
<input type="checkbox"/>	caparra	625	<div style="width: 6%;"></div>
<input type="checkbox"/>	provvigione	388	<div style="width: 4%;"></div>
<input type="checkbox"/>	regalo	384	<div style="width: 4%;"></div>
<input type="checkbox"/>	donazione	271	<div style="width: 3%;"></div>
<input type="checkbox"/>	finanziamento soci	243	<div style="width: 2%;"></div>
<input type="checkbox"/>	interessi	202	<div style="width: 2%;"></div>
<input type="checkbox"/>	presente	187	<div style="width: 2%;"></div>
<input type="checkbox"/>	deposito cauzionale	144	<div style="width: 1%;"></div>
<input type="checkbox"/>	come da accordi	136	<div style="width: 1%;"></div>
<input type="checkbox"/>	prestito infruttifero	49	<div style="width: 0.5%;"></div>
<input type="checkbox"/>	prestito soci	37	<div style="width: 0.4%;"></div>
<input type="checkbox"/>	rimborso soci	16	<div style="width: 0.2%;"></div>
<input type="checkbox"/>	elargizione	8	<div style="width: 0.1%;"></div>
<input type="checkbox"/>	quota rimborso	3	<div style="width: 0.03%;"></div>
<input type="checkbox"/>	usura	2	<div style="width: 0.02%;"></div>



# IBM Content Analytics

## Analisi bonifici bancari : vista connessioni



# IBM Content Analytics

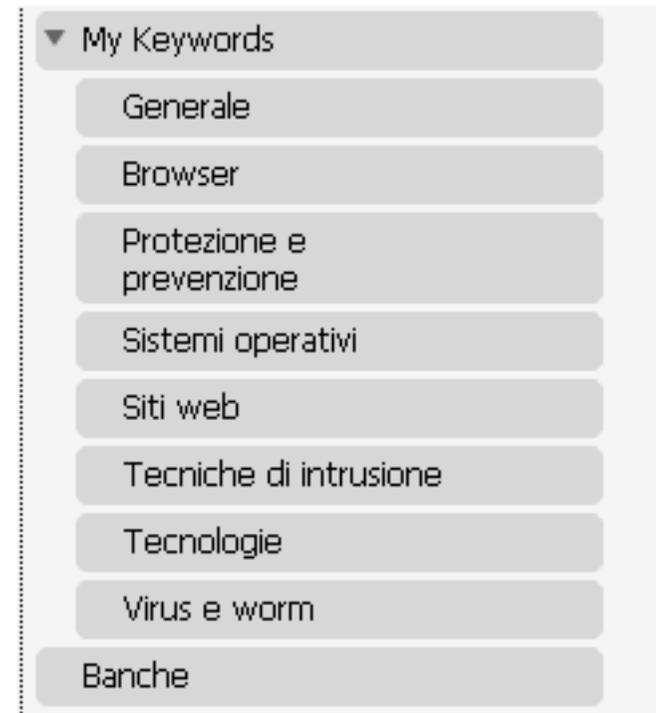
## *Internet banking fraud prevention*

### OBIETTIVO

Realizzare un ambiente che consenta di automatizzare e velocizzare l'esplorazione di una determinata lista di siti internet (lista modificabile a piacere) per trovare tempestivamente informazioni relative a potenziali attacchi informatici a banche o similari dalla rete internet (es.: trojan, viruses, phishing, etc)

