Italian National Agency for New Technologies,
Energy and Sustainable Economic Development

TaLTac in ENAGRID
Toward Parallel Text Mining of Big Data in massive Cultural Textual Corpora

Text Mining & Analytics in ENEAGRID

HPC Cresco 2018 Results

Daniela Alderuccio ï Fiorenzo Ambrosino ï Samuele Pierattini -- DTE-ICT-HPC

Attività Text Mining in Report CRESCO Results 2018

WBS DTE-ICT-HPC https://www.eneagrid.enea.it/DTE_ICT_HPC/jul_2019/DTE_ICT_HPC_WBS.htm

Linea di attività HPC

2. Sviluppi Applicativi (status ATTIVO)

2.1 BIG DATA

2.1.3 Text Mining (TIGRIS, ASTEC, TALTAC, etc.)

2.1.4 Web Crawling (casi d'uso: Finance (cryptocurrencies cfr. MIDAS 2019 (Santomauro, Ponti, Alderuccio, et al.)

Industry 4.0 - Business Intelligence, etc.,

PARALLEL TEXT MINING OF LARGE CORPORA IN GRID ENVIRONMENT - TALTAC IN ENEAGRID INFRASTRUCTURE

Silvio Migliori[1], Daniela Alderuccio[1], Fiorenzo Ambrosino[1], Antonio Colavincenzo[1], Marialuisa Mongelli[1], Samuele Pierattini[1], Giovanni Ponti[1], Andrea Quintiliani[1] ,

Sergio Bolasco[2], Francesco Baiocchi[3], Giovanni De Gasperis[4]

[1]*ENEA – DTE-ICT Division – Roma, Italy*

[2]*Sapienza Università di Roma, Italy* - [3] *Staff TaLTac, Roma, Italy* - [4]*Dip. DISIM Università dell'Aquila, Italy*

ENEAGRID approach
toward Parallel Text Mining of Big Data in massive cultural Textual Corpora.

Keywords: ***Text Mining Software, Cloud Computing, Digital-Humanities, Socio-Economic Sciences, Big Data, Artificial Intelligence, Cybersecurity, Business Intelligence, AltaGamma.***

In: "High Performance Computing on CRESCO Infrastructure: research activity and results 2018", pp. 227-232 - ISBN: 978-88-8286-390-6

ENEAGRID for Research Communities

External user affiliation	
UniSA	2
Uni Madrid	1
UniLaSapienza	5
ISPRA	1
Consorzio RFX	1
UniCamerino	1
UniTorVergata	1
Vinča Institute Belgrade	1
CREATE Consortium	1
UniRC	1
UniFI	1
INM-Germany(Julich)	1
UniNA	1

From e-Science to other Research Communities :

- **Text Mining**
- **e-Humanities**
- **Heritage Science**
- **Big Data**
- **Web Mining**
- **OSINT**
- **…**

ENEAGRID DATA & BIGDATA Sources

The availability of significant amount of computed results c application of the Big Data, Deep learning techniques

Examples of the data available in ENEAGRID/CRESCC

- Climate/energy related forecasts (~100TB)
- **Air pollution models and air quality evaluation (~200 TB, 7x10⁶ files)** collab. with
- **Web crawling (~15 TB, 340x10⁶ web pages)**:: presentation in the Workshop
- Nuclear fusion experiments
- HPC Cluster monitoring system (MySQL, 9 GB recording) ZABBIX
- Bioinformatics
- ....

ENEA G.Bracco, Workshop Big Open Data Analysis, Roma Sede 2/2

ENEA HPC participation in H2020 projects

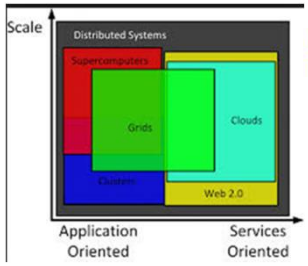
Topics of interest and opportunities: **(2016 -> ETP4HPC member)**

- HPC infrastructures & technologies,applications (material science), distributed data management, I/O performances, collaboration and remote access tools...
- Many different systems for testing and benchmarking

Running projects:

- **H2020 Centre of Excellence: EoCoE -** *http://www.eocoe.eu/*
  - The European Energy oriented Centre of Excellence for computing applications. EoCoE delivers services to exploit the HPC computing power to accelerate the transition to carbon-free energy technologies in industries and society.
  - **Support tools** for several H2020 projects: SEADATACLOUD, NEXTOWER, M4F, INSPYRE, GEMMA... and to EERA-JPNM (Joint Programme on Nuclear Materials): e.g. Report Management

Future Steps toward Parallel Text Mining



Interest in supercomputing is now worldwide, and growing in many new markets (~50% of Top500 computers are in industry).

