

Text Mining per Automatic Intelligence

Universita' di Bologna
4 Dicembre 2006

Ing. Alessandro ZANASI

TEMIS (Fondatore)

Alessandro.zanasi@temis.com

Speaker Background

• Tecnico-Scientifico:

- Ingegnere Nucleare (Univ. Bologna) - *Specializzato in Matematica Applicata (Univ. Paris)*
- *16 anni in IBM (Bologna, Parigi, San Jose'-California)*
- Docente e Ricercatore (Universita' di Bologna, Malta e Southampton) – *Knowledge Management e Data Mining*
- Imprenditore (fondatore di Temis, azienda di sviluppo SW per web intelligence, in Italia, Francia, Germania, UK e USA).

• Sicurezza-Intelligence:

- *Ufficiale Carabinieri (in congedo) - Identificazione parlatore in intercettazioni telefoniche (CCIS, ora RIS)*
- *Responsabile Market Intelligence Analysis per IBM nel Sud Europa*
- *Analista META Group (Int'l Program Director) per applicazioni di Business Intelligence e CRM*
- *Consulente di numerose aziende e governi di tutto il mondo nel settore Sicurezza e Intelligence*
- *Direttore Centro Studi Intelligence e Sicurezza (Malta, Bologna, Modena e Reggio, Bocconi, Parigi, Southampton, Rio de Janeiro)*
- *Membro ESRAB-European Security Research Advisory Board presso European Commission*



(sviluppo SW Text Mining)

- Fondata nel **2000** da esperti ex-IBM , raggiunti nel **2003** da ex-Xerox (MKMS).
- Uffici **Diretti** in Italia, Francia, Germania, UK, USA; **partners** in Australia, Brasile, Repubblica Ceca, Grecia, Olanda, Spagna.
- **Sedici lingue europee + Arabo, Cinese, Coreano, Giapponese.**
- **Italia:** TIM-Telecom Italia Mobile, Regioni Calabria + Campania, Ministero Interno, Trenitalia, Telcal (TelecomItalia+Finsiel), SMD-Stato Maggiore Difesa, Università di Bologna, Area Science Park, Allaxia, Edilintelligence, FederImprese, Synthema...
- **Unione Europea:** Tre progetti finanziati da Commissione Europea, Parlamento Europeo, EuroPol...
- **Francia:** EADS, France Telecom, Renault, @Carrefour, Total, Credit Lyonnais, Peugeot, Xerox...
- **Germania:** G+J (Bertelsmann), ABB, Dresdner Bank, Fraunhofer Institute, HASE, Lion, MDL...
- **Olanda:** BCT
- **Svizzera:** Novartis, LightObjects, Roche...
- **USA:** Pfizer, Conoco, AskMe...
- **Taiwan:** CCC, ...
- **Università:** Italy→Politecnico Milano, Modena, Calabria, Malta, Germany→Dortmund, France→Marseille, Brazil→Rio de Janeiro
- ...ed alcune altre **confidenziali**.

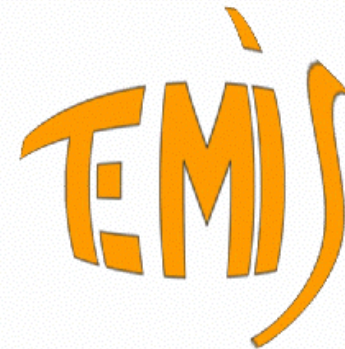


(2005)

Unstructured Information Management Architecture (UIMA)

Partners:

Attensity, ClearForest, Cognos, Endeca, Factiva, Kana, Inquiria, iPhrase, Inxight, nStein, QL2, SAS, Schemalogic, Semagix, SPSS e



Sintesi

1. Il WEB e' la fonte aperta di informazioni che fino a pochi anni fa avremmo considerato strategiche, segretissime, irraggiungibili.
 - **Caso Nord Corea-NC** ha dichiarato di aver costruito armi nucleari leggendo la documentazione su web di AIEA
 - **Caso Al Qaeda-** Al Qaeda ha dichiarato di usare le chatlines per non farsi intercettare...erroneamente
 - ...ma anche indirizzi dei prospects in tutto il mondo dell'Associazione Prosciuttifici di Modena!
2. La difficulta' sta nel recuperare queste informazioni, collegarle, sintetizzarle (ovvero nell'attivita' di *Intelligence*) a causa dei volumi di informazioni messe in gioco (*Information Overload problem*).
3. Il text mining (data mining+linguistica) e' la tecnologia che pemette questa analisi e sintesi in maniera automatica, centrale nel processo di Automatic Intelligence.

Agenda

- 1. Intelligence*
- 2. Automatic Intelligence*
- 3. Problemi dell'Intelligence e Soluzioni di Text Mining*
- 4. Esempi*
- 5. Attuali Ricerche*
- 6. Finanziamenti Europei: ESRAB-ESRP*

1) Intelligence

Intelligence?

James Bond?

No!