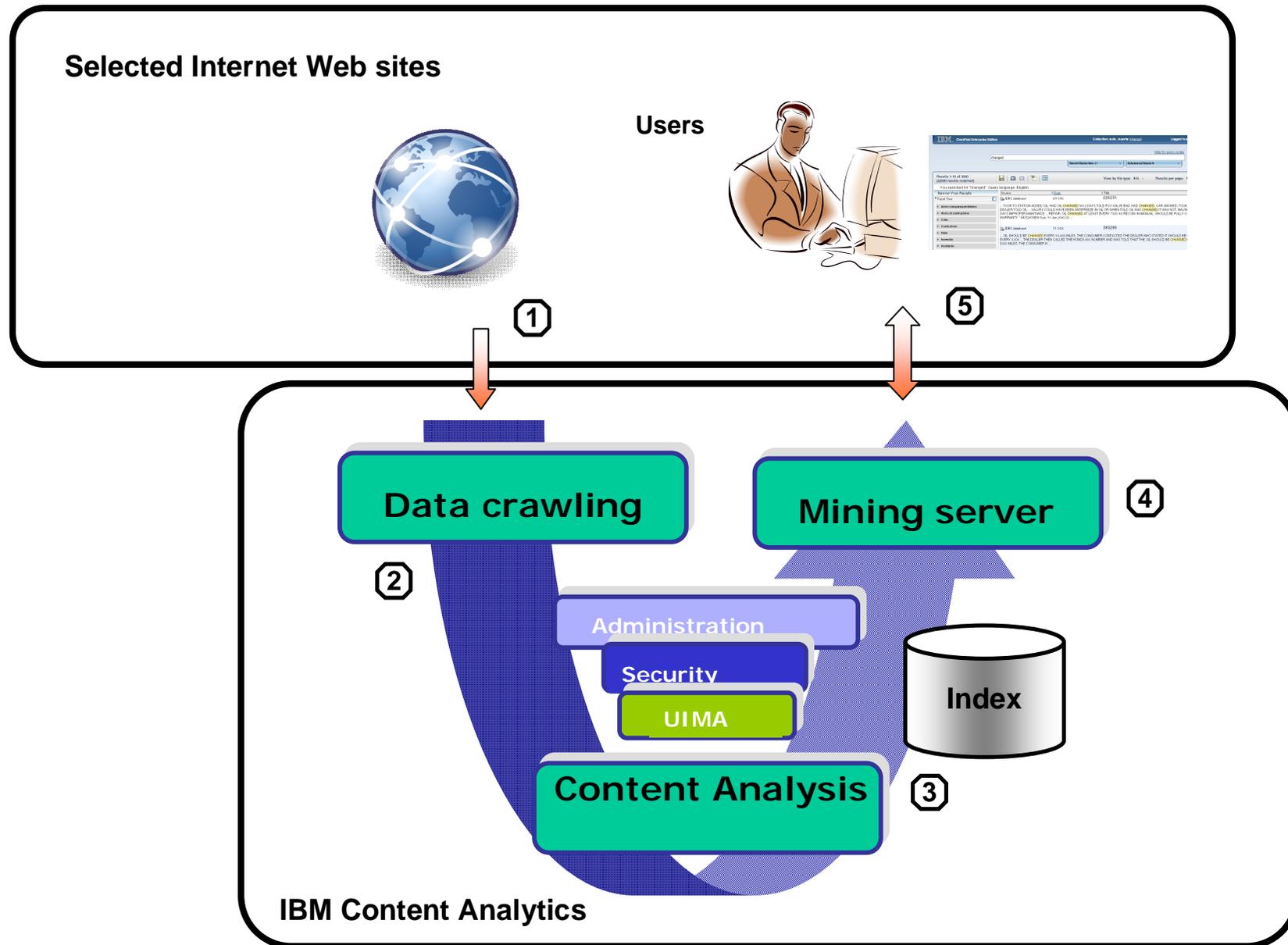
### DESCRIZIONE

- Installazione e configurazione di ICA per effettuare indicizzazione di pagine web da siti selezionati
- Indicizzazione ed analisi di 500 GB di dati
- Estrazione delle informazioni utili attraverso due dizionari contenenti circa 200 parole suggerite dagli esperti di sicurezza del cliente
- Configurazione di ICA per esplorare ed analizzare i contenuti dei web sites monitorati



# IBM Content Analytics

## Internet banking fraud prevention: schema funzionale



Thank  
YOU