Source: Jack Dongarra, «Inaugurazione CRESCO6» 30 maggio 2018

The real power of ENEAGRID will be fully exploited when software will be in MULTI-CORE version

CODE (SW optimized for HW)
To exploit Parallelism

To fully exploit Parallelism in Text Mining in GRID e-Infrastructure

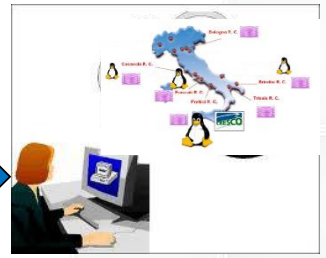
an ADAPTATION OF SOFTWARE APPLICATION IS REQUIRED,
to maximize every form of parallelism within a supercomputer
and use thousands cores simultaneously to solve one large problem

Co-design new CODE Design

WORKLOAD + **CODE** (SW optimized for HW) To exploit Parallelism + **RESOURCES**

Researchers/ Users — **Developers** — **HPC Centers** Access to e-Infrastructure

ENEAGRID
distributed computing environment
"multi-core" architecture
For LARGE CORPORA



ENEAGRID Team

ENEA
Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile

CRESCO ENEAGRID

TaLTaC in ENEAGRID Infrastructure

toward Parallel Text Mining
of Large Text Corpora (from Gigabytes to Terabytes)
In Grid environment

Roma, 15 giugno 2018

Daniela Alderuccio - ENEA DTE-ICT
TaLTaC & Università dell'Aquila

JADT 2018 - International Conference on Statistical Analysis of Textual Data

CRESCO6 in fase di installazione

ENEAGRID/CRESCO 2018 People

http://www.eneagrid.enea.it/people/2018EneaGridPeople.html

Fiorenzo Ambrosino, Giuseppe Aprea, Tiziano Bastianelli, Riccardo Bertini, Irene Bellagamba, Giovanni Bracco, Luigi Bucci, Francesco Buonocore, Marco Caporicci, Michele Caiazzo, Beatrice Calosso, Massimo Celino, Marta Chinnici, Antonio Colavincenzo, Aniello Cucurullo, Pietro D'Angelo, Davide De Chiara, Matteo De Rosa, Daniele Di Mattia, Stefano Ferriani, Gianclaudio Ferro, Claudio Ferrelli, Agostino Funel, Dante Giammattei, Marcello Galli, Simone Giusepponi, Roberto Guadagni, Guido Guarnieri, Michele Gusso, Francesco Iannone, Massimo Marano, Angelo Mariano, Giorgio Mencuccini, Silvio Migliori, Marialuisa Mongelli, Patrizia Ornelli, Simonetta Pagnutti, Filippo Palombi, Salvatore Pecoraro, Antonio Perozziello, Samuele Pierattini, Salvatore Podda, Giovanni Ponti, Andrea Quintiliani, Giuseppe Santomauro, Alberto Scalise, Fabio Simoni, Daniele Visparelli.

Between them a special acknowledgement to the people mostly involved in the daily user support and system management:
Fiorenzo Ambrosino, Antonio Colavincenzo, Agostino Funel, Guido Guarnieri, Filippo Palombi, Giovanni Ponti.

JADT 2018
INTERNATIONAL CONFERENCE ON STATISTICAL ANALYSIS OF TEXTUAL DATA

ENEA ï TaLTaC partnership

Technological Partner

ENEAGRID
digital infrastructure

providing:

- computing power and data storage
- Saas (Software as a Service)
  running Text Mining Tasks in a parallel & distributed environment
- Collaborative research environment

ENEAGRID is open to qualified users.

