

How RAI's Hyper Media News aggregation system keeps staff on top of the news

13th Libre Software Meeting
***Media, Radio, Television and Professional
Graphics***

Geneva - Switzerland, 10th July 2012
Maurizio Montagnuolo

◆ Company presentation

◆ Motivations

◆ Foundations

- Audiovisual content processing for TV streams **analysis**
- Natural Language Processing (NLP) for text **analysis**
- Full text **search** and **retrieval**

◆ Use case implementation

- The RAI Automatic Newscast Transcription System (ANTS)
- The RAI Hyper Media News aggregator
- The RAI Interactive Newsbook

◆ Conclusions and future outlook

The “RAI broadcasts” - a short history

- **National Radio broadcasts since early 30’ (EIAR)**

- 1950: Radio3 was born
- 2012: 10 Radio channels

- **National TV broadcasts since 1954**

- 1961: Rai2 was born
- 1977: Color introduced
- 1979: Rai3 was born with regional transmissions
- 1990: First analog satellite transmissions
- 2005: Youtube official channel
- 2008: DTT introduced (8 channels)
- 2012: 14 SD + 1 HD DTT channels





TV

- About 450,000 hrs
- 320,000 hrs of programming
 - 145,000 hrs: News and Sport
 - 175,000 hrs: Programmes (Entertainment, folk and classic music, theatre, ...)
- 130,000 hrs of Fiction
 - 50,000 hrs: Commercial Films
 - 80,000 hrs: Fiction (TV Series, TV films, Soap operas, ...)

RADIO

- About 1,000,000 hrs recorded on a wide variety of media

PAPER LIBRARY

- 80,000 scripts in Rome
- 15,000 scripts in Firenze
- Evaluation of further RAI archives planned in short time

IMAGE LIBRARY

- 360,000 Photos RAI
- 950,000 Photos ex-ERI, currently RAI TRADE

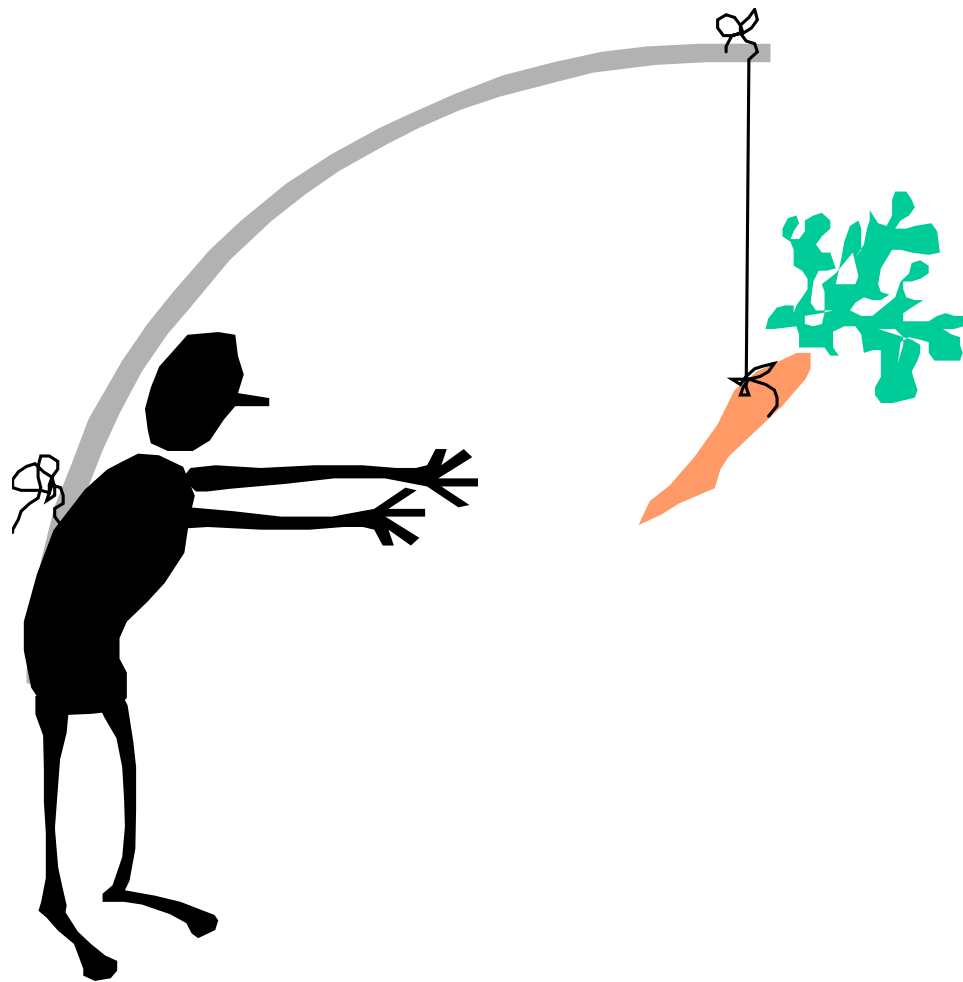
- ◆ **The Centre for Research and Technological Innovation (CRIT)** is responsible for defining, promoting and developing all aspects of research and innovation in the television industry

- <http://www.crit.rai.it/EN/home.htm>

- ◆ **The Centre is active in many EU projects, and collaborates with universities and industries for supporting Master and PhD thesis, as well as developing new standards and services**



Motivations



Why content analysis tools in the media industry?

- ◆ **Digital switch over introduces more channels**
 - More content items produced/published
- ◆ **Cross media production (web, TV,...)**
 - Reuse material in many different ways
- ◆ **Improvements in infrastructure (IT)**
 - Better content accessibility
- ◆ **Recovery of Cultural Heritage**
 - Archive digitisation and annotation
- ◆ **Budget limitations**
 - Archivist/documentalist staff not increasing



CUMULATIVE EFFECT:
a lot more digital items to be managed by the
same staff, and in a quicker way





- ◆ **The world of media is moving fast**
 - More challenges
 - New requirements
- ◆ **Huge data amounts**
 - Up to 10K hours of material for a typical regional news archive
 - Even if a lot is non-textual, we deal with it by
 - ✓ Tags, annotations, closed captioning, speech transcripts, ...
- ◆ **Open source tools for audiovisual content analysis, indexing and retrieval can be a solution**

