

New Tools for New Intelligence: Text Mining and European Commission Funding



(NB: The images presented at the Conference, belonging to real confidential business cases, are not attached)

*"La Inteligencia Competitiva"
Conference
Madrid, November 29, 2007*

Ing. Alessandro ZANASI
*TEMIS Cofounder
ESRIF Member
Alessandro.zanasi @ temis.com*

Executive Summary

1. After 9/11/2001, Intelligence reshaped itself: dispersed, open to several sources and to new IT to solve the **information overload problem** (as Text Mining does).
2. **CI is important in the government**, due to the changing nature of national power → energy production control, industrial and financial power are the new issues → science, technology and business factors and capacity of detecting players and actions
3. European Commission, through ESRAB/ESRIF, defined these new IT tools central in **Security Research**, and through various initiatives (DG JLS, ENT, RELEX, INFO...) will fund (some billions Euro) its development.

Summary of the summary...

*Text Mining,
an **Intelligence** Technology
funded by EC*

Agenda

1. Speaker background

2. Text Mining

3. EC Funding

4. Text Mining in CI: Examples

1. Speaker Background

Speaker Background

• Technical-Scientific:

- *Nuclear Engineer - Universita' di Bologna*
- *Specialized in Financial Engineering and ALM - Paris University*
- *16 years in IBM (Bologna, Parigi, San Jose'-California)*
- *Coordinator of KDD Center at CINECA (Italian Supercomputing Center)*
- *META Group Analyst (Business Intelligence)*
- *Professor at Universita' di Bologna - Data and Text Mining Courses*
- *Reviewer and evaluator for EU funded projects*
- *Founder of TEMIS, SW development company for automatic intelligence.*

• Intelligence/Security:

- *Carabinieri Officer (retired), charged of phone calls interception at Centro Carabinieri Investigazioni Scientifiche (now RIS)*
- *Market Intelligence responsible for IBM in South Europe*
- *Frequente speaker and chairman in int'l conferences dedicated to intelligence (eg: Cadiz, Spain - May 2008)*
- *Intelligence and Security Study Center Scientific Director*
- *ESRAB/ESRIF - European Security Research Advisory Board and Innovation Forum Member*
- *Author of several articles and books on intelligence techniques.*
- *Advisor in Intelligence and Security solutions development and fund raising*

TEMI (SW Company)

- Founded in **2000** by me with (Italian, French, German, US) ex-IBM text mining experts, joined in July **2003** by ex-Xerox team.
- **Direct** Offices in Italy, France, Germany, UK, USA; **partners** in Australia, Brazil, Greece, Holland, Spain.
- **Sixteen (European) + Arabic, Chinese, Korean, Japanese** languages are supported.
- **Italy:** IBM, TIM-Telecom Italia Mobile, Trenitalia, Telcal (TelecomItalia+Finsiel), SMD-Stato Maggiore Difesa, University of Bologna, Area Science Park, Allaxia, Edilintelligence, FederImprese, Firenze Tecnologia, Ministero Interno, Synthema...
- **European Union:** Three European Commission funded projects, European Parliament, Europol,...
- **France:** EADS, France Telecom, Renault, @Carrefour, TotalFina, Credit Lyonnais, Peugeot, Xerox...
- **Germany:** G+J (Bertelsmann), ABB, Dresdner Bank, Fraunhofer Institute, HASE, Lion...
- **Holland:** BCT
- **Switzerland:** Novartis, LightObjects, Roche...
- **USA:** Pfizer, Conoco, AskMe...
- **Universities:** Italy→Politecnico Milano, Modena, Calabria, Malta, Germany→Dortmund, France→Marseille, Brazil→Rio de Janeiro
- ...and some other, **confidential** ones.

2. Text Mining

Text Mining

Target

To “read” quickly huge amount of text,

“discovering” concepts, trends, opinions, hidden knowledge...