ENEA Partners access Collaborative research environment via credentials (username and password) to Supercomputers & ICT services:
VPN, Download Software,
Video conferencing, e-Learning, etc).
http://utict.enea.it/it/servizi-di-base/servizi-di-base/#CredenzAsie

End-User

Digital Humanities Community

- Access TaLTaC software:
  - via remote desktop
  - via user interfaces to ENEAGRID
- Share TaLTaC results :
  - via Virtual Laboratories
- Storage TaLTaC data in user environment :
  - AFS user folder

TaLTaC in ENEAGRID

TaLTaC — Trattamento automatico Lessicale e Testuale per l'analisi del Contenuto di un Corpus

TALTAC is the acronym of "Automatic Lexical and Textual Processing for the Analysis of Content" - http://www.taltac.it/en/taltac1.shtml

Taltac is a **software application** for the automatic analysis of texts according to the logics of both **Text Analysis** (TA) and **Text Mining** (TM). Such an analysis allows to define a quantitative representation of the phenomenon under study, both at the level of text-units (words) and context-units (words). Consequently, both the **language** and the **contents** of the text can be examined. TALTAC employs both **statistical** and **linguistic** resources.

TaLTaC originates from **research** carried out at the Universities of Salerno and Rome "La Sapienza" during the 1990s under the supervision of Sergio Bolasco, Professor of Statistics at the Department of geo-economic, linguistic, statistical and historical studies for regional analysis of "La Sapienza" University. It is the result of the cooperation of researchers and colleagues of several Italian and French universities.

Dal 2000 al 2019 sono state svolte svariate **attività di formazione** sul software TaLTaC: **800 partecipanti**; In particolare, 33 corsi di tutorial (corsi base e avanzati), 4 edizioni di una Scuola internazionale sui "*Metodi di analisi dei dati testuali e text mining*" e 4 corsi in due Master universitari di secondo livello rispettivamente in ***Data Science*** (Roma, Tor Vergata) e in ***Big Data*** (Roma, SAPIENZA).

The TaLTaC software package has been progressively developed to date in three major releases: T1 (2001), T2 (2005) and T3 (2016) T4 (2020).

TALTAC3 was developed in collaboration with the DISIM – Department of Information Engineering, Computer Science and Mathematics at Università degli Studi dell'Aquila

univaq.it Università degli Studi dell'Aquila

TALTAC IS widespread among the Text Analysis Community in Italy and abroad, including 200 entities between university departments, research institutions and other organizations.

A dicembre 2019, TaLTaC^2 è presente in Italia in 120 dipartimenti universitari, in 48 centri di ricerca e istituzioni di interesse nazionale, nonché in alcune università straniere, per un totale di oltre 1200 licenze rilasciate.

A call about the opportunity of using «remotely» the software via ENEA distributed computing facilities

received expressions of interest

from TaLTaC User Community:

40 departements /research institutes answers in 2 days

Area of interests: market, social and opinion research, food, health, political communication, sentiment analysis, etc.

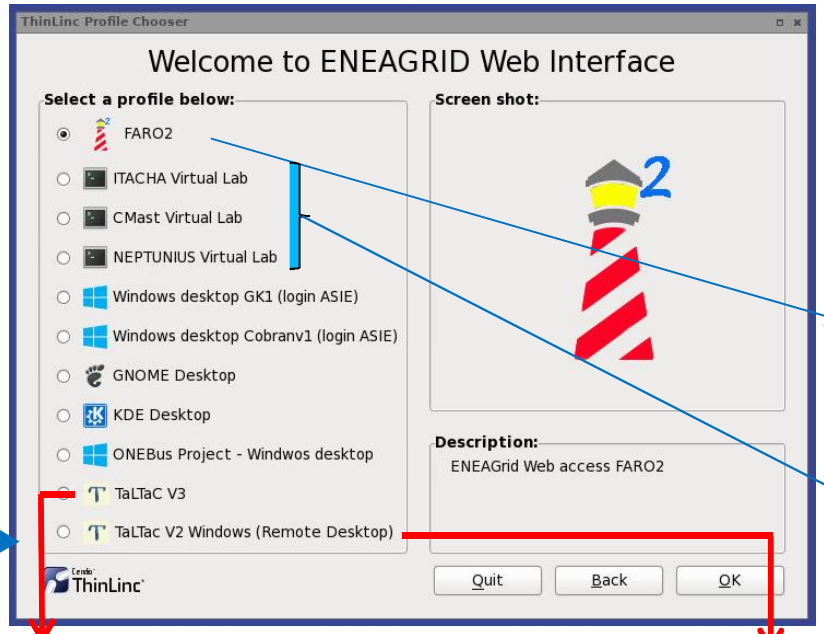
How To Access in ENEAGRID Infrastructure