Speed up the search and retrieval process

Automatic speech understanding

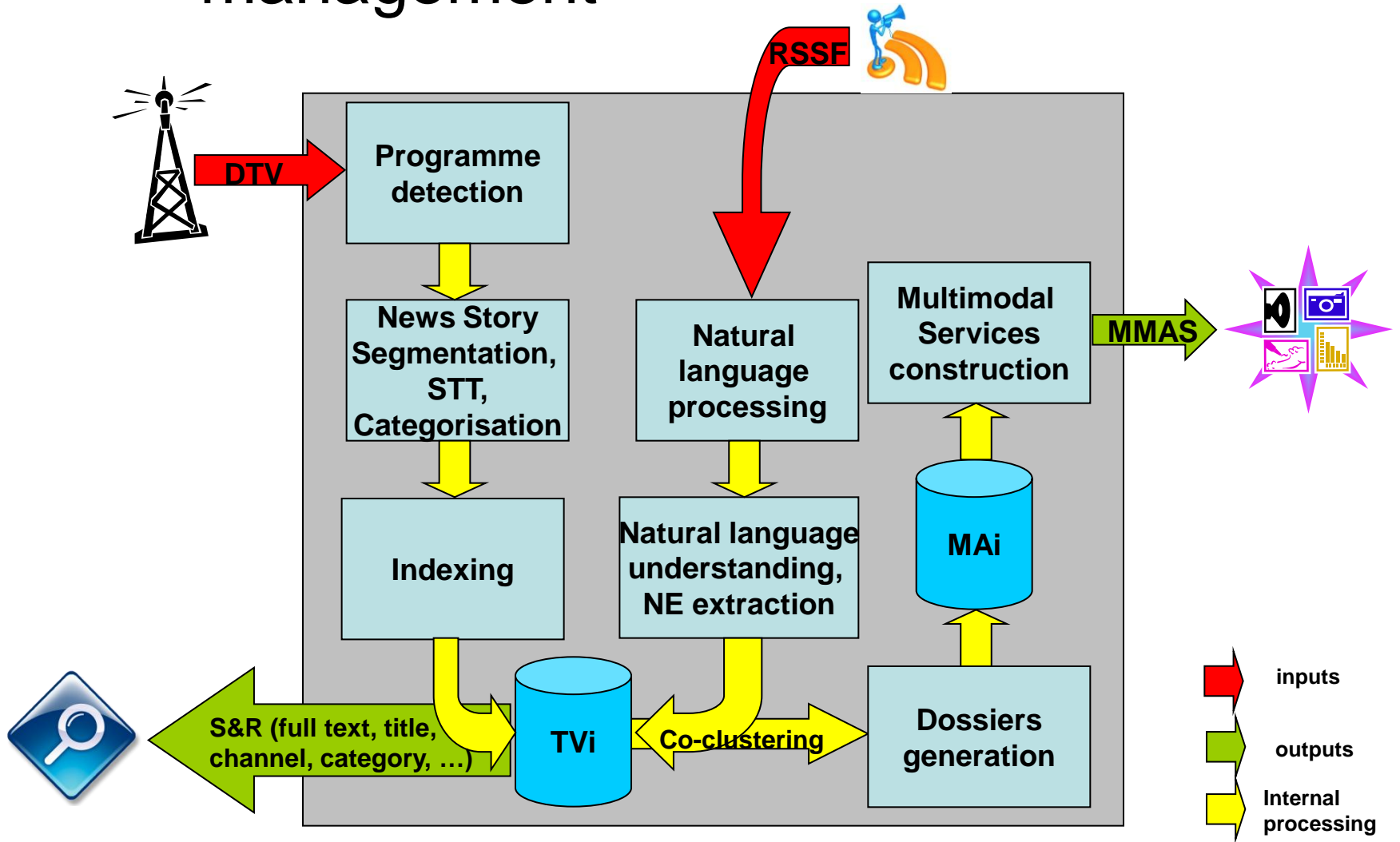
Automatic translation for multi-language news aggregation

Characters extraction and text summarisation

Information extraction and knowledge acquisition



Architecture for multi-modal news management



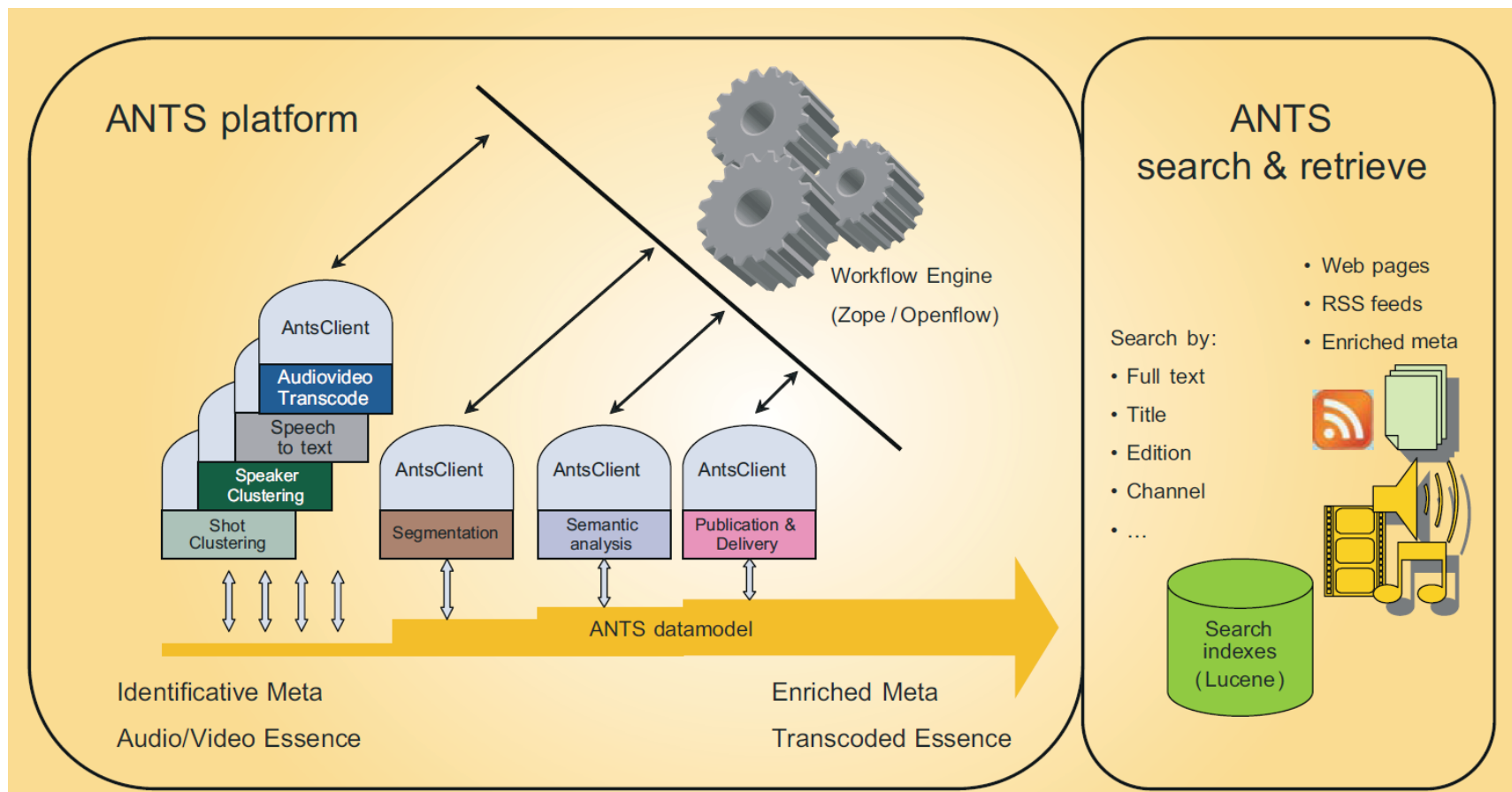
Audiovisual content processing for TV streams analysis