Techniques

Clustering, to group thematically coherent documents

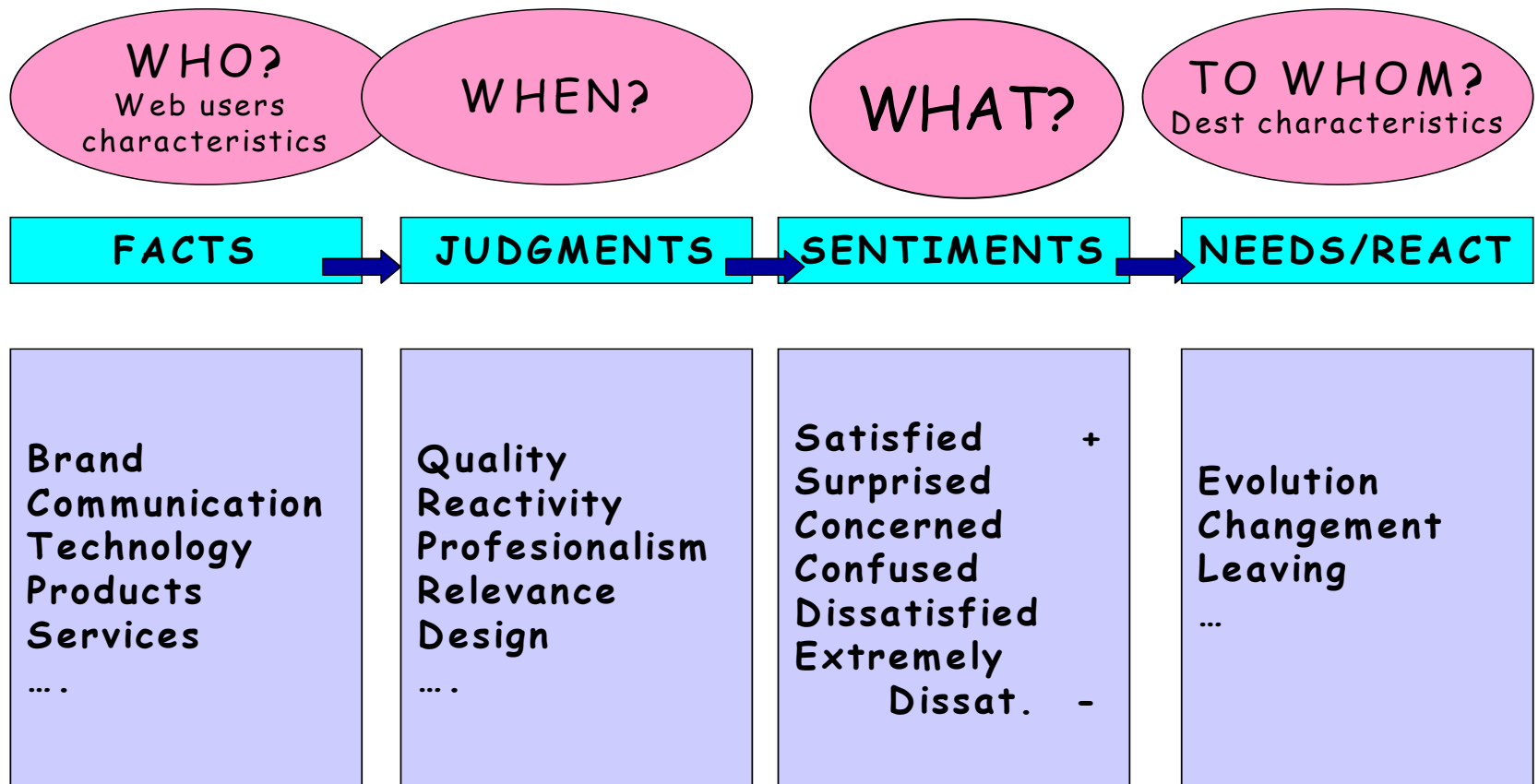
Categorization, to assign documents to predefined categories.

Extraction, to retrieve information from documents (e.g. Feelings)

Packaged Language Model

Skill Cartridges, (*ontologies*), constituted by a specific vocabulary, plus a set of explicit assumptions regarding an intended meaning

If DM is data driven, TM, as *intelligence analysis*, is conceptually driven...