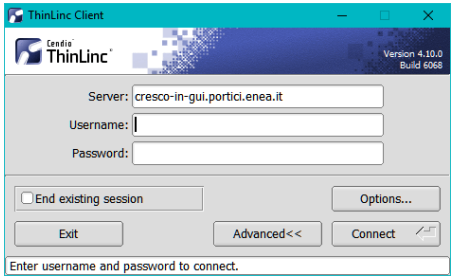
www.cresco.enea.it

Click on cresco-in-gui.portici.enea.it

ENEAGRID credentials:
ENEA.IT/Username
Password

login form integrated with ENEAGRID
https://cresco-in-gui.portici.enea.it





Before accessing at server cresco-in-gui.portici.enea.it download ThinLinc client at:
https://www.cendio.com/thinlinc/features
ThinLinc is a Cendio software making computing resources available to those who need it, when they need it, for a more efficient use of hardware; enabling users to move easily between machines , while still being presented with the same desktop . Since all the horsepower resides in the server hall, users no longer need their own expensive hardware to perform even the most *resource-intensive tasks*

FARO2 - Fast Access to Remote Objects
– Web Access Interface –
Remote *access* to: *CRESCO* software
ENEAGRID v-labs

Virtual Labs

Software

TaLTaC3 (Linux) on Cresco System

TaLTac software is available on CentOs Linux nodes.
Input and Output data can be accessed through the ENEAGRID filesystems and therefore easily uploaded and downloaded.

TaLTaC4
rilasciata a marzo 2020

TaLTaC2 (windows) on Remote Desktop Access

TaLTac software is available on «Windows Server 2012 R2» by remote desktop access to a ENEAGRID machine, reached by the ThincLinc general-purpose and intuitive interface with **ENEA ASIE credentials: ENEA.IT\name.surname and password.**

TaLTaC2 – versione 2.11.2
rilasciata a maggio 2019

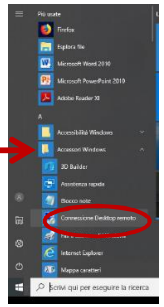
Friendly & Ubiquitous Access to TaLTaC (windows version)

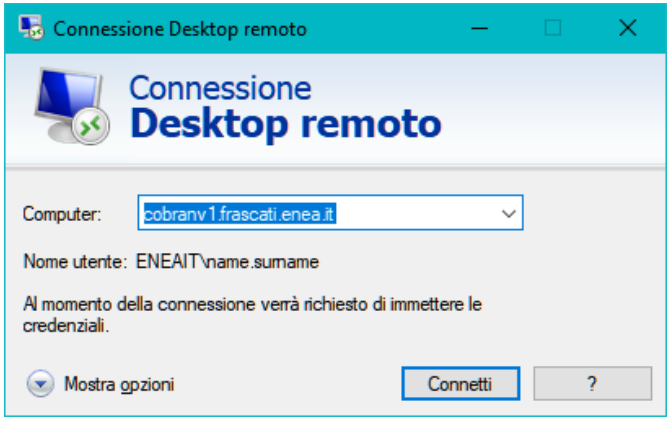
TaLTaC2 (windows) on Remote Desktop Access - ENEA or VPN network