*"...i successi piu' importanti dell'intelligence non provengono da azioni di spionaggio "coperto" ma dalla **lettura paziente di riviste specializzate...** "*

*LBJ, presentando il nuovo Direttore della CIA, Richard M. Helms
30/6/1966*



*"La vittoria in una lunga guerra dipende ...dalla **comunicazione strategica**...e da...una **conoscenza linguistica e culturale** considerevolmente migliorata"*

Donald Rumsfeld - Quadrennial Defence Review 2006

2) Automatic Intelligence

Automatic Intelligence?

Non James Bond ma...

In- -Tel

In-Q-Tel e' un fondo di investimento USA creato dalla CIA nel 1999 per investire in start up di tecnologia per intelligence.

Quali tecnologie emergenti?

Quali per automatic intelligence?

3) Problemi dell'Intelligence...

Il problema che oggi ha l'intelligence e' discernere nel mare di informazioni disponibili (email, instant messaging, chatlines, forums, newsgroup, siti web, banche dati online, blogs...) l'informazione di interesse, nascosta dall'eccesso di informazione

...e soluzioni di «Text Mining»

Il text mining comprende «i metodi e gli strumenti per analizzare in maniera automatica grandi quantità di informazioni testuali», rappresentando l'incontro tra tecnologia informatica, matematica e linguistica

4) Esempi di problemi e soluzioni

(le immagini proiettate, provenienti da progetti reali di nostri clienti, non sono state inserite; contattare l'autore in caso di necessita')

...lettura di lingue esotiche...

*...sovraccarico di informazione
multilingua...*

*...individuazione di legami tra
persone provenienti da data base
diversi ("connecting the dots" ...)*

...e le ragioni del legame...

...estrazione di transazioni e concetti finanziari, tecnologici, economici.....e loro trends.....indirizzi email, n. telefono, conti bancari.....

*...informazioni commerciali.. indirizzi
fisici..*

5) Attuali Ricerche: Estrazione delle caratteristiche personali dell'autore

- Lo stile di questa email (...o qualunque testo) e' simile a quella di un particolare gruppo (politico, sociale, religioso)?*
- La frequenza di associazione di alcune parole e' tipica di qualche persona?*
- Quali le caratteristiche dello scrittore (sesso, cultura, tendenze...)?*

6) Finanziamenti Europei

- 1. La Commissione Europea, attraverso ESRAB/ESRP, finanziaria' (3 Miliardi Euro) progetti legati alla Sicurezza... "volano" per il lancio di un nuovo mercato.*

Un insieme di iniziative coerenti della CE

Council

The European Security Strategy
("A secure Europe in a better World",
December 2003)

European Defence Agency - EDA
(start in July 2004)

2004

2005

2006

2007

time

Preparatory Action for Security research – PASR
(2004-2006)

ESRAB

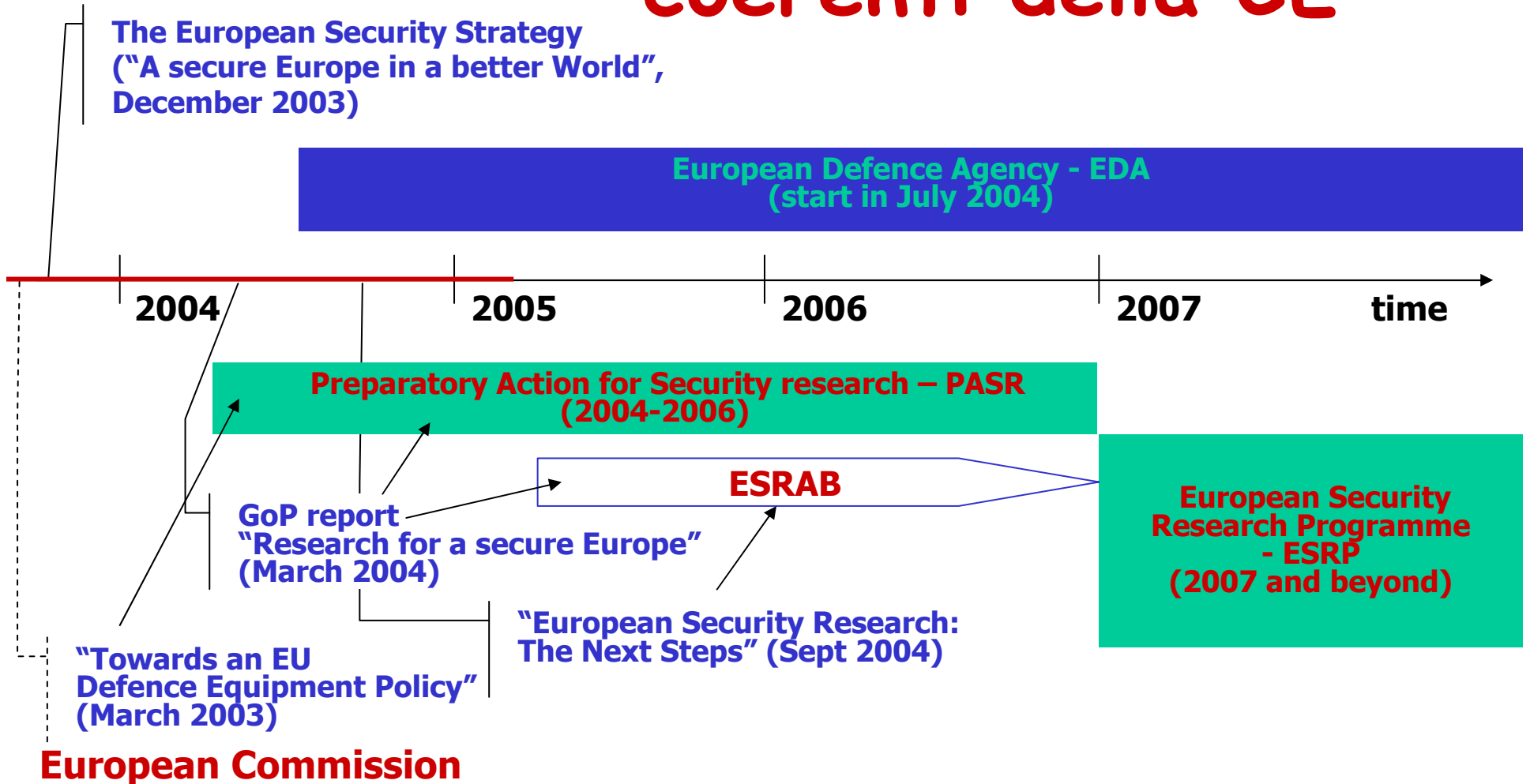
European Security
Research Programme
- ESRP
(2007 and beyond)