...whose concepts are flexibly extracted by Skill Cartridges:

```
<concept name="~DeclarationVerb">
  report
  post
  have
</concept>
```

```
<concept name="~Country">
  U\ .K\ .
  France
  Italy
</concept>
```

```
<concept name="~Revenue">
  revenue
  gain
</concept>
```

"Geos **France** **posted** **revenues** of 1,494 million euros."

```
<concept name="~Company">
  Jungle / Telecom
  Geos / (~Country)?
</concept>
```

```
<concept name="~RevenueExpression">
  ~Revenue / of / ~Money
</concept>
```

```
<concept name="~RevenueRelation">
  ~Company / ~DeclarationVerb / ~RevenueExpression
</concept>
```

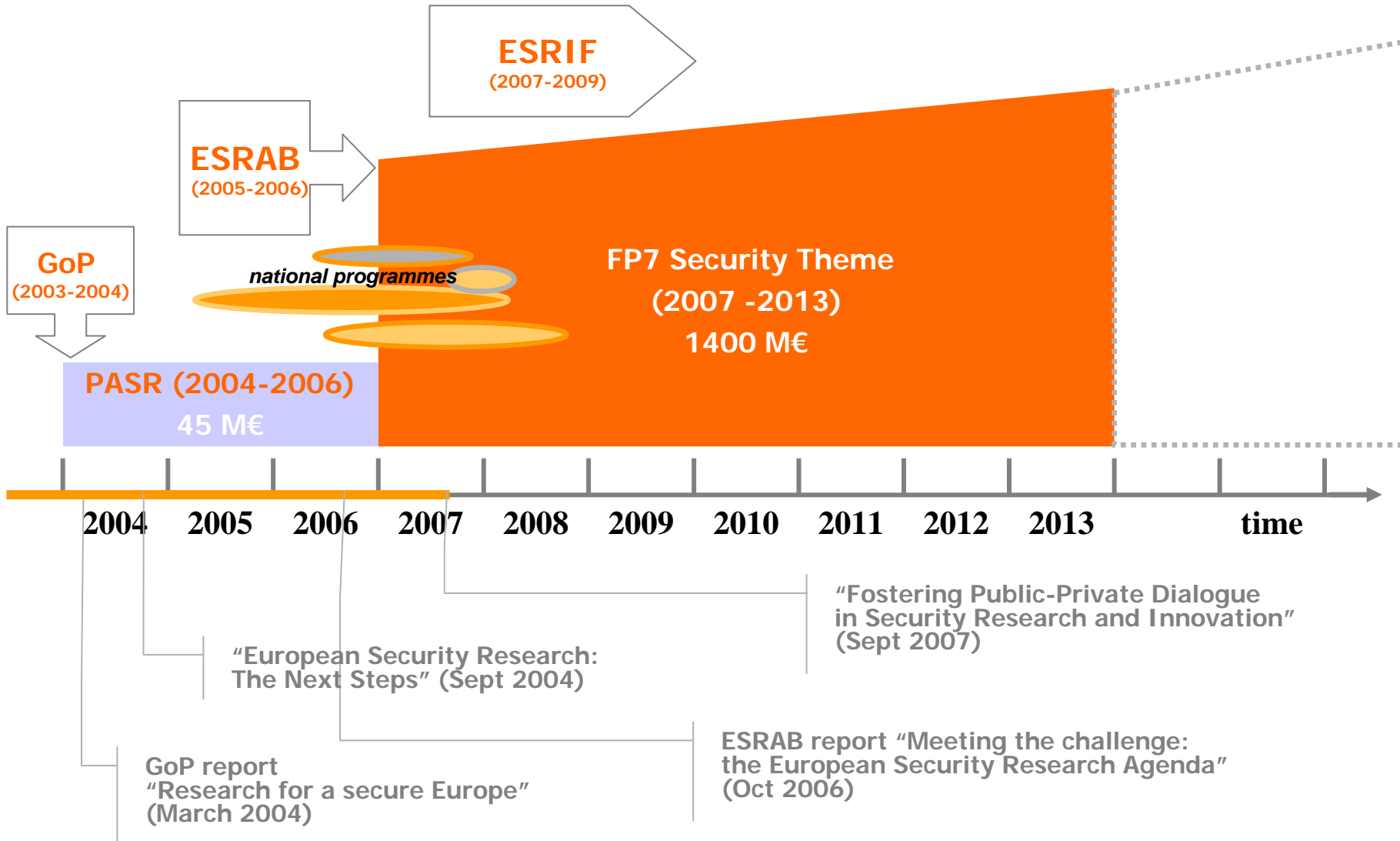
3. EC Funding



a. Why European Security Research ?

- *Unfortunately, Security is a **hot** topic now for European citizens...*
- *So, European Union has decided to fund an initiative*
 - *To use technology to support and enable **security***
 - *To contribute, through technology and innovation, to reinforcing the **competitiveness** of European security industry*

Evolution of European security research



b. ESRAB

European Security Research Advisory Board

- **50** members (25 Member States represented): strategists and experts from stakeholder groups including private & public users
- **Mission**
 - Consult and liaise with Security stakeholders (industry, research, academia, EDA, users) to prepare ESRP funding
 - **Advise on long term strategic vision & research agenda for security**

ESRIF

European Security Research and Innovation Forum

- **60** members (27 Member States + Associates represented):

- **Mission**

Enlarge ESRAB scope and advise on public private partnership

c.ESRP

European Security Research Programme

- *ESRP, from 2007 onwards.*
- *ESRP with an appropriate level of funding, additional to existing national or FP funding (some Euro billions)*
- *ESRP, has its own set of procedures and rules for implementation, participation and funding*

Technologies funded, ranked for importance

Dominio Tecnologico	Aree Tecnologiche Prioritarie
Signal & Information technologies	Data fusion techniques, Data collection/ data classification, Image / pattern processing technology, Information fusion technology, Data and information management technology (DB, ...)