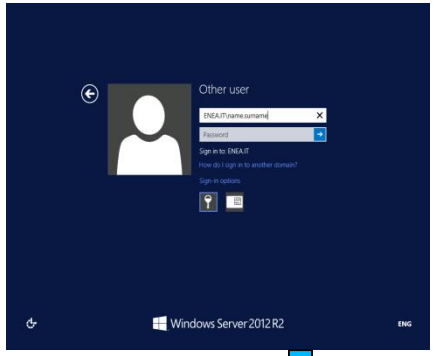
Menu START

-- >ACCESSORI WINDOWS

→Connessione Desktop remoto

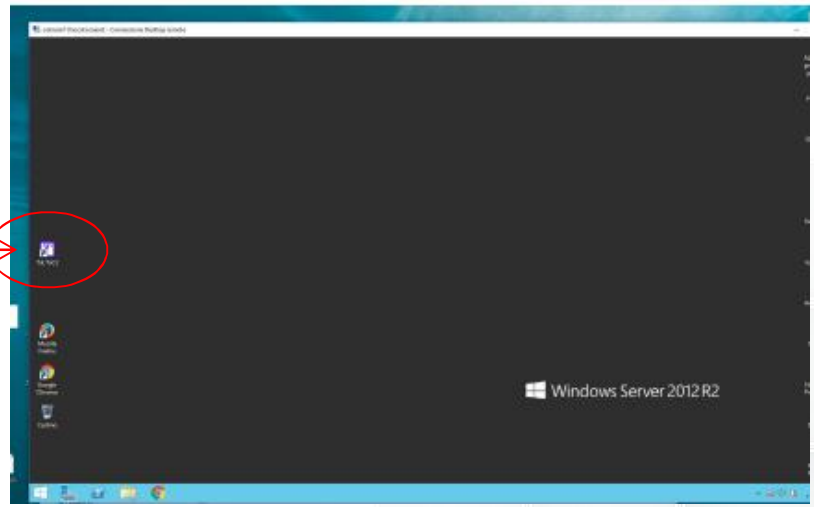






ENEA ASIE credentials:
ENEA.IT\name.surname
Password

TaLTac2



TaLTaC2 (windows) on Remote Desktop Access
TaLTac software is available on «Windows Server 2012 R2»
by remote desktop access to an ENEAGRID machine.
All users involved in the project activities can access the server.
AFS authentication is always required .
(AFS is a distributed network file system facilitating stored server file access between AFS client machines located in different areas)

Friendly -Ubiquitous Access to multicore TaLTaC (Linux version) in ENEAGRID

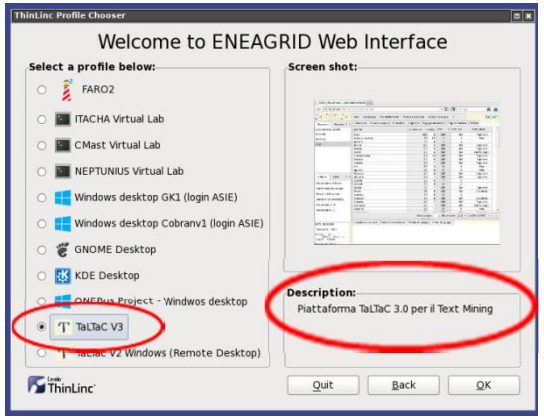
TaLTaC3 (Linux) on Cresco System http://www.cresco.enea.it/

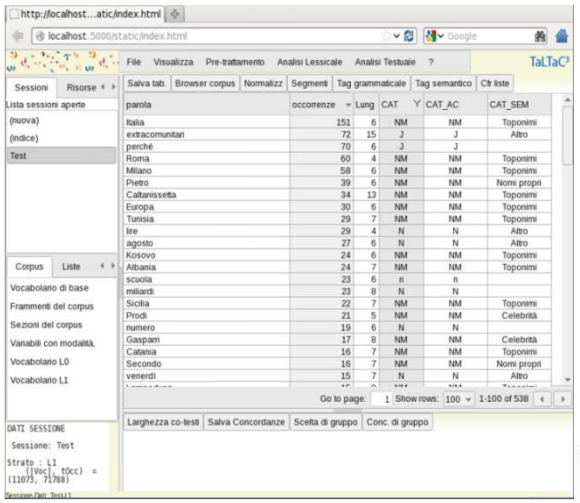
via ThinLinc https://www.cendio.com/thinlinc/features

Click on cresco-in-gui.portici.enea.it

login form integrated with ENEAGRID
https://cresco-in-gui.portici.enea.it:300/main/

welcome screen after logon through a browser





Before accessing www.cresco.enea.it
Download ThincLinc

Report 2019 activities and next step 2020

ENEA

TaLTaC2 ï Window version ï mono-core

TaLTac2 porting and support on windows multi-user machine

Development of the mono-user windows virtual machine on ENEAGRID scalable infrastructure.

