



Trattamento Automatico del Contenuto Audiovisivo per Servizi Innovativi di Informazione

Alberto **Messina***, Maurizio **Montagnuolo** e Roberto **Borgotallo**
Rai - Centro Ricerche e Innovazione Tecnologica
Torino

*Attualmente affiliato anche al
Dipartimento di Informatica dell'Università degli Studi di Torino

1. INTRODUZIONE

La moderna era digitale ha portato nella vita di ciascuno di noi alcuni cambiamenti epocali. Ciò che qualche anno fa, prima dell'avvento di Internet e delle tecnologie digitali, poteva sembrare irraggiungibile dal punto di vista tecnologico, e quindi posizionato idealmente in un quadro che sapeva più di fantascienza che di concreta esistenza, oggi è largamente entrato a far parte della nostra quotidianità. Gli attuali dispositivi di registrazione e riproduzione di largo consumo sono infatti in grado di memorizzare ingenti quantità di dati (da alcuni Gbyte dei riproduttori personali di musica a qualche centinaio di Gbyte per i Personal Video Recorder di ultima generazione) e di riprodurre correttamente pressoché qualsiasi contenuto multimediale.

Sommario

Questo articolo illustra i contenuti di una delle dimostrazioni che si tiene contemporaneamente al Prix Italia 2009 (Torino, 20-26 settembre), e che riguarda le tecnologie di trattamento automatico del materiale audiovisivo per servizi innovativi di informazione. Durante la dimostrazione sono presentate le funzionalità di due sistemi sviluppati dal Centro Ricerche e Innovazione Tecnologica della Rai e supportati dalla direzione Rai Teche: ANTS e Hyper Media News.

L'articolo fornisce inoltre una breve analisi delle motivazioni fondamentali alla base di questi sviluppi, e alcuni cenni descrittivi a proposito delle tecnologie utilizzate dai due sistemi in oggetto, fornendo anche riferimenti bibliografici approfonditi per i lettori più interessati.



La naturale evoluzione culturale che questo cambiamento ha portato con sé ha fatto nascere nella società e nei singoli individui la consapevolezza, e di conseguenza l'abitudine a pensare, che qualsiasi tipo di contenuto multimediale – fotografie, filmati, suoni, testi – possa essere fruito in maniera indifferente al tempo (in qualsiasi momento) e allo spazio (ovunque ci si trovi). Oltre che i contenuti di carattere personale, questa idea di annullamento delle barriere spazio-temporali all'accesso dei contenuti ha cominciato a riguardare anche i contenuti di carattere istituzionale e professionale, quali ad esempio gli archivi radiotelevisivi, in particolare quelli della Rai.

Una delle dimostrazioni organizzate nel periodo in cui si tiene a Torino il Prix Italia (20-26 settembre 2009) illustra le funzionalità di una piattaforma integrata di ricerca e fruizione di contenuti informativi telegiornalistici basata su tecnologie di estrazione ed elaborazione automatica dell'informazione, interamente sviluppata dal Centro Ricerche e Innovazione Tecnologica della Rai, e sponsorizzata dalla Direzione Rai Teche.

2. MOTIVAZIONI

2.1 RIVOLUZIONE DIGITALE

Gli archivi Rai hanno conosciuto una notevole rivitalizzazione a partire da metà degli anni '90, quando il progetto di ridocumentazione degli archivi guidato da Rai Teche, il Catalogo Multimediale [1], ha rappresentato la pietra angolare di un nuovo approccio filosofico e tecnologico allo stesso tempo: la fruizione informatizzata dei contenuti di archivio.

Il Catalogo Multimediale, originariamente, era stato pensato come strumento di supporto alla produzione di contenuti audiovisivi, e perciò dedicato ad una schiera ristretta di utilizzatori, per lo più interni all'azienda stessa.