The RAI ANTS (Automatic Newscast Transcription System) platform provides a set of tools for automated news segmentation, classification, indexing and retrieval

Composed of three main modules

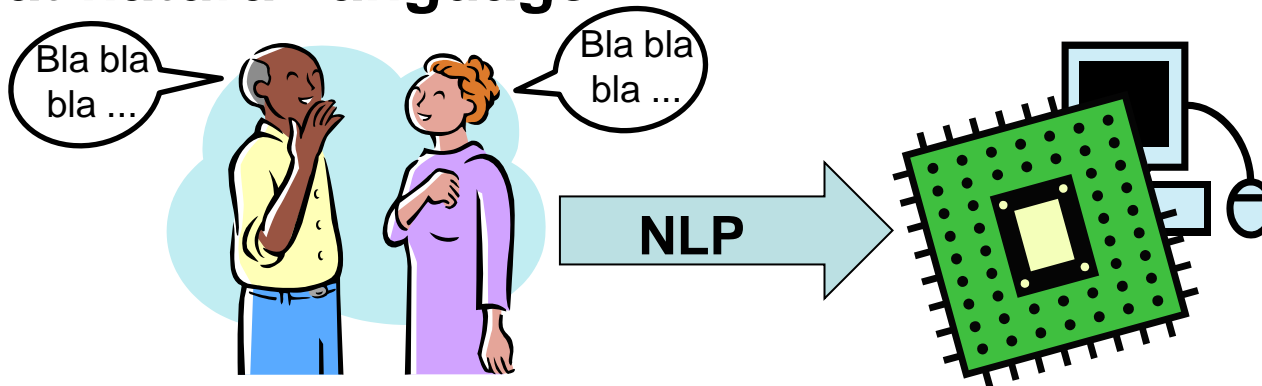
- Programme detection detects the start/end positions of newscasts from the acquired DTV streams
- News story segmentation performs segmentation of the acquired programmes into elementary news stories
- Speech to text analysis extracts text and semantics (i.e., categories, named entities) from the speech content of each story

RAI ANTS architecture



Natural Language Processing (NLP) for text analysis

- ◆ **Natural language** refers to the human ability of understanding ordered sequences of written or spoken words (i.e. phrases)
 - Who? What? Where? When? Why?
- ◆ Language **processing** is the set of algorithms and tools that make machines able to understand and treat natural language





◆ Computers use **number sequences** to communicate

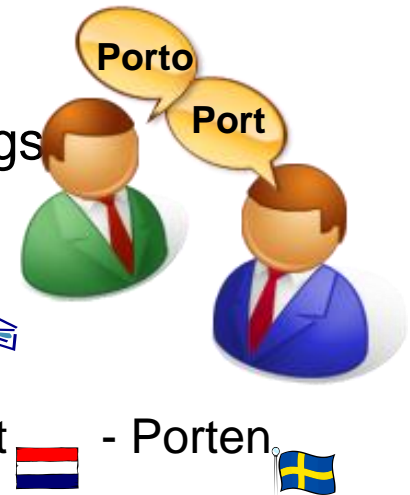
- Numbers are simple
- Numbers are easy to understand
- Numbers do not lie



◆ Humans use **word sequences** to communicate

- Words can be unknown
 - ✓ E.g. *Supercalifragilisticexpialidocious* ?????
- Words can have multiple grammars and meanings