Artificial Intelligence & Decision support	Text-mining / data-mining, IKBS/AI/Expert techniques, Knowledge Management, Modelling and simulation, Optimisation and decision support technology
Sensor Equipment	Cameras, radar sensor equipment, NRBC sensors (in particular biological and chemical threat detection technologies), passive IR sensors equipments
Sensor technologies	Hyperspectral / multispectral sensors, Hyperspectral / multispectral processing, autonomous small sensors / smart dust technologies, IR sensor technologies, Terahertz sensors, Optical sensors Technologies, Acoustic sensors - passive
Communication equipment	Re-configurable communications, Mobile secured communications, Communications network management and control equipment, network supervisor, Network and protocol independent secured communications, Information security, Secured, wireless broadband data links for secured communications, Protection of communication networks against harsh environment
Human sciences	Human behaviour analysis and modelling, Population behaviour, Human factors in the decision process, Teams, organisations and cultures
Information Security Technologies	Encryption and key management, Data mining, Access control, Filtering technologies , Authentication technologies, Encryption technologies (cryptography)
Computing Technologies	Protocol technology, SW Architectures, Secure computing techniques, High performance computing, High integrity and safety critical computing, Software engineering
Information Warfare / Intelligence systems	Infrastructure to Support Information Management & Dissemination, Cyber security policy management tools, Optimisation, Planning & Decision Support systems

4. Text Mining in CI: Examples

TM allows:

*...extraction of financial, economic
and technological transactions and
concepts*

and their trends.....

*email addresses, phone numbers,
bank accounts.....*

...people links detection, with info coming from heterogeneous data bases ("connecting the dots" ...)

...and the reasons of the link...

*Some problems into which we are...
and some current solutions*

...exotic languages reading...

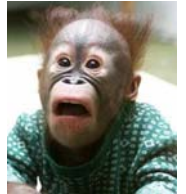
*...multilanguage information
overload ...*

So...