Grazie all'impetuosa evoluzione tecnologica e scientifica accaduta negli anni seguenti l'esordio del Catalogo Multimediale, le moderne tecnologie dell'informazione permettono oggi a tutti gli utenti della rete di utilizzare schemi di fruizione dei contenuti multimediali basati sul paradigma di "Ritrovamento

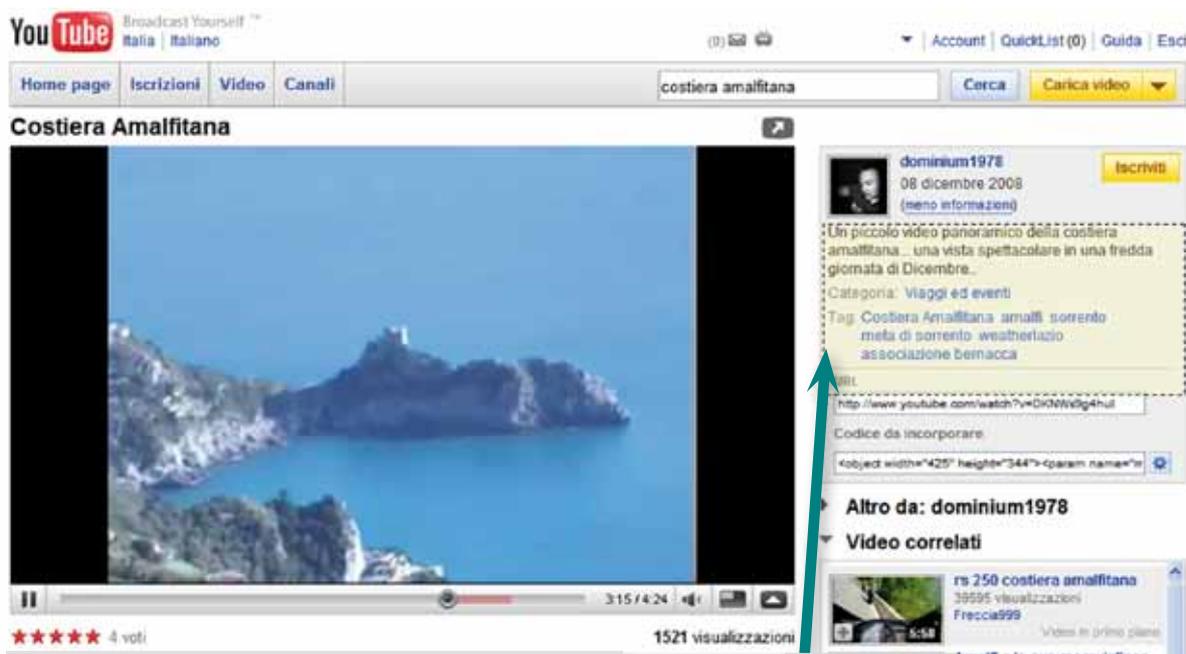


Fig. 1 - Esempio di contenuto indicizzato su YouTube.

annotazione



e Recupero”^{Nota 1}, ovvero sia seguendo un processo costituito da due passi fondamentali: la specifica di un’interrogazione e la sua sottomissione ad un sistema di ricerca, seguita dal recupero e fruizione degli oggetti ritrovati dal sistema di ricerca.

Questo processo, apparentemente semplice e diretto, presenta tuttavia dei requisiti di funzionamento non banali. Per funzionare correttamente, infatti, il sistema di ricerca deve conservare al suo interno degli indici in grado di organizzare in maniera efficiente e quanto più possibile completa l’informazione contenuta negli oggetti di archivio, similmente alla funzione che un indice analitico svolge per le pubblicazioni bibliografiche. Questo requisito implica che qualche forma di analizzatore intelligente osservi a priori il contenuto e costruisca questo indice seguendo dei criteri definiti e possibilmente oggettivi.

La forma più comune esistente in natura di tale analizzatore è rappresentata da ciascuno di noi, quando è intento ad annotare il proprio album digitale di fotografie su Flickr^{Nota 2} o il proprio canale di videoclip su YouTube^{Nota 3}, utilizzando schemi di etichette (tag) che evocano il contenuto di ciò che stiamo descrivendo, e che possono essere utili per far ritrovare gli stessi contenuti ad altri utenti di questi sistemi (figura 1).