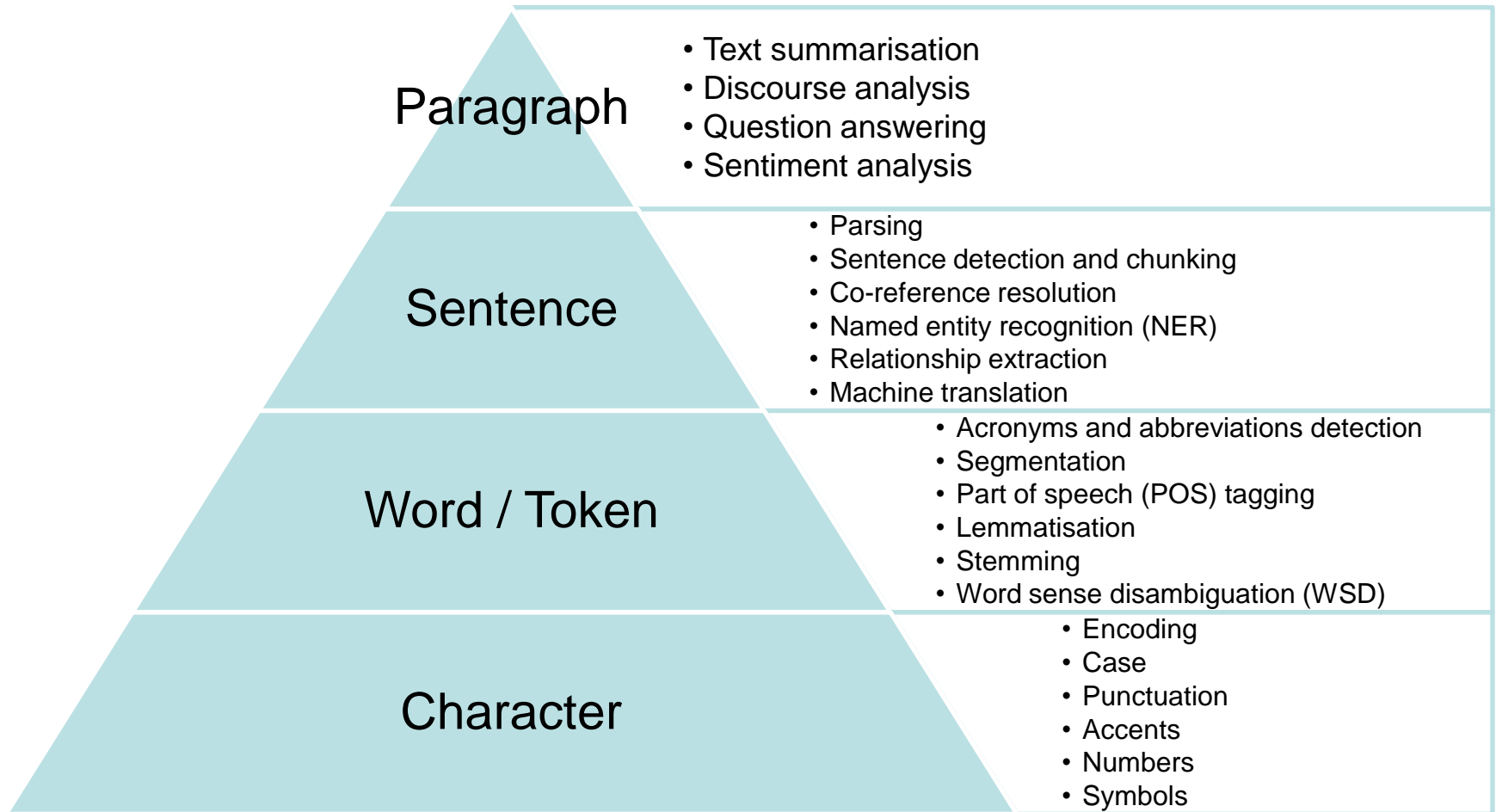
✓ To port  - Port wine  - Usb port



- Words are multilingual

✓ Port  - Porto  - Puerto  - Luka  - Poort  - Porten 

NLP pyramid tasks



- ◆ **Different tools and libraries available**
- ◆ **Different implementations and programming platforms**
 - C, C++, C#, Java, Python, Perl, Ruby, ...
- ◆ **Different usage licenses**
 - GPL, LGPL, MIT, Apache, ...
- ◆ **Further detail on Wikipedia**
 - List of natural language processing toolkits
 - ✓ http://en.wikipedia.org/wiki/Natural_language_processing_toolkits

- ◆ **Machine learning toolkit released under the Apache License 2.0**
- ◆ **Pre-built models for several languages**
 - Danish, German, English, Spanish, Dutch, Portuguese, Swedish
- ◆ **Set of training tools for building further language models**

Sentence detector

Tokenization

Name
Finder

Docu
ment
Categorizer

Part of Speech
Tagger

Chu
nker

Parser

Corefe
rence
Resol
ution

- ◆ **Sentence detection does not mean only splitting at punctuation marks**
 - E.g. - Ms. - Mrs. - www.raai.tv - 1,000,000
- ◆ **Sentence tokenisation needs to**
 - Separate possessive endings or abbreviated forms from preceding words, e.g. Maurizio 's, can 't,...
- ◆ **Separate punctuation marks, quotations, brackets (...) from words**
 - Maurizio lives in Turin (Italy). → Maurizio lives in Turin (Italy) .
- ◆ **A word might have multiple pos tags depending on its context.**
- ◆ **Named entities might be of multiple types**

- ◆ Full text enterprise search server based on Lucene
- ◆ XML/HTTP, JSON Interfaces
- ◆ Distributed as Web application (war)
 - Platform independent, HTTP controlling
- ◆ Web administration interface
 - Access, management, testing
- ◆ Easy configuration via XML files
 - Definition of indexes, data types, operations, etc.
- ◆ Index replication and distribution
- ◆ Search results caching

◆ Software

- Operating system: Windows, Linux, Mac, ...
- Java Development Kit (JDK) v1.5 or greater
- Apache Ant (not required for standard installation)
- Java EE Application Server
 - ✓ Jetty, Apache Tomcat, JBoss,...
- Java Database Connectivity (JDBC) for database interaction

◆ Hardware requirements depends on the size and complexity of the data

- RAM affects indexing, optimisation and searching performance
- The size of the documents (number of documents, fields per document, fields size,...) affects storage requirements