TaLTaC3 / TaLTaC4 ï Linux version

Versione 3:

- ✓ Versione multipiattaforma (ambiente Python2, Redis la rendono utilizzabile su Linux e quindi CRESCO)
- ✓ Multithread (i processi di calcolo sono separati dagli altri processi di accesso ai dati e rendering)
- ➢ No - multicore
- ➢ No - binary (il sorgente viene decriptato a run-time)
- ➢ Funzionalità ridotte, in termini di analisi, rispetto alla versione 2

Versione 4:

- ✓ Versione multipiattaforma (Python3)
- ✓ Multicore
- ✓ Codice pre-compilato

Integrazione in ENEAGRID

- ➢ *Taltac3: disponibile in ENEAGRID tramite accesso ThinLinc, utilizzo solo per testing su nodi interattivi. Nel 2019 lo sviluppo è stato interrotto per il passaggio a Taltac4*
- ➢ *Taltac4: In fase di sviluppo da parte del team Taltac (rilascio previsto a Marzo 2020); Installazione ENEAGRID di un modello infrastrutturale: Profilo Thinlinc, browser per rendering della web-app GUI, sottomissione batch su nodi interattivi Cresco6*

Staff TaLTaC

Prove di installazione sul server portici, per ottimizzare l'installazione

Tolta password di attivazione, ora vi è un binario compilato, compatibile con la CentOS 6.4.

Installazione su CRESCO 6 in stand-by per ultimazione TaLTaC vers 4

https://www.taltac.com/

Aggiornamento cronologia di sviluppo di TaLTaC:

– **2011-2015 verso TaLTaC3:** studio sulle diverse tecnologie multipiattaforma Lucene, Python, PyPy, Jython, Electron.
– **2016-2017 Beta release TaLTaC 3.0:** Redis, monocore, PyPy, Python2.
TaLTaC 3.0: Un software multi-lessicale e uni-testuale ad architettura web
TaLTaC 3.0. A Multi-level Web Platform for Textual Big Data in the Social Sciences
– **2017-2018 TaLTaC 3.0:** Redis, multicore, PyPy, Python 3.5.
TaLTaC in ENEAGRID Infrastructure
– **2019-2020 Beta release TaLTaC 4.0:** filesystem, multicore, PyPy 7-Python 3.7.

TaLTaC2 - da Maggio 2019 è disponibile la versione 2.11.2

TaLTaC4 - da marzo 2020 sarà disponibile la versione 4

info@taltac.com

Versione di TaLTaC multi-piattaforma e multi-core

Giovedì 20 febbraio 2020, avrà luogo la presentazione della **nuova release** di **TaLTaC 4.0** a **Roma,**, in **Viale Regina Elena 332**, presso il **Dipartimento di Scienze Statistiche** della Sapienza Università di Roma, Edificio CU002, della ex Facoltà, (III piano, aula 3), nel corso di un workshop su "***Il Text Mining con TaLTaC al tempo dei Big Data***". Per il programma dell'evento *clicca qui*. Per partecipare occorre iscriversi, inviando una mail a sergio.bolasco@uniroma1.it .

From Results 2018 to Future Directions

Next Step

ENEAGRID offers to researchers computation and storage resources and services in a ubiquitous and remote way.
ENEAGRID integrates a cloud computing environment and exports:

a) remote software (i.e. TaLTaC);
b) remote storage facilities (with OpenAFS file system).
c) Virtual Labs: thematic areas accessible via web, where researchers can find set of software and documentation regarding specific research areas.
All these activities are based in ENEA e-Infrastructure.