I livelli di qualità richiesti per la documentazione dell’archivio radiotelevisivo professionale sono tuttavia di gran lunga più alti di quelli che possono essere assicurati da una breve descrizione evocativa o da qualche etichetta linguistica.

La funzione fondamentale che l’archivio svolge di sorgente di contributi per le nuove produzioni impone, infatti, che i contenuti siano indicizzati temporalmente, cioè lungo la loro dimensione di riproduzione (esempio di figura 2), e che siano specificate

informazioni dettagliate sui partecipanti alla scena. Inoltre è estremamente utile fornire un indice di accesso multimediale al contenuto, costituito da immagini rappresentative delle inquadrature, per guidare l’utente del sistema nella sua selezione.

Grazie alla nostra esperienza di “annotatori” piuttosto naïf, possiamo immaginare quanto sia tedioso e ripetitivo il lavoro della nostra controparte professionale, l’archivista documentatore, e quanto di conseguenza questa attività possa essere lunga, soggetta ad errori e disuniformità, e talvolta anche piuttosto costosa.

Nota 1 - Search & retrieval, nella terminologia tecnico/scientifica internazionale

Nota 2 - www.flickr.com

Nota 3 - www.youtube.com

Sigla iniziale (Dal Programma: TG2 DOSSIER STORIE del 2009-08-02)
00:00 02/08/2009, CMM News
Comentario musicale.

Partecipano:

Sommario (Dal Programma: TG2 DOSSIER STORIE del 2009-08-02)
00:00 02/08/2009, CMM News

Su sottofondo musicale, e anche in commento speaker, la conduttrice Mattei annuncia i temi del servizio che andranno in onda durante la puntata. **descrizione**

Inquadrature

Partecipano:

Maria Concetta Mattei () come Conduttori

Conticello contro la mafia palermitana (Dal Programma: TG2 DOSSIER STORIE del 2009-08-02)
00:00 02/08/2009, CMM News

Servizio: "Palermo amara". Su sottofondo musicale, commento speaker sulla situazione in cui vive Vincenzo Conticello, titolare dell'Antico focacciera San Francesco nel centro storico di Palermo, che ha denunciato e fatto arrestare i suoi estorsori e sul calo del numero dei clienti del suo ristorante a causa della sua posizione contro la malavita locale. Intervista a Conticello. Momenti di un processo in cui Conticello accusa i suoi estorsori.

crediti

Partecipano:

Maria Luisa Liberatori () come Montaggio
Vincenzo Conticello (ristoratore) come Intervistati
Daniela Orsello () come Servizio di

Fig. 2 - Esempio di annotazione analitica dal Catalogo Multimediale Rai.



2.2 NUOVI MODELLI DI FRUIZIONE

A rendere ancora più critica la situazione, gli attuali modelli di fruizione dei contenuti multimediali sulla rete impongono che questi siano resi disponibili quanto più velocemente ed efficacemente possibile, e attraverso una vasta scelta di modalità di ricezione. Come effetto della possibilità di interagire direttamente con i produttori e distributori di contenuti, si assiste inoltre al fenomeno della cosiddetta long tail [2]^{Nota 4}, la quale instaura meccanismi di domanda/offerta che rendono inefficace e non più praticabile il modello di trasmissione "da uno verso molti", ovvero il broadcasting tradizionale, per preferire modalità di interazione basate sul web (figura 3) o modalità ibride basate su dispositivi in grado contemporaneamente di ricevere il segnale digitale radiodiffuso e connettersi a servizi Internet.

Questi fenomeni hanno come effetto cumulativo quello di imporre vincoli produttivi senza precedenti per i produttori e distributori di contenuti, in particolare quello di fornire accesso incondizionato ai contenuti, cioè sulla base di preferenze espresse attraverso interrogazioni o dedotte dal comportamento implicito degli utilizzatori. Ciò, in ultima analisi, si traduce in un requisito stringente di minimizzare il lasso di tempo intercorrente tra la produzione dei contenuti e la loro indicizzazione.