◆ Testing performance

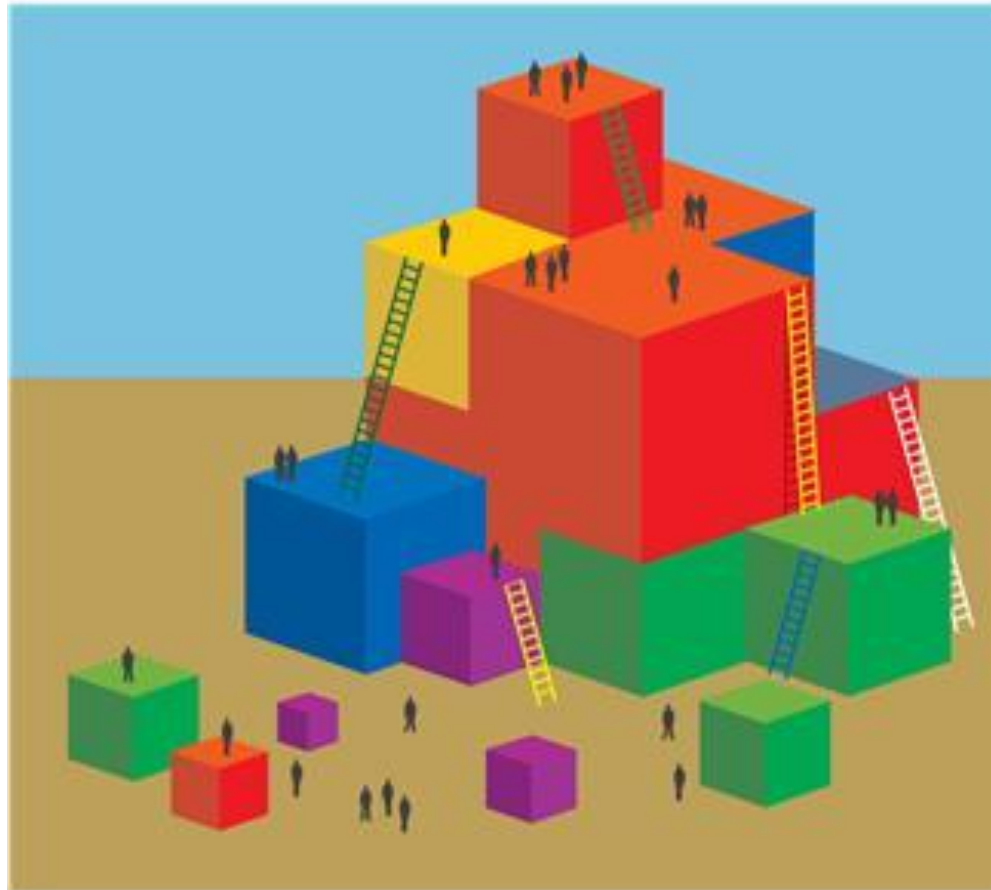
- SolrMeter

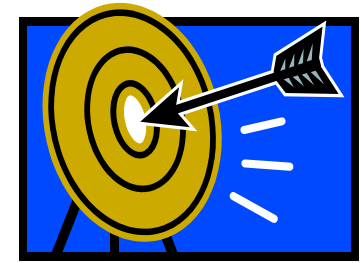
- ◆ ***solr.xml***: defines the number of cores (indexes) available
 - <http://wiki.apache.org/solr/CoreAdmin>
- ◆ ***schema.xml***: defines all of the details about which fields your documents can contain, and how those fields should be dealt with when adding documents to the index, or when querying those fields.
 - <http://wiki.apache.org/solr/SchemaXml>
- ◆ ***solrconfig.xml***: is the file where to put most of the parameters for configuring the Solr cores (query handlers, highlighting, faceting, etc)
 - <http://wiki.apache.org/solr/SolrConfigXml>

Solr - Data Import handler (DIH)

- ◆ **The Data Import Handler is the component for importing data from external sources (e.g. XML archives, databases,..)**
- ◆ **Read and Index data from xml/(http/file) based on configuration**
- ◆ **Read data residing in relational databases**
 - Build Solr documents by aggregating data from multiple columns and tables according to configuration
 - Update Solr according to DB updates
 - ✓ Detect inserts/update deltas (changes) and do delta imports
- ◆ **Make it possible to plugin any kind of data source (ftp, scp, ...) and any other format (JSON, CSV, ...)**

Use case implementation





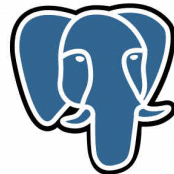
- ◆ Define and develop methods and systems for automated content **analysis, documentation, indexing** in the media domain
- ◆ Example target: news
 - Explosive growth of available informative assets
 - ✓ Professional, e.g. Newspapers, press agencies, Radio & TV
 - ✓ Amateur, e.g. UGC, social networks, personal blogs
 - Heterogeneity of sources, e.g. the Internet, radio, TV, print media, legacy archives, ...



ANTS, HMN & Interactive Newsbook main components



PostgreSQL





ANTS, HMN & Interactive Newsbook - core features

◆ Fully automated multimodal content-analysis tools for data extraction and mining

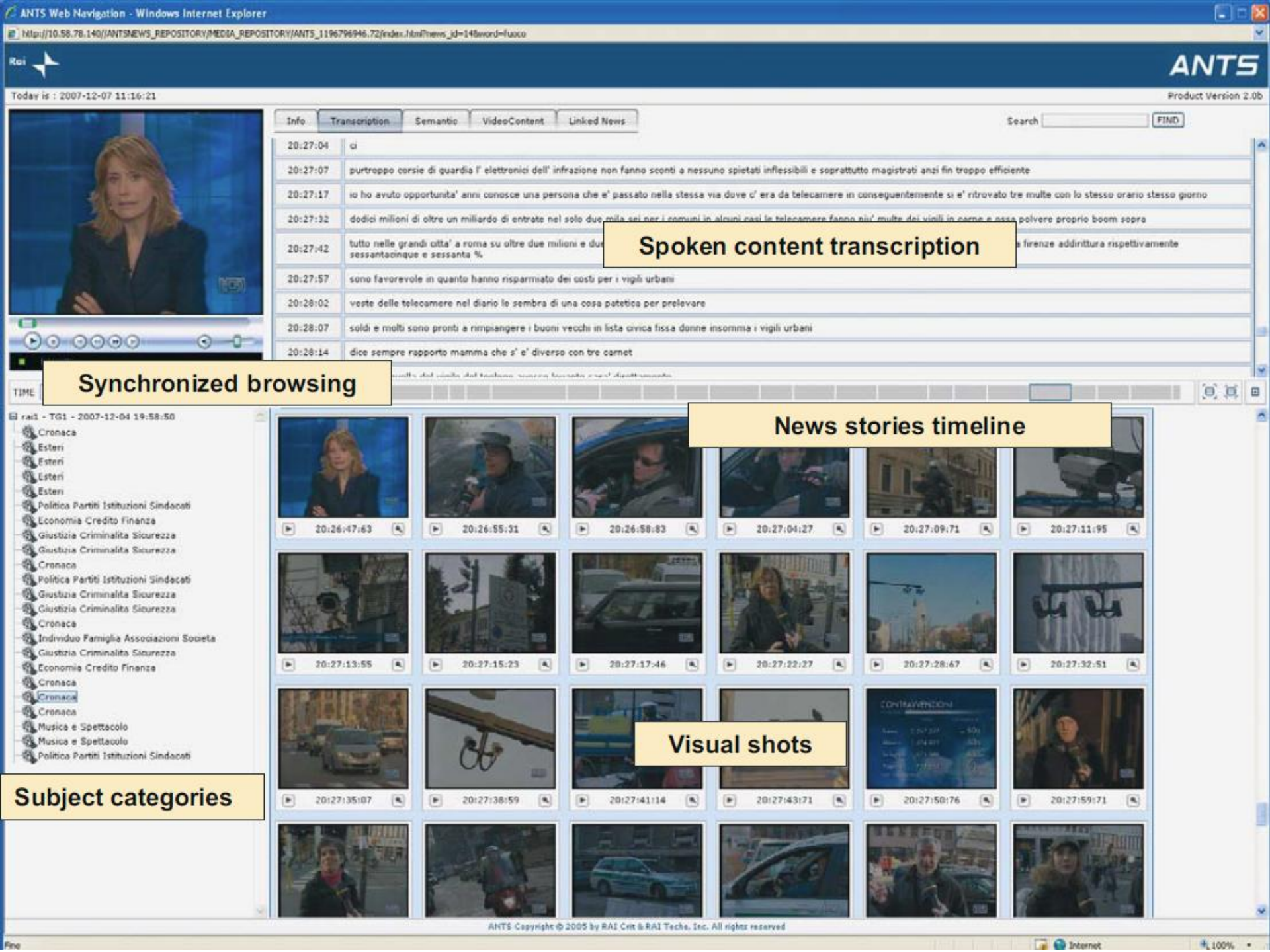
- TV news programmes detection, segmentation and indexing
- RSS feeds analysis, hierarchical linking and indexing

◆ Aggregation of multimodal news items by affinity measure

- A novel (generalised) measure for assessing affinity
- Aggregations are contextualised within automatically extracted information
 - ✓ Entities (i.e. persons, places and organisations)
 - ✓ Temporal span
 - ✓ Categorical topics
 - ✓ Social networks popularity and audience scores.