*if interested in asking for European
funding for Security Research...*

...please leave your business card !

Text Analytics Inc.



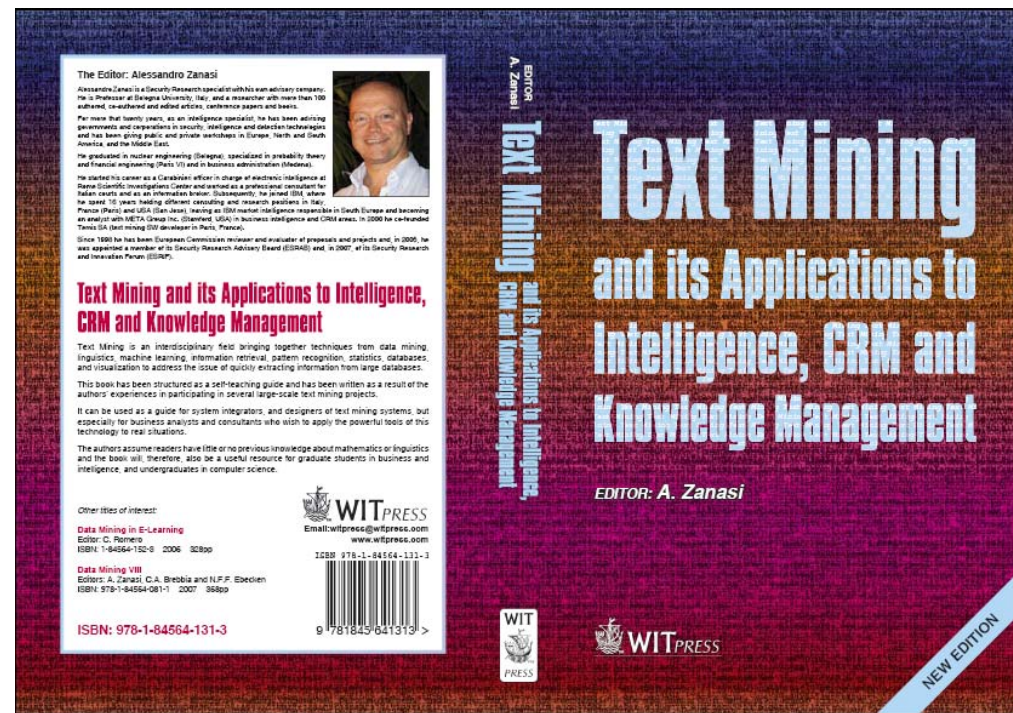
**New York
London
Tokio**

**Dr. John SMITH
CEO**

To know more...

Text Mining and its Applications to Intelligence...

Edited by: A. ZANASI,



Primarily intended for business analysts and statisticians across multiple industries such as retail, telecommunications, financial services, government and pharmaceuticals, this book provides an introduction to the types of problems encountered and current available text mining solutions. It will also be suitable for use on undergraduate courses specializing in this area, and on business oriented graduate and postgraduate courses.

Contributors: Natasa MILIC (Microsoft), Maria Teresa PAZIENZA (Univ.Roma), Fabrizio SEBASTIANI (CNR, Italy), Federica MANDREOLI (Univ.di Modena), Dunja MLADENIC (J.Stefan Institute-Slovenia), Dan SULLIVAN (Redmont Technologies), Mascia FERRARI (Univ. Modena and Reggio), Luc GRIVEL (CNRS, France), Alessandro ZANASI (Temis), Kai LEBETH (Dresdner Bank), Gunter PETERS (G+J/Bertelsmann, Germany), Alessandro POLITI (Ministry of Defense, Italy), Stefano DE ROSSI (TIM, Italy), Leandro WIVES (Universidade do Rio Grande, Brazil), J. FLUCK, H. DENEKE, C. GIEGER (Fraunhofer Institute), F.NERI (Synthema), S.ANENIEV (Megaputer).

...or contact:

Alessandro.zanasi @ temis.com

Tel: (+39) 349 41 31 718