La migrazione verso il mondo web pone un'altra problematica fondamentale, ancora in gran parte se non completamente irrisolta: l'integrazione dell'informazione presente sul web (per esempio su Wikipedia^{Nota 5} o sui siti di informazione online) con l'informazione proveniente dai contenuti multimediali pubblicati sui portali di accesso WebTv.

Nota 4 - Con questa espressione si indica in statistica la proprietà generale di alcune funzioni di densità di probabilità di possedere delle "code", cioè delle zone periferiche rispetto all'intervallo di variabilità più probabile, nelle quali il valore assunto dalla funzione degrada lentamente. Esempi di tali funzioni sono le distribuzioni di Pareto. Metaforicamente, l'effetto "long tail" nel mercato dell'informazione starebbe ad indicare un fenomeno per il quale singoli oggetti potenzialmente vendibili in maniera indipendente a molti compratori possono generare un volume di affari paragonabile a quello generato da pochi oggetti popolarissimi.
Nota 5 - it.wikipedia.org.



Fig. 3 - Esempio di portale WebTv.



Soprattutto nell'ambito dell'informazione quotidiana (telegiornali), questa possibilità di integrazione rappresenta un fondamentale contributo alla pluralità e completezza dell'informazione, nonché alla possibilità di costruire basi storiche di informazione con valenza archivistica.

2.3 I TRE PUNTI IMPORTANTI

Volendo riassumere in alcuni punti fondamentali l'analisi condotta fin qui, potremmo individuare i seguenti elementi imprescindibili:

- 📖 L'indicizzazione dei contenuti multimediali è generalmente essenziale al loro ritrovamento e alla loro fruizione, indipendentemente dal modello di distribuzione;
- 📖 Le modalità di fruizione basate sul web, grazie all'interattività, impongono vincoli produttivi estremamente più stringenti rispetto al passato;
- 📖 La migrazione verso il dominio web implica la necessità di integrare l'informazione proveniente da sorgenti eterogenee e indipendenti.

Questi tre punti cardinali costituiscono la mappa di riferimento nella quale sono situati gli studi e i prototipi oggetto della presente dimostrazione, che saranno brevemente illustrati nel seguito di questo articolo.

3. IL SISTEMA ANTS/HYPER MEDIA NEWS

Per contribuire a risolvere alcune delle problematiche illustrate nella prima parte di questo articolo da alcuni anni il Centro Ricerche e Innovazione Tecnologica della Rai sta studiando e realizzando sistemi di supporto all'attività di annotazione dei contenuti basati sull'applicazione di tecniche automatiche di elaborazione ed estrazione dell'informazione. I principali sistemi in oggetto sono ANTS (Automatic Newscast Transcription System)[3][4] e Hyper Media News [5][6], che saranno brevemente illustrati nei paragrafi seguenti, e che sono l'effettivo oggetto della dimostrazione.

3.1 ANTS

ANTS è un sistema integrato per l'annotazione automatica di contenuti telegiornalistici. Le principali funzionalità del sistema sono le seguenti:

- 📖 Trascrizione automatica del parlato in testo
- 📖 Segmentazione del contenuto in unità informative elementari
- 📖 Classificazione per contenuto delle unità informative elementari
- 📖 Estrazione di elementi semantici dalla trascrizione del parlato

La trascrizione del parlato in testo avviene attraverso l'impiego di un software che modella i fenomeni acustici di emissione vocale con semplici processi Markoviani^{Nota 6}, sfruttando le caratteristiche statistiche del segnale e le caratteristiche statistiche dei fonemi e delle loro relazioni di precedenza e successione tipiche dell'espressione naturale di una certa lingua. Inoltre il software fa uso di un modello linguistico che conserva la statistica di co-occorrenza dei termini di un dizionario di riferimento, al fine di ulteriormente migliorare i risultati ottenibili. Per ulteriori dettagli in merito a questa tecnologia si consiglia la lettura di [7] e [8].

La segmentazione del contenuto in unità informative elementari avviene grazie allo sfruttamento di alcune regole osservate nella maggior parte dei telegiornali, e alla loro implementazione algoritmica.