◆ Integration with external resources

- Public Internet search engines, Legacy digital libraries
- Integration of otherwise disconnected resources



Search

Info Transcription Semantic VideoContent Linked News

20:27:04	ci
20:27:07	purtroppo corsie di guardia l' elettronici dell' infrazione non fanno sconti a nessuno spietati inflessibili e soprattutto magistrati anzi fin troppo efficiente
20:27:17	io ho avuto opportunita' anni conosce una persona che e' passato nella stessa via dove c' era da telecamere in conseguentemente si e' ritrovato tre multe con lo stesso orario stesso giorno
20:27:32	dodici milioni di oltre un miliardo di entrate nel solo due mila sei per i comuni in alcuni casi le telecamere fanno più multe dei vigili in zona e essa polvere proprio boom sopra
20:27:42	tutto nelle grandi citta' a roma su oltre due milioni e due sessantacinque e sessanta %
20:27:57	sono favorevole in quanto hanno risparmiato dei costi per i vigili urbani
20:28:02	veste delle telecamere nel diario le sembra di una cosa patetica per prelevare
20:28:07	soldi e molti sono pronti a rimpiangere i buoni vecchi in lista civica fissa donne insomma i vigili urbani
20:28:14	dice sempre rapporto mamma che s' e' diverso con tre carnet

Spoken content transcription

Synchronized browsing

- rai1 - TG1 - 2007-12-04 19:58:50
- Cronaca
- Esteri
- Esteri
- Esteri
- Esteri
- Politica Partiti Istituzioni Sindacati
- Economia Credito Finanza
- Giustizia Criminalita Sicurezza
- Giustizia Criminalita Sicurezza
- Cronaca
- Politica Partiti Istituzioni Sindacati
- Giustizia Criminalita Sicurezza
- Giustizia Criminalita Sicurezza
- Cronaca
- Individuo Famiglia Associazioni Società
- Giustizia Criminalita Sicurezza
- Economia Credito Finanza
- Cronaca
- Cronaca
- Musica e Spettacolo
- Musica e Spettacolo
- Politica Partiti Istituzioni Sindacati

Subject categories

News stories timeline

Grid of video thumbnails with timestamps:

- 20:26:47:63
- 20:26:55:31
- 20:26:58:83
- 20:27:04:27
- 20:27:09:71
- 20:27:11:95
- 20:27:13:55
- 20:27:15:23
- 20:27:17:46
- 20:27:22:27
- 20:27:28:67
- 20:27:32:51
- 20:27:35:07
- 20:27:38:59
- 20:27:41:14
- 20:27:43:71
- 20:27:50:76
- 20:27:59:71

Visual shots

News Archive

May 2012

Mo	Tu	We	Th	Fr	Sa	Su
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2
3						

Italia alle urne per le amministrative

Last News Update : 2012-05-08 03:11:45

Da domani mattina alle 8 fino a lunedì' alle 15 si voterà per le amministrative 2012, con una pausa notturna fra le 22 e le 7. Saranno rinnovati, fra gli ... [Leggi on-line su RAI.it](#)

593 521 bersani boris bretagna casini chigi comuni elezioni francia friuli genova giulia governo greca olandese imu interno italia johnson lega londra mario maroni monti palermo pasok pd pdf per rai roma sarkozy sicilia venezia



Riconoscimenti

Giovedì 25 Novembre 2010

Premio Nostalgia di Futuro

In ricordo di Giovanni Giovannini, viene assegnato alla Rai per l'alto contributo di innovazione tecnologica nel campo del trattamento automatico dell'informazione multimediale (progetti Hyper Media News e ANTS).

Hanno ritirato il Premio Giorgio Dimino e Alberto Messina, della Direzione Strategie Tecnologiche - CRIT di Torino, e Roberto Rossetto, della Direzione Tecche Rai.

[Leggi la notizia on-line](#)

Mercoledì 09 Giugno 2010

Premio ICMT

Per l'approccio innovativo con cui viene affrontata la convergenza tra differenti tipologie di contenuto e per il contributo ad un settore di rilevanza cruciale qual è quello dell'informazione.

Il premio è stato consegnato dal presidente della Repubblica **Giorgio Napolitano**, nell'ambito della Giornata dell'Innovazione

[Leggi la notizia on-line](#)

News Categories

Categories

- ▶ Ambiente e Natura
- ▶ Consumi Servizi
- ▶ Cronaca
- ▶ Economia e Finanza
- ▶ Esteri
- ▶ Giustizia e Sicurezza
- ▶ Individuo e Famiglia
- ▶ Industria e Imprese
- ▶ Lavoro Previdenza
- ▶ Musica e Spettacolo
- ▶ Politica
- ▶ Sanità Salute
- ▶ Scienze Tecnologie
- ▶ Sport
- ▶ Trasporti



Andreotti sta meglio e vorrebbe tornare a casa

Last News Update : 2012-05-05 17:18:18

Il senatore a vita: "Io dato per morto? Porta bene, allunga la vita". Era stato ricoverato giovedì al Gemelli per una grave insufficienza respiratoria. Le sue condizioni sono buone, ma ... [Leggi on-line su RAI.it](#)

70 28 andreotti gemelli giulio roma wikipedia

Autobus sbanda sull'A13, 5 morti

Last News Update : 2012-05-07 14:53:10

Ha sbandato a destra improvvisamente, come un animale impazzito, e ha chiuso la sua folle corsa adagiato su un fianco, ai margini di una canaletta e con un conto di ... [Leggi on-line su RAI.it](#)

80 68 aprilia arma associazione bologna carabinieri comune jesolo lazo legnaro padova pulman roma terme



Rss Web News

Adnkronos/IGN

Affaritaliani.it

Agenzia di Stampa Asca

Agenzia di Stampa Italtpress

AGI - Agenzia Giornalistica Italia

AGO Agenzia Giornalistica Online

America Oggi

Last TV News

TG2 13:00 - 18/01



VOTO SCHLESWIG, MERKEL

Last News Update : 2012-05-07 19:24:57

Rai VOTO SCHLESWIG, MERKEL TIENE MA ARRETRA La Cdu di Angela Merkel è in leggero vantaggio sulla Spd nelle elezioni nel piccolo Land dello Schleswig-Holstein, ma la coalizione con la Fdp che era al governo esce sconfitta. Il partito della cancelliera, secondo i dati diffusi dalla tv statale Ard, otterrebbe il 30,5%, mentre la formazione socialdemocratica guadagna 4 punti e si attesta al 29,5%. I liberaldemocratici della Fdp si fermano all'8,5%, mentre i verdi raggiungono uno storico 13,6%. La Cdu dovrà cercare altri alleati per continuare a governare. Ancora un successo per i Pirati, all' 8%, mentre la Linke è ferma al 2,5%.