GoP report
"Research for a secure Europe"
(March 2004)

"European Security Research:
The Next Steps" (Sept 2004)

"Towards an EU
Defence Equipment Policy"
(March 2003)

European Commission



Tra i 50 membri ESRAB:

Supply/Industrie:

Novartis, Alcatel, SAGEM, Siemens, EADS, Thales, Telefonica, Finmeccanica, Fincantieri, Ericsson, BAE...

Demand/Governi:

EDA, Europol, NATO, DGA, Ministero Ricerca (Francia, Polonia), BKA, Fraunhofer Institute, Ministero Difesa (Grecia, Ungheria, Portogallo), Min.Interni (Italia, UK, Olanda), ...

Think Tanks/Esperti di Sicurezza (It):

Prof.Silvestri
Ing.Zanasi

ESRP

European Security Research Programme

- ESRP si avvia nel 2007 con un suo livello di finanziamento (**€ 3 miliardi**, in aggiunta ai finanziamenti esistenti nazionali o FP)
- Ne potranno beneficiare:
 - Aziende
 - Istituzioni
 - Ricerca

Per saperne di piu':

The Editor: Alessandro Zanasi

Alessandro Zanasi is co-founder and executive vice president of TEMIS SA, and Professor at Modena University, Italy.

For more than twenty years, as an intelligence specialist, he has been consulting for corporations and governments, and giving public and private workshops in Europe, and North and South America.

He graduated in nuclear engineering (Bologna, thesis in systems theory), specialized in probability theory and financial engineering (Paris VI) and in business administration (Modena).

He started his career as a Carabinieri officer in charge of electronic intelligence at Rome Scientific Investigations Center, and worked as a professional consultant for Italian courts and as an information broker. Subsequently, he joined IBM, where he spent 15 years holding different consulting and research positions in Italy, France (Paris) and USA (San Jose), leaving as IBM market intelligence responsible in South Europe. He then became an analyst with META Group Inc. in business intelligence and CRM areas.

Since 1998, he has been European Commission reviewer and evaluator of proposals and projects.

Text Mining and its Applications to Intelligence, CRM and Knowledge Management

Series: Advances in Management Information Volume 2

Organizations generate and collect large volumes of textual data. Unfortunately, many companies are unable to capitalize fully on the value of this data because information implicit within it is not easy to discern.

Primarily intended for business analysts and statisticians across multiple industries, this book provides an introduction to the types of problems encountered and current available text mining solutions.

Partial Contents: Text processing and information retrieval; Application integration in applied text mining; ROI in text mining projects; Open sources automatic analysis for corporate and government intelligence; Customer feedbacks and opinion surveys analysis in the automotive industry; Text mining based knowledge management in banking; Text mining in life sciences; Link analysis in crime pattern detection.

Other titles of interest:

**Handling Missing Data
Applications to Environmental Analysis**
Editors: G. LATINI and G. PASSERINI
ISBN: 1-85312-992-5 2004 200pp+CD-ROM

**Data Mining VI
Data Mining, Text Mining and Their Business Applications**
Editors: A. ZANASI, C.A. BREBBIA and N.F.F. EBEBKEN
ISBN: 1-84564-017-9 2005 568pp

ISBN: 1-85312-995-X



EDITOR
A. Zanasi

Text Mining and its Applications to Intelligence, CRM and Knowledge Management

Text Mining and its Applications to Intelligence, CRM and Knowledge Management



WITPRESS

Email: witpress@witpress.com
www.witpress.com

ISBN 1-85312-995-X



WITPRESS

EDITOR
A. Zanasi

Coordinate:

alessandro.zanasi@temis.com

Tel: 349 41 31 718

Domande?