Le tre regole empiriche implementate sono:

- a. il soggetto parlante più frequente coincide con il telecronista da studio (anchorman);
- b. in corrispondenza di ogni apparizione dell'anchorman inizia una nuova unità informativa elementare;
- c. eventuali cambi di inquadratura sull'anchorman coincidono con l'inizio di unità informative elementari.

Nota 6 - Un processo Markoviano discreto è un insieme di stati e di possibili transizioni di stato nel quale la probabilità di transizione da uno stato x ad uno stato y dipende solamente dallo stato x e non dalla storia precedente. Questi processi sono altrimenti detti processi "senza memoria".



La classificazione per contenuto delle unità elementari rilevate si basa sulla modellazione statistica della probabilità condizionata che la presenza di alcuni termini chiave nel contenuto parlato implicino un certo genere di argomenti (ad esempio, la parola "Papa" implica generalmente che l'argomento dell'unità in cui essa compare sia di carattere religioso).

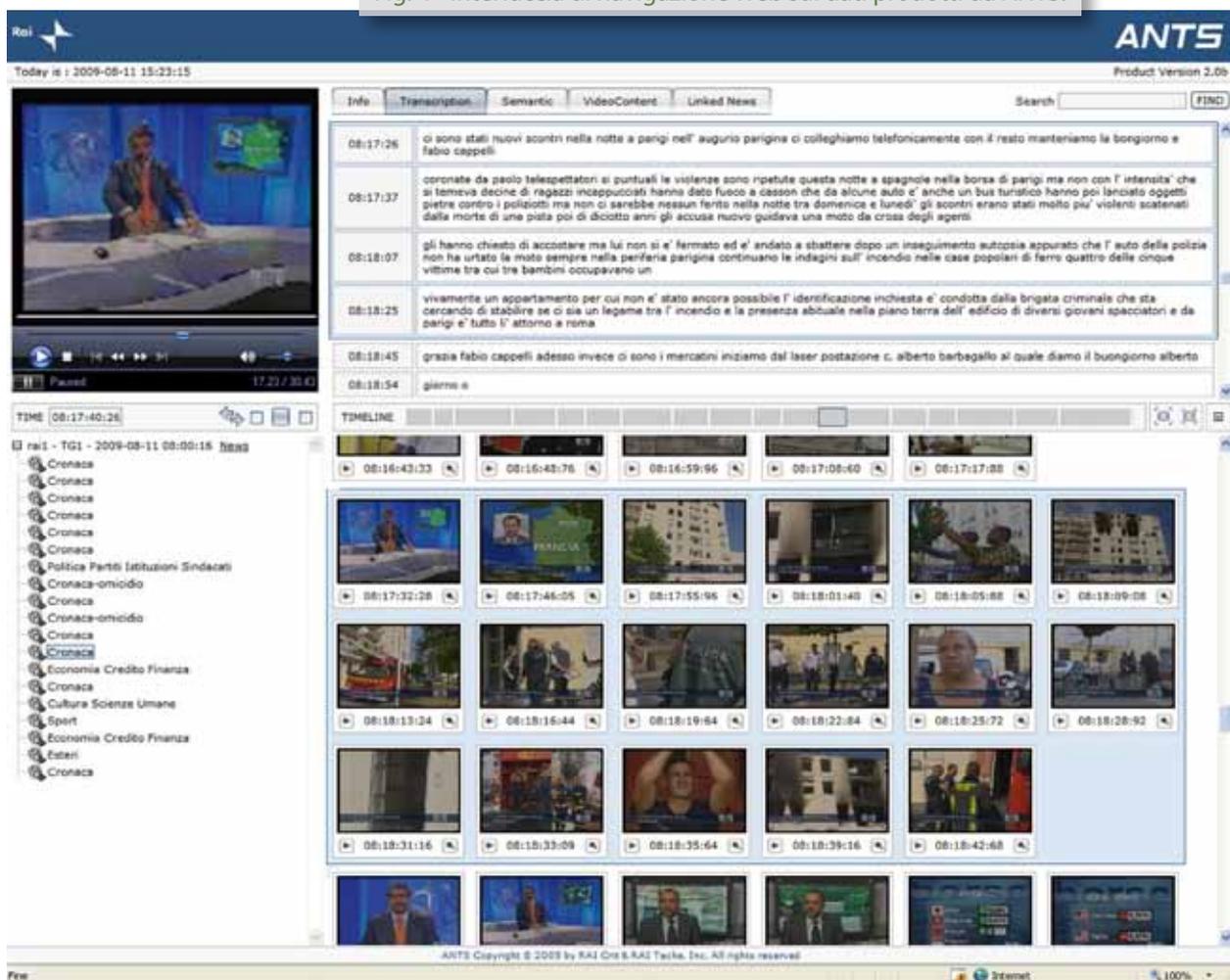
L'estrazione di elementi semantici (quali ad esempio nomi di persone, luoghi) dal contenuto parlato avviene attraverso una modellazione statistica della probabilità condizionata che certe categorie di parole precedano o seguano elementi di questo genere [9]. Per esempio la probabilità che la parola "X" sia un nome di persona dato che è preceduta