News Panel

Total Matching News : 118

PRESIDENTE HOLLANDE

RAI.it - 2012-05-07 15:31:10

Rai PRESIDENTE HOLLANDE A ELISEO IL 15 MAGGIO Il neo presidente francese François Hollande si insedierà all'Eliseo il 15 maggio. Domani, festa nazionale, ci sarà il primo incontro con il presidente uscente Sarkozy, nel corso della cerimonia per ricordare la...[\(continua\)](#)

PUTIN PRESIDENTE, CERIMONIA AL CREMLINO

RAI.it - 2012-05-07 15:29:11

Rai PUTIN PRESIDENTE, CERIMONIA AL CREMLINO "Oggi entriamo in una nuova tappa dello sviluppo nazionale e gli anni a venire saranno determinanti per il destino della Russia". Lo ha detto il presidente Putin, che si è insediato al Cremlino. Questo...[\(continua\)](#)

PRESIDENTE HOLLANDE

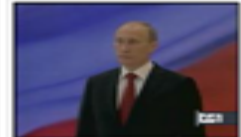
RAI.it - 2012-05-07 15:24:31

Rai PRESIDENTE HOLLANDE A ELISEO IL 15 MAGGIO Il neo presidente francese François Hollande si insedierà all'Eliseo il 15 maggio. Domani, festa nazionale, ci sarà il primo incontro con il presidente uscente Sarkozy, nel corso della cerimonia per ricordare la...[\(continua\)](#)

Putin ha giurato al Cremlino. Fermati 120 oppositori

RAI.it - 2012-05-07 15:20:58

E' al terzo mandato da presidente. Berlusconi tra i 3000 ospiti alla cerimonia. Mosca blindata: irruzione della polizia nel caffè ritrovo dei dissidenti. ...[\(continua\)](#)



TG1 - Servizio Num: 16 - Rai1 - 2012-05-07 13:44:35

e allora vediamo come e' cominciata la terza volta di putin al cremlino abbiamo sentito oggi il giuramento alla cerimonia sorellina mosca con tremila invitati tra loro anche berlusconi corrispondente giovanni masotti e' blindata gravi incidenti della ... [\(continua\)](#)



TG1 - Servizio Num: 14 - Rai1 - 2012-05-07 13:42:25

quanto ai risultati delle elezioni europee e in una nota diffusa ieri a tarda sera preferite consiglio ha affermato bisogna rispettare il rigore ma sono urgenti misure per la crescita contatti telefonici poi ieri tra presente il confine a ... [\(continua\)](#)



TG1 - Servizio Num: 9 - Rai1 - 2012-05-07 13:39:27

le elezioni amministrative in italia i saggi dei quasi mille comuni chiamati a scegliere sindaco e consiglio comunale chiuderanno alle quindici subito dopo cominceranno le operazioni di scrutinio in caso di ballottaggio si voterà domenica ventì lunedì ventuno maggio ... [\(continua\)](#)



TG1 - Servizio Num: 7 - Rai1 - 2012-05-07 13:36:44

io ora la germania dunque la cancelliera merkel la nuovi interlocutori a cominciare dal presidente francese parlando di noi e noi e' parlato ieri sera alle ore in diretta subito dal corrispondente marco vanelli a caccia ne'

Rai Televideo

100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120

TELEVIDEO 14:07 Prog. 13 - 14:08

DAL MONDO

FRANCIA: HOLLANDE E' IL NUOVO PRESIDENTE

MOSCA: PUTIN SI INSEDIA AL CREMLINO

GERM. SI VOTA PER NUOVO PARLAMENTO

VOTO SCHLESWIG: MERKEL A ELISEO

GERM. TARGO E MERKEL AL BALLETTAGGIO

MOSCA: PUTIN SI INSEDIA AL CREMLINO

Mega Channel Online

Ultim'ora

DAL MONDO

ITALIA

FRANCIA

MOSCA

GERMANIA

SCHLESWIG

SERBIA

SILVIO

UE

VLADIMIR

FRANCIA

MOSCA

GERMANIA

SCHLESWIG

SERBIA

SILVIO

UE

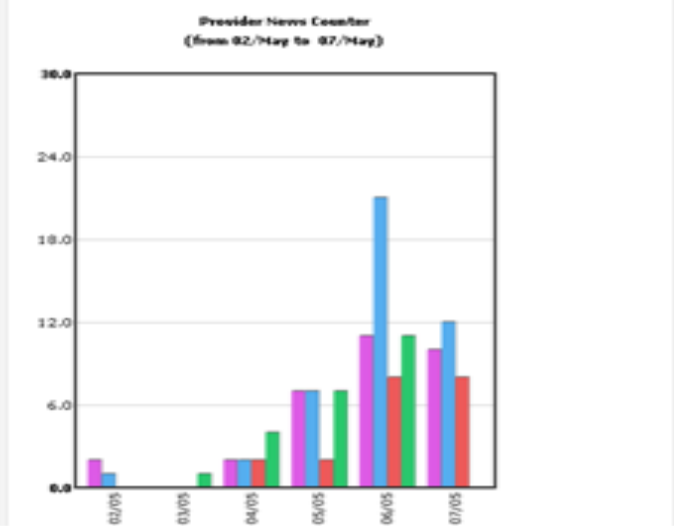
VLADIMIR

Leggi la notizia direttamente su RAI.it

Dossiers

- Francia. Domani voto decisivo
 - Francia, Francois Hollande al 51,9%
 - Francia: Berlino, lavoreremo insieme
 - Elezioni in Serbia, affluenza al 53,6%
- [Visualizza gli altri dossiers \(30\)](#)

al angela atene berlusconi cdu cremlino democrazia elezioni eliseo europa fran **francia** francois germania **grece** hollande holstein land merkel MOSCA nea nicolas pasok presidenziali putin rai russia samaras sarkozy schleswig serbia silvio ue vladimir



Filter Panels

Current Selection

Score Date

AND OR

All RAI 1 RAI 2 RAI 3 RAI NEWS

From: 2012-03-28 To: 2012-04-12

Search by Entity
(press any key to open suggestions)

Search by Content
(put free text here)