Research collaborations with Experts Communities interested in Parallel Text Mining of Massive Volume of Text Data (PB & TB-sized corpora) in: e-Humanities, Heritage Science, Social Sciences, Big Data, Web Mining, Business Intelligence, Open Source Intelligence, Artificial Intelligence, Cybersecurity, etc

To fully exploit parallelism in Text Mining in GRID e-Infrastructure

and to **Create and activate new Knowledge from (cultural and socio-economic) Big Data**

Research in progress is related to:

- Knowledge extraction from Internet (web, social media, electronic papers, tweets or online news) case studies: Finance (Cryptocurrencies) ï Fashion (Luxury)
- Web Data Mining, Storage and Analysis of Open Data extracted from Open Sources
- Adaptation of software application (TaLTaC, SBS, etc.)

Taltac in ENEA Cloud

Integrating ICT inside Digital Cultural Research

ñTaLTaC in CLOUDòis a joint ENEA-TaLTaC project

for the set-up of an ICT portal on the ENEA distributed e-Infrastructure - The ñText Mining & Analyticsò Platform organized in specific Virtual Laboratory

hosting TaLTaC Software

Users will access TaLTaC software (Windows and Linux versions)

in a remote and ubiquitous way and the computational power (800 Teraflops) of ICT ENEA distributed resources, as a single supercomputer.

The aim of this joint ENEA-TaLTaC project is **integrating ICT inside Digital Cultural Research**.

to enable Digital Researcher in :

- Text Mining & TaLTaC software User Community
- e-Humanities
- Economic & Social Sciences

with **remote access to TaLTaC software through ENEAGRID Infrastructure,**

In this Project ENEA is both technological partner and TALTAC User.

The “Text Mining & Analytics” Platform

DTE-ICT HPC – ENEAGRID

offers a digital collaborative environment and an integrated platform of tools and resources for Research Collaborations, Sharing Knowledge and digital resources and Storing textual data.

The “Text Mining & Analytics” Platform

Text Mining & Analysis in GRID environment will provide :

i) Collection of Open Data from Social and Web (crawling tasks);

ii) Data and computing capacity;

iii) Data storage facilities;

iv) Parallel Text/Data Mining & Analysis (data & task parallelism).

v) on-line and ubiquitous access «always and anywhere ON», to software and computational resources in ENEAGRID, regardless of the location of the specific machine, of the employed hardware/software platform and regardless of the corpus size (from Gigabytes to Terabytes),

vi) Virtual Research Environments (virtual labs) & Collaboration tools (network management, video conferencing and voip services, cloud computing, ecc.);

iv) a simple user interface for users (web access) of software of linguistic Analysis for large corpora on ENEAGRID :

(i.e. TALTAC2 windows version and of TALTAC3 linux version)

"Data-driven approaches can complement the traditional method
in detecting trends of continuity and change in large-scale textual corpora.

Data Science enables cross disciplinary research
exploring the interplay between Social Science, Humanities, and large-scale data-driven AI."

Computational approaches *can establish meaningful relationships*
between a given signal **in large-scale textual corpora** *and verifiable historical moments,*

but the ***understanding*** *of the* ***implications*** *of these findings for people cannot be automated,*
and will always be ***the realm of the humanities and social sciences****,*
and never that of machines.

Nello Cristofanini – Professor of AI University of Bristol

Artificial Intelligence, Machine Learning, Media Content Analysis, Big Data, Epistemological and Ethical Implications of Data-Driven Science and Society

"Leggere 180 milioni di parole" da Tuttoscienze – La Stampa 16 maggio 2018 -large-scale analysis of historical newspapers, modern news, social media content and images

Co-Authors

DTE-ICT Authors:

Silvio Migliori, Andrea Quintiliani, Daniela Alderuccio, Fiorenzo Ambrosino, Maria Luisa Mongelli, Samuele Pierattini, Giovanni Ponti

TaLTaC Authors:

Sergio Bolasco – uniroma1

Francesco Baiocchi - ISTAT

Giovanni De Gasperis - univaq

Thanks for your attention!