dalle parole "Presidente della Repubblica" è significativamente maggiore della probabilità che una generica parola "Y" presa a caso possa esserlo.

I risultati dei processi delineati sono pubblicati in maniera sinottica attraverso un'interfaccia di navigazione web, esemplificata in figura 4.

Grazie alla sua completa automatizzazione e ottimizzazione, il sistema ANTS sopprime a due delle criticità identificate nei precedenti paragrafi: l'indicizzazione e l'efficienza e velocità di accesso ai contenuti. L'integrazione dei contenuti televisivi con le fonti di informazioni indipendenti è invece oggetto del sistema Hyper Media News.

Fig. 4 - Interfaccia di navigazione Web sui dati prodotti da ANTS.





3.2 HYPER MEDIA NEWS

Hyper Media News è un sistema in grado di integrare le informazioni generate automaticamente da ANTS con le informazioni presenti sul web e rese disponibili attraverso i siti di informazione quotidiana online^{Nota 7}. Il principio di funzionamento di questo sistema si basa su un'innovativa tecnologia per l'aggregazione ibrida di oggetti, cioè il raggruppamento di oggetti di tipologia differente quali sono gli articoli testuali ritrovabili sui siti web di informazione e le unità elementari di informazione televisiva generate da ANTS. Per maggiori dettagli tecnici si rimanda alla lettura di [5] e [6].

La tecnologia sviluppata permette anche di selezionare gli elementi rappresentativi dei raggruppamenti trovati, così da permettere l'implementazione di servizi di informazione innovativi, quali:

- 📖 Servizi RSS^{Nota 8} multimediali
- 📖 Servizi Hot News e Topic Tracking
- 📖 Navigazione visuale avanzata

I servizi RSS multimediali sono costituiti dalla pubblicazione di canali RSS^{Nota 9} contenenti i raggruppamenti trovati dal sistema, organizzati in maniera sequenziale nel tempo oppure indicizzati sulla base delle preferenze degli utenti, come esemplificato dalla figura 5.

Nota 7 - Il sistema Hyper Media News, assieme ad ANTS, è già stato dimostrato con successo in diverse occasioni durante conferenze scientifiche internazionali quali: Semantic and Media Technologies (SAMT) 2008 (<http://samt2008.uni-koblenz.de/>), International Conference on Multimedia and Expo (ICME) 2008 (<http://www.icme2008.org>), International Conference on Data Mining (ICDM) 2008 (<http://icdm08.isti.cnr.it/>), Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) 2008 e 2009 (rispettivamente <http://wiamis2008.itec.uni-klu.ac.at/> e <http://wiamis2009.qmul.net/>).

Le tecnologie di base del sistema sono state anche presentate durante seminari accademici presso l'Università degli Studi di Torino e il Politecnico di Torino.

Nota 8 - Really Simple Syndication

Nota 9 - Un canale RSS è una risorsa web formattata secondo le specifiche RSS, un dialetto XML, che rappresenta liste di elementi in ordine cronologico di pubblicazione. Per maggiori dettagli si veda www.rss-specifications.com.

Fig. 5 - Servizi RSS multimediali forniti da Hyper Media News.