Filter by TV Channel

Filter by Date
From: 2012-03-28 to: 2012-04-12

Top 50 Persons

Allegri Andrea Masiello Annan Belinelli **Benedetto** Calderoli Calearo Cini **Corrado Passera** Di Durante Elia **Elsa Fornero Emilio Fedele** Fidel Castro Filippo Neri **Fornero Francesco Belsito** Galliani Gb Giorgio Chinaglia Giovanni Toti **Imu** Kofi Annan Luigi Lusi Luis Enrique Manchester City Maradona Marcegaglia **Mario Monti** Massimiliano Latorre Moratti Musy **Passera** Pier Ferdinando Casini **Pietro Renzo** **Renzo Bossi Roberto Maroni** Robinho Rosi Mauro Rosy Mauro Rutelli Salvatore Girone Schifani Sergio Marchionne Susanna Camusso **Suu Kyi** Totti Ddi

Top 50 Organisations

Al Qaeda **Alcoa** Apple Azamara **Azienda Bennet Bmw** Brigade Rosse Citroen Cnn **Coldiretti Concommercio** Credit Equitalia **Fan Club** Federconsumatori Fiamme Gialle Fiat Industrial Fincantieri **Finmeccanica Forza Italia** Foxconn **Gdf** Goldman Sachs **Guardia Di Finanza** Inps Interpol Italia Dei Valori Landini **Legambiente** Morgan Stanley **Ocse** Partito Democratico Qaeda **Regione Lazio** Regione Lombardia Rizzoli Samsung Siae Snam **Sony** **Telecom** Telecom Italia Trenitalia Ubi Banca **Uefa** Unipol Wall Street Journal Yahoo Youtube

Top 50 Locations

Abruzzo Ambiente America Anci Andrea Arte Asia Atalanta Aung Barca Barça **Belsito** Birmania Bot Chiesa

Search Result Lists

- ▶ EUROVISION News Exchanges
- ▶ RAI Hyper Media News Topics
- ▶ RSS Items
- ▶ TV News Stories
- ▶ RAI Multimedia Catalogue
- ▶ Google

Search Result Statistics

EUROVISION News Exchanges – Percentage Coverage by Country

1 222

- ▶ RAI Hyper Media News Topic Trend Analysis
- ▶ News Timeline Charts
- ▶ Category Pie Charts
- ▶ Source Pie Charts



Current Selection

- Score Date
- AND OR
- All RAI 1 RAI 2 RAI 3 RAI NEWS Others

Remove All
 From: 2012-06-24 To: 2012-07-09
 Wimbledon

Search by Entity

(press any key to open suggestions)

Search by Content

(put free text here)

Filter by Date

From: 2012-06-24 to: 2012-07-09

Top 50 Persons

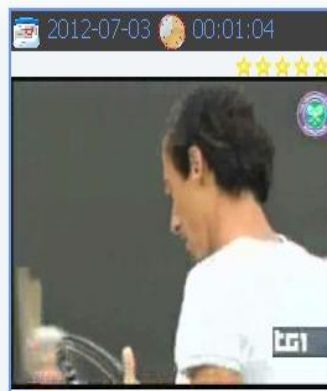
- Aga Agnieszka Radwanska
- Andy Murray** Azarenka
- Bellucci Berdych Camila
- Camila Giorgi** Cilic Clijsters
- David Ferrer Elena Baltacha
- Errani** Fabio Fognini
- Francesca Schiavone** Giorgi Ivanisevic
- Juan Martin Karin Knapp Karlovic
- Keothavong Kerber
- Knapp**
- Kohlschreiber Kvitova
- Laura Robson Lisicki Lukas Rosol
- Mahut Maria Sharapova Marin Cilic
- Nadia Petrova** Novak Djokovic
- Paszek Pete Sampras
- Petra Kvitova
- Pironkova Rafa Nadal
- Roberta Vinci** Rosato Rosol Sampras

- ▶ EUROVISION News Exchanges
- ▶ RAI Hyper Media News Topics
- ▶ RSS Items
- ▼ TV News Stories

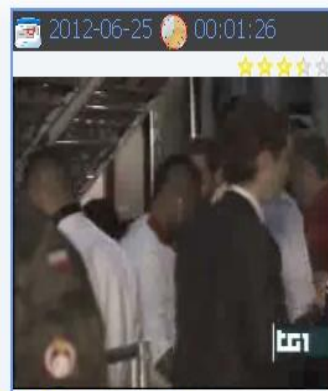
« Previous 1 Next » (Displaying 1 to 3 of 3 results)



Rai 1 TG1 20:00 (22)
 denis lo svizzero roger federer ha vinto per la settima volta il torneo di **wimbledon** in finale ha battuto il britannico andrew murray in quattro set



Rai 1 TG1 08:00 (18)
 a **wimbledon** e' finita ko francesca schiavone capo e' sempre stata contro la campionessa uscente del torneo la ceca petra ritrovati quest' ultima frontiera



Rai 1 TG1 13:30 (13)
 i due giocatori che hanno sbagliato i rigori piu' realista e sane che titola nessuno pronto per il torneo di **wimbledon** legge

- ▶ RAI Multimedia Catalogue
- ▶ Google

- ◆ **The media industry is moving fast**
- ◆ **New markets, new trends, new technologies for end-users mean new challenges and new requirements and a lot more content items**
- ◆ **Indexing, integrating and accessing multimodal content in an efficient way is a crucial factor**
- ◆ **It's time to adopt the more mature results of automation systems and open source solutions in our production infrastructures**

**The future is:
make data (re) use and metadata
cheaper and quicker**



◆ RAI

- Centre for Research and Technological Innovation (CRIT)
 - ✓ <http://www.crit.raai.it/EN/home.htm>
- Automatic Newscast Transcription System (ANTS)
 - ✓ http://tech.ebu.ch/docs/techreview/trev_2008-Q1_ants-dimino.pdf
- Hyper Media News (HMN)
 - ✓ <http://www.crit.raai.it/EN/attivita/archivi/e-Archivi-2.pdf>
- Interactive Newsbook
 - ✓ <http://www2012.org/proceedings/companion/p389.pdf>

◆ Apache OpenNLP

- Homepage (download, license, documentation,...)
 - ✓ <http://opennlp.apache.org/>
- Models
 - ✓ <http://opennlp.sourceforge.net/models-1.5/>

◆ Apache Solr

- Homepage (download, features, documentation, Wiki,...)
 - ✓ <http://lucene.apache.org/solr/>
- AJAX Solr library (demo, documentation, download)
 - ✓ <https://github.com/evolvingweb/ajax-solr>



maurizio.montagnuolo@rai.it



Thank you for
your attention