References

1. D. Alderuccio, S. Migliori, A. Quintiliani, F. Ambrosino, A. Colavincenzo, M. Mongelli, S. Pierattini , G. Ponti, S. Bolasco, F. Baiocchi, G. De Gasperis. TaLTaC in ENEAGRID Infrastructure in *Proceedings JADT 2018 (Journées Internationales d'Analyses statistique des Données Textuelles) 14th Int. Conference on Statistical Analysis of Textual Data)*- pp. 501-508 - Roma 15.6.2018 c/o CNR, (2018)
2. S. Bolasco, F. Baiocchi, A. Canzonetti, G. De Gasperis (2016). TaLTaC3.0, un software multi-lessicale e uni-testuale ad architettura web, in D. Mayaffre, C. Poudat, L. Vanni, V. Magri, P. Follette (eds.), *Proceedings of JADT 2016,* CNRS University Nice Sophia Antipolis, Volume I, pp. 225-235. (2016).
3. S. Bolasco, G. De Gasperis (2017). TaLTaC 3.0 A Web Multilevel Platform for Textual Big Data in the Social Sciencesòin C. Lauro, E. Amaturo, M.G. Grassia, B. Aragona, M. Marino. (eds.) *Data Science and Social Research - Epistemology, Methods, Technology and Applications (series: Studies in Classification, Data Analysis, and Knowledge Organization)* Springer Publ., pp. 97-103. (2017).
4. G. Ponti et al. (2014). The Role of Medium Size Facilities in the HPC Ecosystem: The Case of the New CRESCO4 Cluster Integrated in the ENEAGRID Infrastructure In: *Proceedings of the International Conference on High Performance Computing and Simulation,* HPCS (2014), ISBN: 978-1-4799-5160-4. (2014).
5. G. Ponti, D. Alderuccio, G. Mencuccini, A. Rocchi, S. Migliori, G. Bracco, P. Negri Scafa (2017). Data Mining Tools and GRID Infrastructure for Text Analysis in *ñPrivate and State in the Ancient Near EastòProceedings of the 58th Rencontre Assyriologique Internationale,* Leiden 16-20 July 2012, edited by R. De Boer and J.G. Dercksen, Eisensbrauns Inc. - LCCN 2017032823 (print) | LCCN 2017034599 (ebook) | ISBN 9781575067858 (ePDF) | ISBN 9781575067841. (2017).
6. ENEAGRID http://www.ict.enea.it/it/hpc -
7. Laboratori Virtuali http://www.ict.enea.it/it/laboratori-virtualixxx/virtual-labs
8. TIGRIS Virtual Lab http://www.afs.enea.it/project/tigris/indexOpen.php
9. TaLTaC: www.taltac.it

Report CRESCO Results 2018

PARALLEL TEXT MINING OF LARGE CORPORA IN GRID ENVIRONMENT - TALTAC IN ENEAGRID INFRASTRUCTURE

Silvio Migliori[1], Daniela Alderuccio[1], Fiorenzo Ambrosino[1], Antonio Colavincenzo[1], Marialuisa Mongelli[1], Samuele Pierattini[1], Giovanni Ponti[1], Andrea Quintiliani[1] , Sergio Bolasco[2], Francesco Baiocchi[3], Giovanni De Gasperis[4]

[1]*ENEA – DTE-ICT Division – Roma, Italy*

[2]*Sapienza Università di Roma, Italy*

[3] *Staff TaLTac, Roma, Italy*

[4]*Dip. DISIM Università dell'Aquila, Italy*

In this work we presented an ENEAGRID approach toward Parallel Text Mining of Big Data in massive cultural Textual Corpora.

ENEAGRID provides computation, storage resources and services for Text Mining Communities in a ubiquitous and remote way.

ENEA distributed e-Infrastructure and cloud services enable the management of research process in Economic-Social Sciences, Digital Humanities and Cyber-security, providing technology solutions and tools to academic departments and research institutes.

With these research activity ***ENEA opens to collaborations with other Research Communities interested in Parallel Text Mining of Massive Volume of Text Data (PB &TB-sized corpora) in e-Humanities, Heritage Science, Social Sciences, Big Data, Web Mining, Business Intelligence, Open Source Intelligence, Artificial Intelligence, etc.***

Keywords: Text Mining Software, Cloud Computing, Digital-Humanities, Socio-Economic Sciences, Big Data, Artificial Intelligence, Cybersecurity, Business Intelligence, AltaGamma.