I servizi di Hot News sono costituiti dalla pubblicazione di rapporti giornalieri contenenti le notizie più importanti della giornata, con collegamenti ai servizi RSS contenuti e annotati automaticamente con parole chiave (tag). Il sistema mantiene un archivio storico di questi rapporti, per eventuale consultazione futura. Un esempio di pagina di Hot News è riportata nella figura 6.

Il sistema offre agli utenti più avanzati una modalità di navigazione visuale basata sulla rappresentazione dei raggruppamenti in forma di grafo, come esemplificato dalla figura 7. L'utente può selezionare visualmente sottoinsiemi del grafo, che corrispondono a sottoargomenti dell'argomento principale. E' possibile, a partire dai nodi selezionati, fruire delle clip televisive e degli articoli web aggregati, in un'unica interazione sinergica.

4. DESCRIZIONE DELLA DIMOSTRAZIONE

La dimostrazione che si tiene contemporaneamente al Prix Italia 2009 consta sostanzialmente dell'illustrazione delle funzionalità del sistema ANTS e Hyper Media News, anche attraverso la partecipazione diretta del pubblico, che potrà sperimentarne direttamente l'utilizzo attraverso una comune interfaccia web.

5. CONCLUSIONE

Le moderne tecnologie e l'evoluzione dei paradigmi di fruizione dell'informazione pongono nuove sfide ai produttori e distributori di contenuti, soprattutto in termini di efficienza, usabilità e completezza dell'informazione. Le tecnologie all'avanguardia



Fig. 6 - Il servizio Hot News. I due argomenti classificati come più importanti della giornata del 10 agosto 2009.



nella gestione automatizzata del contenuto audiovisivo sviluppate negli ultimi dal Centro Ricerche e Innovazione Tecnologica della Rai costituiscono un tassello importante per colmare queste lacune, come anche riconosciuto a livello internazionale dalla comunità tecnica e scientifica del settore.

Al Prix Italia, i progettisti ed ideatori di questi sistemi sono lieti di rendere partecipe il pubblico di queste innovazioni, cogliendo allo stesso tempo l'importante opportunità di conoscere e valutare le reazioni e le valutazioni dei potenziali utilizzatori di questi servizi.

Fig. 7 - Navigazione visuale avanzata.

Esodo, giornata da bollino rosso





BIBLIOGRAFIA

1. R. Del Pero, G. Dimino, M. Stroppiana, "Catalogo Multimediale: l'esperienza Rai", *Elettronica e Telecomunicazioni*, n. 1, Aprile 2000.
2. C. Anderson, "The Long Tail" *Wired*, October 2004.
3. A. Messina, R. Borgotallo, G. Dimino, L. Boch, D. Airola Gnota, "An Automatic Indexing System for Television Newscasts", *IEEE ICME 2008*, Hannover, June 2008.
4. R. Borgotallo, G. Dimino, A. Messina, "ANTS: a complete system for automatic news programme annotation based on audiovisual content and text analysis", *EBU Technical Review* nr. 313, Geneva, March 2008.
5. A. Messina, M. Montagnuolo, "A Generalised Cross-Modal Clustering Method Applied to Multimedia News Semantic Indexing and Retrieval", *18th International Conference on World Wide Web*, Madrid, April 2009.
6. M. Montagnuolo, M. Ferri, A. Messina, "HM-News: an Integrated System for Searching and Browsing Hypermedia News Content", *HyperText 2009*, Torino, June 2009.
7. L.R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition", *Proceedings of the IEEE*, February 1989, Volume: 77, Issue: 2, page(s): 257-286.
8. F. Brugnara, M. Cettolo, M. Federico, D. Giuliani, "A system for the segmentation and transcription of Italian radio news", in *Proceedings of RIAO, Content-Based Multimedia Information Access*, Paris, France, 2000.
9. R. Basili, M. Cammisa, E. Donati, "RitroveRAI: A Web Application for Semantic Indexing and Hyperlinking of Multimedia News", in *Proc. of "International Semantic Web Conference"*, *Lecture Notes in Computer Science*, LNCS 3279, 97-111, Springer, 2005.